

Browser-Typografie

Untersuchungen zur Lesbarkeit von Schrift
im World Wide Web

Dissertation

zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.)

Zugelassen am 30.11.2005 durch den Promotionsausschuss
für die Fakultät Kulturwissenschaften
der Universität Dortmund

Vorgelegt von

Dipl.-Journalist Martin Liebig
geb. 24.06.1970
in Wuppertal

Vorgelegt am

15.11.2006

Gutachter

Prof. Dr. Ulrich Pätzold, Universität Dortmund
Prof. Dr. Günther Rager, Universität Dortmund

Martin Liebig
Glünderstraße 14
30167 Hannover
info@blickwerkstatt.de

für uns

Kurzzusammenfassung

Das World Wide Web ist innerhalb von knapp anderthalb Jahrzehnten zum Massenmedium avanciert. Jeder zweite Bundesbürger im Alter von über 13 Jahren ging 2006 regelmäßig online – die meisten davon mit dem Ziel, aktuelle Informationen über das regionale, überregionale und das Weltgeschehen ausfindig zu machen. Entsprechend hoher Zugriffsraten erfreuen sich Anbieter explizit journalistisch konzipierter General-Interest-Websites. Marktführer sind nach anerkannter Statistik die Print-Ableger „bild.de“, „Spiegel online“ und „stern.de“.

Allen multimedialen Potenzialen des Internets zum Trotz findet journalistische Vermittlung im Web nach wie vor überwiegend in schriftlicher Form statt. Da mutmaßlich nur ein Bruchteil von Online-Nutzern Webseiten ausdruckt, um diese von Papier zu rezipieren, liegt die Kreation einer lesefreundlichen Bildschirm-Typografie im Kerninteresse von Anbietern journalistischer Information im Web.

Was aber zeichnet guten Schriftsatz im Web aus? Nach herrschender Meinung entfaltet Typografie auf Monitoren zumindest teilweise andere Gesetzmäßigkeiten, als sie für den Papierdruck in den vergangenen Jahrhunderten entwickelt und teilweise empirisch untermauert wurden. Lesen am selbststrahlenden Medium Bildschirm gilt als anstrengend für die beständig direktem Licht ausgesetzten Augen. Verschiedene Studien legen den Schluss nahe, dass das Lesen vom Monitor zwar nicht unbedingt langsamer vonstatten geht als von Papier, dabei aber flüchtiger und weniger detailfokussiert ist. Auch sind Schriftzeichen am Monitor in aller Regel erheblich gröber aufgelöst als auf Papierdrucken, was ihre Rezeption potenziell erschwert. Nicht zuletzt lassen sich Unterschiede zwischen papierbasierter und Web-Typografie am Programmtyp „Browser“ fest machen, der insbesondere die Auswahl und Kontrolle von Schriftarten und Schriftgrößen erheblich schwieriger geraten lässt als im Print-Layout.

Es besteht mithin Bedarf nach einer Beschreibung von Gesetzmäßigkeiten einer mediumspezifischen, einer allgemeinen Monitor- und einer spezifischen Webbrowser-Typografie. Es ist nur ein scheinbares Indiz für gesicherte Erkenntnisse in diesem Gebiet, dass die typografischen Gestaltungen führender journalistischer Web-Angebote sich derzeit zum Teil frappierend ähneln. Denn die Meinungen in der einschlägigen Literatur gehen bisweilen erheblich auseinander hinsichtlich wichtiger Aspekte wie Schriftwahl, Setzung von Zeilenbreiten oder der Definition von Zeilenabständen auf Webseiten. Auch die empirische Forschung zum Thema Web-Typografie macht zwar Fortschritte, hat aber in vielen Teilbereichen sehr widersprüchliche oder wenig erhellende Ergebnisse zutage gefördert. Vereinzelt typografische Wirkfaktoren wie etwa der Zeilenabstand sind bis heute empirisch praktisch unhinterfragt.

Die meisten der vorliegenden einschlägigen Studien kranken insbesondere an einem zentralen Defizit: Sie widmen sich sehr isoliert einem vereinzelt typografischen Faktor, beispielsweise der Schriftart, ohne dessen Wechselbeziehungen mit anderen relevanten mikrotypografischen Ausprägungen zu berücksichtigen. Diese wechselseitige Interaktion von immerhin

vier anerkannt wichtigen typografischen Einzelfaktoren erforschte daher ich in einem Experiment, in dem ich Lesegeschwindigkeiten unter den Bedingungen von insgesamt 90 typografischen Konstellationen maß – indem ich nämlich wechselseitig 5 Zeilenbreiten mit 2 Schriftarten in 3 verschiedenen Schriftgrößen mit 3 verschiedenen bemessenen Zeilenabständen kombinierte und präsentierte ($5 \times 2 \times 3 \times 3 = 90$). An diesem als Feldstudie konzipierten Online-Experiment nahmen im Januar 2006 über 1400 Menschen teil.

Überraschenderweise förderte das Experiment nicht nur typografische Erkenntnisse zutage, sondern führte zur Identifikation zweier Wirkfaktoren im Leseprozess am Bildschirm, deren Einfluss ich in dieser Eindeutigkeit nicht erwartet hatte: Lebensalter der Nutzer und Konstruktionsart des Monitors. So lasen Teilnehmende im Alter zwischen 19 und 35 Jahren die Experimentaltexte im Schnitt erheblich schneller als die minderjährigen und insbesondere die über 55-jährigen Probanden. Zudem ergab eine Varianzanalyse, dass Texte an Flachbildschirmen nicht nur etwas schneller rezipiert werden als an Röhrenbildschirmen. Es zeigte sich auch, dass einige typografische Befunde sich je nach Monitorgattung teilweise ausgeprägter darstellten als in der Gesamtbetrachtung über alle Bauarten hinweg – sich diese Befunde teilweise aber auch nur an einem der beiden Bildschirmtypen überhaupt manifestierten.

Insgesamt waren die auffälligsten Befunde,

- *dass der wichtigste typografische Wirkfaktor offenbar die Zeilenbreite ist.*

Die Ergebnisse meiner Studie legen nahe, dass Zeilen, die durchschnittlich 40 bis 50 Schriftzeichen Platz bieten, die besten Lesezeiten induzieren. Kürzer bemessene Zeilen führen dagegen zu einer empfindlichen Verschlechterung der Lesezeiten. In abgeschwächter Form gilt dies auch für Zeilenbreiten ab vermuteten 65 Anschlägen aufwärts. Zeilenbreiten zwischen 50 und 65 Anschlägen scheinen etwas schlechter lesbar als moderat kürzer bemessene, aber unter Einschränkungen geeignet. Eine Wechselwirkung des Faktors „Zeilenbreite“ mit dem Faktor „Zeilenabstand“ scheint, wiewohl in der Literatur gelegentlich behauptet, nicht zu existieren. Auch die Faktoren „Schriftart“ und „Schriftgröße“ interagieren nach meinen Erkenntnissen in keiner wechselseitigen Kombination mit der Zeilenbreite. All die beschriebenen Befunde zeigten sich an Flachbildschirmen ausgeprägter als an Röhrenbildschirmen.

- *dass ein weiterer, sehr wichtiger typografischer Einflussfaktor die Schriftgröße ist.*

Schrift in Webbrowsern sollte nach den Ergebnissen meiner Studie in ihrer Größe so definiert sein, dass zur Darstellung der Kleinbuchstaben mindestens 7 Bildschirmpunkte (Pixel) in der Vertikalen zur Verfügung stehen. Kleinere Bemessungen führen zu merklicher Verlangsamung der Lesegeschwindigkeiten, eine Erhöhung auf 8 Pixel beschleunigt den Rezeptionsprozess zumindest an Flachbildschirmen nicht weiter. Es ergab sich zudem eine starke Wechselwirkung zwischen Schriftgröße und Schriftart; so verbesserten sich die Lesezeiten für die Schriftart *Times* merklich mit wachsender Schriftgröße, während die *Verdana* in der größten gemessenen Schriftgröße (8 Pixel x-Höhe) langsamer rezipiert wurde als in der nächstkleineren (7 Pixel x-Höhe). Weitere Interaktionen fanden sich nicht für den Faktor Schriftgröße.

- *dass der Zeilenabstand so gut wie keinen Einfluss ausübt auf die Lesbarkeit.*

Laut Varianzanalyse spielt es praktisch keine Rolle, wie groß der Zeilenabstand von Web-Schrift bemessen ist, sofern dieser etwa das 2,5- bis 3-fache der jeweiligen Schrift-x-Höhe beträgt. Lediglich die *Verdana* scheint an Flachbildschirmen moderat bessere Lesezeiten zu induzieren mit linearer Vergrößerung des Zeilenabstands auf das bis zu 1,5-fache der numerischen Schriftgröße.

- *dass die Bedeutung der Schriftart-Wahl augenscheinlich überschätzt wird.*

Die in der Literatur oft unterstellte schlechtere Lesbarkeit serifentragender Schriftarten mit ausdifferenzierten Strichstärken gegenüber strichmonotoneren, serifenlosen Schriftarten bestätigte sich nicht. Die Anbietung der Serifen-Schriftart *Times* induzierte im Experiment praktisch identische Lesegeschwindigkeiten wie Präsentation der serifenlosen Schriftart *Verdana*.

- *dass die Bedeutung der Schriftart-Wahl eventuell doch nicht überschätzt wird.*

In einer das Experiment abschließenden Befragung bat ich die Teilnehmenden, Schulnoten zu vergeben für „Attraktivität“ sowie vermutete „Lesbarkeit“ von vier Schriftarten: *Times*, *Verdana*, *Frutiger* sowie *Arial*. Dabei schnitten *Verdana* und *Arial* unter beiden Kriterien mit der Durchschnittsnote 2 deutlich besser ab als die *Frutiger* (Note 3, ebenfalls praktisch analog für „Aussehen“ und „Lesbarkeit“). Die *Times* hingegen, in der Geschwindigkeitsmessung noch gleichauf mit der *Verdana*, erhielt für „Attraktivität“ wie „Lesbarkeit“ im Schnitt eine 3,5.

Die objektive, an der messbaren Lesegeschwindigkeit festgemachte Lesbarkeit von Schriftarten korrespondiert also offenbar nicht zwangsläufig mit deren durch die Nutzer erfahrenen „subjektiven Lesbarkeit“. Die *Times* erwies sich in meinem Experiment als gut lesbar – wahrgenommen als solche wurde sie nicht, im Gegenteil. Vielmehr scheinen Lesende dazu zu tendieren, Schriftarten als „gut lesbar“ zu empfinden, die sie „attraktiv“ finden – oder umgekehrt. Eine Wechselwirkung von Schrift-Ästhetik und Rezeptions-Adäquatheit scheint zumindest virulent und bedarf weiterer Hinterfragung und Messung.

Inhalt

Kurzzusammenfassung	I
Vorwort	XI
0 Einleitung	1
0.1 Das World Wide Web als journalistisches Medium.....	1
0.2 Ziel und Aufbau dieser Arbeit.....	4
0.3 Der Begriff der Typografie.....	6
0.4 Usability und Typografie.....	10
0.5 Typografie als Gegenstand der Journalistik.....	14
0.6 Bildschirm- und Webbrowser-Typografie.....	20
0.6.1 Typografie auf dem Monitor.....	21
0.6.1.1 Entstofflichung der Typografie: das Medium Monitor.....	21
0.6.1.2 Bildschirmmedien: Monitor-Konsum oder Ausdruck?.....	23
0.6.1.3 Die Flüchtigkeit des www oder: Wer hebt schon Webseiten auf?.....	23
0.6.1.4 Von der Bildschirm- zur Web-Typografie.....	25
0.6.2 Das Konzept des Webbrowsers.....	25
0.6.2.1 HTML.....	27
0.6.2.2 Das Prinzip der logischen Auszeichnung.....	28
0.6.2.3 Der „Sündenfall“: physische Textauszeichnung.....	29
0.6.2.4 Der „zahnlose Tiger“? Das W3C.....	31
0.6.2.5 Netscape, Microsoft und der „Browser-War“ der 1990er Jahre.....	31
0.6.3 Der steinige Weg: Typografie in Webbrowsern.....	33
0.6.3.1 Phase 1: Ausschließlich logische Textauszeichnung.....	34
0.6.3.2 Phase 2: Die ersten physischen Textauszeichnungen.....	34
0.6.3.3 Phase 3: Schriftvorschlag und Spaltensatz durch Tabellen-Layouts.....	35
0.6.3.4 Phase 4: Cascading Style Sheets (CSS).....	36
0.6.3.5 Die Unwägbarkeiten bestehen fort.....	40
0.6.4 Die „Marktanteile“ der Browser.....	42
1 Sprachfixierung und Leseprozess	43
1.1 Die Kulturleistung der Sprachfixierung.....	43
1.2 Die Systeme der Sprachfixierung.....	43
1.3 Der Prozess der Textdekodierung.....	46
1.3.1 Die Gestaltpsychologie als Grundlage.....	46
1.3.2 Die Theorie von „Wortbild“ und „Wortsilhouette“.....	49
1.3.3 Die Dekodierung von Einzel-Zeichen.....	52
1.3.3.1 Das Schablonen-Modell.....	52
1.3.3.2 Das Modell der Merkmalsanalyse.....	54
1.3.4 Zeichenrezeption und Kontextwissen.....	56
1.3.5 Sakkaden.....	58
1.4 Typografie im alphabetischen System.....	59
2 Lesen vom Papier, Lesen am Bildschirm	61
2.1 Besonderheiten des Bildschirm-Lesens.....	61
2.1.1 Leseabstand.....	61
2.1.2 Flimmern und Blendungen.....	61
2.1.3 Sitzhaltung und Betrachtungswinkel.....	62
2.1.4 Monitorauflösung.....	62
2.1.5 Weiteres Vorgehen.....	63
2.2 Die vermeintliche Entdeckung der Langsamkeit.....	63
2.2.1 Die Studien der 80er und frühen 90er Jahre.....	63

2.2.2 Studien aus der „Internet-Epoche“	65
2.2.2.1 Wissenserwerb – online und von Papier.....	66
2.2.2.2 Papier und Bildschirm – eine vergleichende Studie.....	66
2.2.2.3 Ziefles Studien.....	68
2.2.3 Würdigung der Ergebnisse: Papier versus Bildschirm.....	69
2.3 Der Leseprozess – Scannen, Skimmen, Lesen.....	70
3 Typografische Faktoren.....	73
3.1 Die Elemente von Schrifthanordnungen.....	73
3.2 Interdependenzen der typografischen Faktoren.....	73
3.3 Das Vorgehen in diesem Kapitel.....	74
3.3.1.1 Schwerpunkt der Betrachtung: Brottext.....	75
3.3.1.2 Nicht berücksichtigt: empirische Studien vor 1990.....	75
3.3.1.3 Die Rolle der printbasierten Lesbarkeitsforschung.....	76
3.3.1.4 Noch eine Bemerkung zur nicht-empirischen Quellenlage.....	78
3.4 Schriftart.....	78
3.4.1 Die Schriftart als typografisches Phänomen.....	78
3.4.1.1 Schrift als ästhetisches Gestaltungselement.....	79
3.4.1.2 Der Aspekt der Lesbarkeit.....	80
3.4.1.3 Und was ist „gute Lesbarkeit“?.....	81
3.4.1.4 Das Verhältnis Lesbarkeit – Schriftästhetik.....	82
3.4.1.5 Geschichte der Schrift.....	82
3.4.1.6 Mikrotypografische Begriffe.....	83
3.4.1.7 Schriftarten – eine deutsche Kategorisierung.....	84
3.4.2 Manipulation von Schriftarten auf HTML-basierten Webseiten.....	91
3.4.2.1 Problem Nummer 1: Schriftvorschlag statt Schriftfestlegung.....	92
3.4.2.2 Problem Nummer 2: die Bildschirmauflösung.....	93
3.4.3 Empfehlungen und Vorgaben aus der Literatur.....	99
3.4.3.1 Zwischenbemerkung: Verunstaltete Schrift – oder neuer Font?.....	101
3.4.3.2 Konsequenzen für die Schriftwahl im Web.....	102
3.4.3.3 Empfehlungen zur Schriftmischung.....	106
3.4.4 Alternativen zu den Standardschriften.....	107
3.4.4.1 True Doc und OpenType.....	108
3.4.4.2 Typografie-Bilder.....	110
3.4.4.3 Satz mit Adobe Flash.....	112
3.4.5 Am Horizont: Neue Windows-Schriften.....	113
3.4.6 Empirische Studien zur Schriftart am Bildschirm.....	114
3.4.6.1 Lesegeschwindigkeit und Leseverständnis: 4 Schriftarten im Vergleich.....	114
3.4.6.2 Lesedauer von 12 Schriftarten.....	115
3.4.6.3 Individuelle Einschätzung der Lesbarkeit durch die Probanden.....	116
3.4.6.4 Schriftart-Präferenzen von Nutzern.....	116
3.4.6.5 Einschätzungen von Schrift-Charakteristika durch Nutzer.....	117
3.4.6.6 Die Schriftarten Times, Arial und Courier im Vergleich.....	118
3.4.6.7 Die Schriftarten Verdana und Georgia in verschiedenen Vergleichskonstellationen.....	118
3.4.6.8 Bevorzugte Schriftarten im Web.....	119
3.4.6.9 Times versus Schmuckschrift.....	120
3.4.6.10 Zwei neue Schriftarten im Test.....	120
3.4.6.11 Schriftglättung mit „ClearType“.....	121
3.4.6.12 Ein älterer Schriftglättungs-Test.....	121
3.5 Schriftgröße.....	122
3.5.1 Die Schriftgröße als typografisches Phänomen.....	122
3.5.1.1 Der Punkt: Die wackelige Maßeinheit der Typografen.....	122
3.5.1.2 Unverwüstlich: der Bleikegel.....	124
3.5.1.3 Auch das noch: die optische Größe.....	126

3.5.2	Manipulation von Schriftgrößen in HTML-basierten Browsern.....	126
3.5.2.1	Die Definitionseinheiten in CSS.....	128
3.5.2.2	Der „Pixel“ – eine neue Maßeinheit?.....	129
3.5.3	Empfehlungen und Vorgaben aus der Literatur.....	133
3.5.3.1	Die Frage nach der Maßeinheit.....	133
3.5.3.2	Jenseits der Maßeinheit: Wie groß soll's denn nun sein?.....	135
3.5.4	Empirische Erkenntnisse zur Schriftgröße am Bildschirm.....	137
3.5.4.1	Schriftgrößen und Schriftarten am Bildschirm.....	137
3.5.4.2	Tempomessung mit Anti-Aliasing.....	138
3.5.4.3	8 Schriftarten in 3 Größen: Lesbarkeit und Attraktivität.....	139
3.5.4.4	Vier Schriftarten im Test.....	141
3.5.4.5	Schriftpräferenzen von Kindern.....	142
3.5.4.6	Lesegeschwindigkeit, -effizienz und Schriftpräferenzen älterer User.....	142
3.5.5	Zwischenfazit: Schriftarten und Schriftgrößen.....	143
3.6	Zeilenabstand.....	144
3.6.1	Der Zeilenabstand als typografisches Phänomen.....	144
3.6.2	Manipulation des Zeilenabstands in HTML-basierten Browsern.....	145
3.6.3	Empfehlungen und Vorgaben aus der Literatur.....	145
3.6.4	Empirische Erkenntnisse zum Zeilenabstand am Bildschirm.....	146
3.6.5	Zwischenfazit: Zeilenabstand.....	147
3.7	Zeilenbreite.....	147
3.7.1	Die Zeilenbreite als typografisches Phänomen.....	147
3.7.2	Manipulation der Zeilenbreite in HTML-basierten Webbrowsern.....	147
3.7.3	Empfehlungen und Vorgaben aus der Literatur.....	149
3.7.3.1	Allgemeine Vorgaben.....	149
3.7.3.2	Der Aspekt der Zeilenfall-Ästhetik.....	151
3.7.4	Empirische Studien zur Zeilenbreite am Bildschirm.....	153
3.7.4.1	Die Times in verschiedenen Zeilenbreiten.....	153
3.7.4.2	6 Zeilenbreiten im Vergleich.....	154
3.7.4.3	Vergleich zweier Zeilenbreiten.....	155
3.7.4.4	Drei Zeilenbreiten im Abgleich.....	155
3.7.4.5	25 und 80 Anschläge im Vergleich.....	156
3.7.4.6	6 Zeilenbreiten im Abgleich.....	156
3.7.4.7	Zeilenbreiten und Leseverständnis.....	158
3.7.4.8	Zeilenbreiten – wie sie auf Kinder und Erwachsene wirken.....	158
3.7.4.9	Vier Zeilenbreiten im Vergleich.....	159
3.7.4.10	Textverständnis in zwei Zeilenbreiten.....	160
3.7.5	Zwischenfazit: Zeilenbreiten.....	161
3.8	Schriftausrichtung.....	161
3.8.1	Die Schriftausrichtung als typografisches Phänomen.....	161
3.8.2	Manipulation der Schriftausrichtung in HTML-basierten Browsern.....	162
3.8.3	Empfehlungen und Vorgaben aus der Literatur.....	162
3.8.4	Empirische Erkenntnisse zur Schriftausrichtung am Bildschirm.....	163
3.8.5	Zwischenfazit: Schriftausrichtung.....	164
3.9	Auszeichnungsvarianten.....	164
3.9.1	Auszeichnungsvarianten als typografische Phänomene.....	164
3.9.2	Manipulation der Auszeichnung in HTML-basierten Browsern.....	165
3.9.3	Empfehlungen und Vorgaben aus der Literatur.....	166
3.9.4	Empirische Erkenntnisse zur Textauszeichnung am Bildschirm.....	169
3.9.4.1	Kursivsatz: Times, Arial und Courier.....	169
3.9.4.2	Kursivsatz: Verdana.....	169
3.9.4.3	Textfettung.....	169
3.9.5	Zwischenfazit: Textauszeichnungen.....	170

3.10 Laufweite	170
3.10.1 Die Laufweite als typografisches Phänomen.....	170
3.10.2 Manipulation der Laufweite auf HTML-basierten Webseiten.....	172
3.10.3 Empfehlungen und Vorgaben aus der Literatur.....	172
3.10.4 Empirische Studien zur Laufweite von Bildschirmschriften.....	172
3.11 Schriftfarbe und Farbgrund	172
3.11.1 Schriftfarbe und Farbgrund als typografische Phänomene.....	172
3.11.2 Manipulation von Farbe in HTML-basierten Browsern.....	173
3.11.3 Empfehlungen und Vorgaben aus der Literatur.....	174
3.11.4 Empirische Erkenntnisse zu Schriftfarbe am Bildschirm.....	177
3.11.4.1 Rote Schrift auf weißem Grund.....	177
3.11.4.2 Verschiedene Farbkombinationen im Vergleich.....	178
3.11.4.3 Schwarze Schrift auf planem und texturiertem Farb-Hintergrund.....	178
3.11.4.4 Differenzen zwischen grauer Schriftfarbe und grauen Farbgründen.....	179
3.11.5 Zwischenfazit: Schriftfarben.....	180
4 Das Experiment	181
4.1 Notwendigkeit einer Studie	181
4.2 Sinn und Machbarkeit einer Browser-Lesbarkeitsstudie	183
4.3 Lesbarkeit und ihre Operationalisierung	184
4.3.1 Legibility und Readability: Sichtbarkeit und Lesbarkeit.....	185
4.3.2 Die Legibility: Bekannte Forschungsdesigns.....	186
4.3.3 Ein Vorrang für die Readability.....	187
4.3.4 Mögliches Kriterium guter Lesbarkeit: Wissens-Zuwachs.....	187
4.3.4.1 Operationalisierbarkeit des Wissens-Zuwachses.....	188
4.3.4.2 Grenzen der Operationalisierbarkeit des Wissens-Zuwachses.....	188
4.3.4.3 Zwischenfazit: Wissens-Zuwachs als Lesbarkeits-Kriterium.....	190
4.3.5 Mögliches Kriterium guter Lesbarkeit: körperliche Beanspruchung.....	191
4.3.5.1 Operationalisierbarkeit des Grades körperlicher Beanspruchung.....	191
4.3.5.2 Zwischenfazit: körperliche Beanspruchung als Lesbarkeits-Kriterium.....	192
4.3.6 Mögliches Kriterium guter Lesbarkeit: Urteil der Probanden.....	192
4.3.6.1 Das Urteil der Probanden – kein sinnvoller Ansatz.....	193
4.3.6.2 Das Urteil der Probanden – jenseits der Lesbarkeit durchaus interessant.....	193
4.3.7 Mögliches Kriterium guter Lesbarkeit: Lesegeschwindigkeit.....	193
4.3.7.1 Operationalisierbarkeit der Lesegeschwindigkeit.....	193
4.3.7.2 Der erste Pretest.....	195
4.3.8 Nachbesserungsbedarf.....	197
4.3.9 Eingeführte Methoden zur Messung der Lesegeschwindigkeit.....	198
4.3.9.1 Ein „guter“ Lesbarkeitstest: Die Kriterien nach Paterson und Tinker.....	199
4.3.9.2 Die Messmethodik nach Tinker und Paterson.....	200
4.3.9.3 Modifikationen der Paterson-/Tinker-Methode.....	201
4.3.9.4 Ein weiterer Ansatz zur Messung der Lesegeschwindigkeit.....	203
4.3.10 Entscheidung für eine „Stolperwort“-Studie.....	203
4.4 Methode der Wahl: experimentelle Beobachtung	204
4.4.1 Das Experiment als empirisches Forschungsinstrument.....	206
4.4.2 Deskription, Exploration oder Hypothesenprüfung?.....	207
4.4.3 Labor- versus Feldbeobachtung.....	208
4.4.4 Das Online-Experiment als spezielle Form der Feldbeobachtung.....	210
4.4.4.1 Chancen und Potenziale von Online-Erhebungen.....	210
4.4.4.2 Grenzen und Risiken von Online-Erhebungen.....	211
4.5 Experimentaldesign	213
4.5.1 Projekt: Online-Studie.....	213
4.5.2 Wie misst man Dutzende von typografischen Konstellationen?.....	215

4.5.2.1	Eigentlich unmöglich: Einrichtung einer Kontrollgruppe.....	217
4.5.2.2	Lösungsansätze bisheriger Studien.....	218
4.5.2.3	Systematisierung der bisherigen Ansätze.....	220
4.5.2.4	Bewertung der Ansätze.....	222
4.5.2.5	Ein heikler Ausweg: Ein Proband – mehrere Testtexte.....	223
4.5.2.6	Zuteilungsverfahren im Experiment – eine Abwägung.....	224
4.5.2.7	Schwerwiegende Frage: Noch aussagekräftig?.....	226
4.6	Gemessene Faktoren und Attribute.....	227
4.6.1	Faktor Schriftart.....	228
4.6.1.1	Warum den Faktor Schriftart messen?.....	228
4.6.1.2	Anzahl und Ausgestaltung der gemessenen Attribute.....	229
4.6.2	Faktor Schriftgröße.....	229
4.6.2.1	Warum den Faktor Schriftgröße messen?.....	229
4.6.2.2	Anzahl und Ausgestaltung der gemessenen Attribute.....	230
4.6.3	Faktor Zeilenbreite.....	233
4.6.3.1	Warum den Faktor Zeilenbreite messen?.....	233
4.6.3.2	Anzahl und Ausgestaltung der gemessenen Attribute.....	233
4.6.4	Faktor Zeilenabstand.....	234
4.6.4.1	Warum den Faktor Zeilenabstand messen?.....	234
4.6.4.2	Anzahl und Ausgestaltung der gemessenen Attribute.....	234
4.6.5	Zusammenfassung der Faktoren und Attribute.....	236
4.6.6	Nicht gemessene Faktoren.....	237
4.7	Die Experimentaltexte.....	237
4.7.1.1	Die inhaltliche Gestaltung der Testtexte: Vorgaben.....	238
4.7.1.2	Die Länge der Texte.....	239
4.7.1.3	Anzahl der Texte.....	242
4.7.1.4	Anzahl und Mess-Sensibilität der Stolperwörter.....	243
4.7.1.5	Textinhalte.....	244
4.7.1.6	Die Positionierung der Stolperwörter in den Texten.....	246
4.7.1.7	Auswahl und Reihenfolge der Texte.....	247
4.7.1.8	Der zweite Pretest.....	247
4.7.2	Technische Ausgestaltung des Experiments.....	248
4.7.2.1	Verworfen Alternative: autonome Programmdateien.....	249
4.7.2.2	Verworfen Alternative: HTML/CSS.....	249
4.7.2.3	Die Alternative: Adobe Flash.....	250
4.7.2.4	Kriterien des Aufrufs in PHP.....	254
4.7.3	Die Programmierung.....	258
5	Allgemeine Ergebnisse.....	259
5.1	Eingesetzte statistische Methoden.....	259
5.2	Erhebungszeitraum.....	259
5.3	Akquirierung der Teilnehmenden.....	259
5.4	Rücklauf.....	260
5.5	Strukturdaten der Teilnehmenden.....	262
5.5.1	Altersstruktur.....	264
5.5.2	Bildschirmmaße.....	266
5.5.3	Betriebssysteme.....	267
5.5.4	Bildschirmarten.....	267
5.5.5	Ein erster Blick auf die gemessenen Zeiten.....	268
5.6	Notwendigkeit einer Datenbereinigung.....	269
5.6.1	Kriterium Mehrfachteilnahme.....	270
5.6.2	Kriterium Fehlklicks.....	271
5.6.3	Kriterium Lesezeit.....	274
5.6.4	Gleichverteilung der Konstellationen.....	276

5.6.5	Endauswahl der verwerteten Daten.....	277
5.6.6	Ergebnis der Datenbereinigung.....	277
5.6.7	Alternativen der Datenbereinigung?.....	279
5.7	Welche Konstellation hat „gewonnen“?.....	280
6	Die Ergebnisse im Detail.....	282
6.1	Die statistische Signifikanz.....	282
6.1.1	Begriff und Bedeutungsgehalt.....	282
6.1.2	Notierungsmethoden der statistischen Signifikanz.....	283
6.1.3	Die „heilige Kuh“ Signifikanz.....	283
6.2	Die Varianzanalyse.....	284
6.2.1	Die Resultate der Varianzanalyse.....	285
6.2.2	Bildschirmart und Lebensalter als „Überraschungsfaktoren“.....	287
6.2.3	Exkurs: Bildschirmarten.....	291
6.2.3.1	Röhrenbildschirme (CRT).....	291
6.2.3.2	Flachbildschirme.....	292
6.3	Der Dunnett T3-Test.....	293
6.3.1	Prinzip des Tests.....	293
6.3.2	Darstellung der Dunnett-Ergebnisse in dieser Arbeit.....	294
6.4	Schriftarten.....	296
6.4.1	Isoliertes Ergebnis.....	296
6.5	Schriftgrößen.....	297
6.5.1	Isoliertes Ergebnis.....	297
6.5.1.1	Ergebnisse am Flachbildschirm.....	298
6.5.1.2	Ergebnisse am Röhrenbildschirm.....	299
6.5.2	Zwischenfazit: Schriftgrößen.....	299
6.5.3	Schriftgröße und Schriftart im Zusammenwirken.....	300
6.5.3.1	Ergebnisse an an allen Monitorarten.....	300
6.5.3.2	Ergebnisse nach Bildschirmarten.....	301
6.6	Zeilenabstand.....	303
6.6.1	Isoliertes Ergebnis.....	303
6.6.2	Zeilenabstand und Schriftart im Zusammenwirken.....	304
6.6.3	Zeilenabstand, Schriftart und Bildschirm im Zusammenwirken.....	304
6.6.4	Zeilenabstand und Schriftgröße im Zusammenwirken.....	305
6.7	Zeilenbreite.....	306
6.7.1	Isoliertes Ergebnis.....	306
6.7.1.1	An allen Bildschirmen.....	306
6.7.1.2	Ergebnisse am Flachbildschirm.....	307
6.7.2	Zwischenfazit: Zeilenbreiten.....	308
6.7.3	Zeilenbreite und Schriftart im Zusammenwirken.....	308
6.7.4	Zeilenbreite und Schriftgröße im Zusammenwirken.....	309
6.7.4.1	Am Flachbildschirm.....	310
6.7.4.2	Am Röhrenbildschirm.....	311
6.7.5	Zeilenbreite und Zeilenabstand im Zusammenwirken.....	311
6.7.6	Zeilenbreite, Zeilenabstand und Schriftart im Zusammenwirken.....	312
6.7.6.1	Verdana.....	312
6.7.6.2	Times.....	313
6.7.6.3	Gesamtfazit: Zeilenbreiten.....	313
6.8	Die Schrift-Benotung.....	314
6.8.1	Bemerkung: Die Schulnote als ordinale Skala.....	315
6.8.2	Paarweiser Abgleich der Noten.....	316
6.8.3	Attraktivität = Lesbarkeit?.....	316

6.8.4 Die Noten im Detail.....	317
6.8.5 Die These von der „gefühlten Lesbarkeit“	318
6.8.6 Die These von der Gewohnheit.....	319
7 Zusammenfassung und Ausblick.....	320
8 Quellen.....	326
9 Abbildungsverzeichnis.....	342
10 Anhang.....	343
10.1 Messergebnisse der 90 Schriftkonstellationen.....	343
10.2 Digitale Dokumente.....	344
10.2.1 Das Experiment zum Ausprobieren im Web.....	344
10.2.2 Flash- und PHP-Dateien auf der CD-ROM.....	344
10.2.3 SQL-Dateien auf der CD-ROM.....	344
10.2.4 Diese Arbeit als PDF auf der CD-ROM.....	344
10.2.5 Bevölkerungsstatistik auf CD-ROM.....	344
Ehrenerklärung.....	345

Vorwort

Ein Übermaß an Zuwendung kann Zuneigung zerstören. Gleiches gilt, und zwar zweifellos in ausgeprägterem Maße, für vernachlässigte Beziehungs-Pflege. Heute, in den ersten November-Tagen des Jahres 2006, gehören diese beiden Weisheiten auf den Prüfstand.

Immerhin: Mein Verhältnis zu jenem faszinierenden und schönen Befassungsgegenstand der Schriftgestaltung im World Wide Web empfinde ich derzeit als angespannt, aber insgesamt intakt. Die Browser-Typografie, das Thema der vorliegenden Arbeit, hat meine Aufmerksamkeit zwar vehement eingefordert in den vergangenen Monaten. Meine Zuneigung zur Welt der Schriftzeichen hat in dieser Zeit zwar oft gelitten, erloschen ist sie zum Glück nicht.

Umso überfälliger ist der hilflose Dank an jene Person, der ich fortgesetzt typografische Hörner aufsetzte in den abgelaufenen Monaten. An diesen wunderbaren, liebevollen und tapferen Menschen, auf dessen Haupt-Kosten mein Flirt – oder sollte ich nicht sagen: mein fortgesetzter Seitensprung – mit der Web-Typografie ging: meine über alles geliebte Silke. Sie hat mir in diesem Jahr 2006 so viel Kraft verliehen, so viel Unterstützung geschenkt und so viel unfassbare Nachsicht und Geduld aufgebracht, dass es mich beschämt. Zumal ich weiß, dass der Rand des Erträglichen gelegentlich nicht nur erreicht, sondern überschritten war.

Mehr noch: Diese wunderschöne rothaarige Frau war sich nicht einmal zu schade, strategische Allianz zu bilden mit ihrem Widerpart: der vorliegenden Arbeit in ihrer Entstehung. Sie las kritisch gegen, was ich schrieb, diskutierte mal mehr, mal weniger abseitige Detailprobleme mit Langmut und verzichtete auf ihren geliebten Sommerurlaub im Süden. Wenn ich all diese Entbehrungen irgendwann nur ansatzweise vergelten kann, werde ich es versuchen.

Danken will ich aber auch und ausdrücklich meinem Kollegen Jürgen Dunker. Die vielen Anregungen, Hinweise und Ermutigungen, für die er sich Zeit nahm, haben das in dieser Arbeit vorgestellte Experiment in seiner Konzeption entscheidend mitgeprägt – und letztlich mit den Ausschlag dafür gegeben, dass ich den Angang zu dieser Dissertation überhaupt wagte.

Ein ganz besonderer Dank gebührt auch dem *Niedersächsischen Fußballverband (NFV)*, und hier vor allem meinem wichtigsten Ansprechpartner, dem stellvertretenden Verwaltungsdirektor Walter Burkhard. Ohne die Kooperationsbereitschaft des Verbands wäre die Akquisition vergleichbar vieler Versuchsteilnehmer, über deren Mitwirkung ich mich im Januar 2006 freuen durfte, zweifellos schwieriger, wenn nicht sogar unrealisierbar gewesen.

Ein großer Dank geht zudem an meine Kollegin Christine Fackiner, die eine für mich schwer nachvollziehbare „Win-Win-Situation“ konstruierte, indem sie diese Dissertation nicht nur auf Schwächen durchforstete, sondern sich nach getaner Arbeit sogar noch für die Aushändigung des Manuskripts bedankte. Dies tat mein Vater zwar nicht – aber auch er nahm sich einiger Seiten dieser Arbeit in der ihm eigenen Akribie an. Auch ihm danke ich für seine Hilfe.

Hannover, im November 2006

0 Einleitung

0.1 Das World Wide Web als journalistisches Medium

Jeder zweite Bundesbürger im Alter von mehr als 13 Jahren ging 2006 mindestens einmal monatlich online¹, im Schnitt sogar an 4,8 Tagen pro Woche². Ein mehr als beachtlicher Wert, datiert man die Geburtsstunde des Internets auch nur grob auf die beginnenden 1990er Jahre: „Selbst das Lieblingsmedium der Deutschen, das Fernsehen, benötigte einen längeren Zeitraum, bis es die Hälfte der Erwachsenen in (damals West-)Deutschland erreichte“³.

Das Internet ist in der abgelaufenen Dekade definitiv zum Massen- und Alltagsmedium aufgestiegen. Aus den weltweit gerade einmal rund 50 Webservern (also Computern, die potenziell Inhalte zum Online-Abruf bereithalten), die im Jgeahre 1993 festgestellt wurden⁴, waren Ende 2005 über 70 Millionen geworden⁵, 2006 dürften es bereits mehr als 80 Millionen sein – eine mehr als stattliche Menge, auch wenn nach den Erhebungen des renommierten britischen Web-Dienstleisters *Netcraft* davon auszugehen ist, dass nur die Hälfte dieser Server „aktiv“ ist, also wirklich relevante Inhalte beherbergt und nicht – beispielsweise in Gestalt einer jener berüchtigten „Web-Baustellen“ – mehr oder minder unbeachtet der Belegung mit „Content“ harrt⁶. Dass die Geschichte des Internets – beziehungsweise des Vorläufers „ARPANET“ – sogar bis in die 1960er Jahre und auf eine Initiative des US-Verteidigungsministeriums zurückgeht (und zwar mit dem Ziel des friedlichen Forschungsaustauschs und nicht, wie eine immer noch beliebte Legende besagt, um den Datentransfer im Falle eines Atomkriegs zu gewährleisten⁷), dass in den 1970er Jahren erste Universitäten und wissenschaftliche Institute das ursprünglich militärische Netz mitzunutzen begannen, und dass der Startschuss für die Popularisierung des Web fast folgerichtig von einer Forschungseinrichtung ausging, ist oft und gerne beschrieben worden⁸ und soll an dieser Stelle nicht zum wiederholten Male nachgezeichnet werden. Es geht in der vorliegenden Arbeit vorrangig um das Internet von heute: das Internet des Jahres 2006.

Und dazu ist folgendes bekannt. Nach wie vor der meistgenutzte Dienst im Internet ist E-Mail (78 % der Befragten der ARD-/ZDF-Online-Studie gaben an, die elektronische Post mindestens einmal wöchentlich zu nutzen). Wie viele der Online-Nutzer in Deutschland es beim Empfangen und Senden von E-Mails bewenden lassen, geht aus der Studie nicht hervor; Fakt

¹ Vgl. van Eimeren/Frees 2006, S. 404

² Vgl. van Eimeren/Frees 2006, S. 411

³ van Eimeren et al. 2004, S. 350

⁴ Vgl. Tolksdorf 2003, S. 1

⁵ Vgl. Zakon 2005

⁶ Vgl. Netcraft Ltd 2006

⁷ Vgl. Hafner/Lyon 2000, S. 10

⁸ Vgl. vor allem Abbate 2000, S. 43ff; vgl. ferner Hillstrom 2005, S. 5ff.; vgl. ferner Hafner/Lyon 2000, S. 11ff.; vgl. ferner Hauben/Hauben 1997, S. 13ff.; vgl. ferner Veen 2001, S. 14ff.; vgl. ferner Zehnder 1998, S. 28ff.

ist aber: 50 Prozent der User suchen laut Eigenaussage mindestens einmal pro Woche „zielgerichtet bestimmte Angebote“, 45 Prozent surfen „einfach so im Internet“, 35 Prozent betreiben Homebanking⁹. All diese Tätigkeiten aber sind praktisch unverrichtbar, ohne dass man sich jenes zweiten großen Internet-Dienstes neben E-Mail befleißigte, der begrifflich gerne mit dem „Internet“ gleichgesetzt wird¹⁰, aber tatsächlich eben nur ein Teil des „Netzes der Netze“ bildet: das World Wide Web, kurz Web oder auch „www“ – für Kreuzberger das „Glanzstück der Dienste im Netz“¹¹.

In selbigem existiert beispielsweise die „netZeitung“ (netzeitung.de) – ein, wie der Name bereits suggeriert, zumindest im redaktionellen Selbstverständnis journalistisch informierendes Portal mit Vollredaktions-Anspruch, das insofern bis heute ein einzigartiges Projekt darstellt im deutschsprachigen Web, als es ohne Pendant auf dem Markt der Printmedien existiert¹². Denn häufiger und zahlreicher unter den im weiteren Sinne journalistischen Angeboten mit „General Interest“-Ausrichtung sind da schon die sogenannten „Print-Ableger“ im Web, die aber samt und sonders eigenständige Redaktionen unterhalten und nur teilweise Beiträge des gedruckten „Mutterblatts“ übernehmen: „Spiegel online“ (spiegel.de) beispielsweise, „Focus online“ (focus.de) oder „FAZ.net“ (faz.net) – Angebote, für die sich inzwischen der Sammelbegriff der „Online-Zeitung“ etabliert hat, obgleich es sich bei den Mutterblättern teilweise auch um Magazine handelt¹³. Inhaltlich unterhalten werden diese Angebote größtenteils von Journalisten, die aufs Publizieren im Internet spezialisiert sind; längst ist der „Online-Journalismus“ auch Bestandteil hochschulgebundener und der betrieblichen Lehr- ausbildung¹⁴ sowie Gegenstand isolierter wissenschaftlicher Betrachtung¹⁵.

Nun muss ein Angebot nicht notwendigerweise auf eine Nachfrage reagieren. Gerade um die Jahrtausendwende galt vielfach die schiere Parole, man habe eben „im Internet zu sein“, das bloße Vorhandensein einer Präsenz galt als imagerelevant, der Inhalt war zunächst nachgeordnet. Doch journalistische Angebote im Internet allgemein, vermutet aber insbesondere im Web scheinen durchaus eine Nachfrage zu bedienen. Denn nahezu jeder Zweite begibt sich – laut Eigenaussage – online „häufig“ oder „gelegentlich“ auf die Suche nach „aktuellen Nachrichten“ (45 Prozent, und das ist der Spitzenwert), 38 Prozent rufen „aktuelle Regionalnachrichten und -informationen“ auf, 40 Prozent „Veranstaltungstipps und Freizeitinformationen“, 37 Prozent „aktuelle Serviceinformationen (Wetter, Verkehr)“. Schenkt man den erhobenen Daten Glauben, ist der sprichwörtlich ablenkbare oder intellektuell anspruchslosere Konsument im Web klar in der Minderheit: „Sportinformationen“ – diese verbuche ich hier

⁹ Vgl. van Eimeren/Frees 2006, S. 408

¹⁰ Vgl. Böhringer et al. 2003, S. 369;

¹¹ Kreuzberger 1997, S. 13

¹² Vgl. Maier 2004, S. 680

¹³ Vgl. Braun 2006, S. 69

¹⁴ Vgl. Hoofacker 2004; vgl. ferner Lackerbauer 2003; vgl. ferner Schmider 2003; vgl. ferner Alkan 2002; vgl. ferner Wieland/Spielkamp 2003; vgl. ferner Heijnk 2002; zu Standards der Ausbildung vgl. Dörmann/Pätzold 1998, S. 59ff.

¹⁵ Vgl. Meier 2002

mutig zunächst einmal unter der Rubrik „Unterhaltung“ – rufen nur 29 Prozent der Befragten „häufig“ oder „gelegentlich“ ab, „Unterhaltungsangebote“ (spezifiziert als „Comedy, Spiele“) werden gerade einmal in 7 Prozent aller Fälle auf den Schirm geholt¹⁶. Wie gesagt: Dies sind die Angaben von befragten Usern. Ob sich dabei wirklich alle Nutzer, die das Web vornehmlich aus Zerstreungs-Interesse bemühen, als solche zu erkennen gaben, muss spekulativ bleiben.

Auch die Zahlen der „Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern“ (IVW) mit Sitz in Berlin stützen die These, dass journalistische Dienstleistung im Web durchaus nachgefragt ist¹⁷. Seit 1997 kontrolliert das Institut zusätzlich zu Print-Titeln wie Tages- und Wochenzeitungen sowie Magazinen und Zeitschriften auch die Verbreitungen von Online-Angeboten; und die März-Daten 2006 (dieser Monat wird hier beispielhaft angeführt, da er außerhalb von Ferienzeiten lag und das Großereignis Fußball-Weltmeisterschaft noch keine dominierende Rolle in der Berichterstattung übernommen hatte), weist in der Sparte „redaktioneller Content“ beispielsweise für „bild.de“ 465 Millionen Seitenaufrufe auf (das entspricht 15 Millionen sogenannte „Page-Impressions“ pro Tag!). Und selbst, wenn man „Bild“ eher in der Sparte „Unterhaltung“ verortet – auch die Zugriffszahlen der Ableger seriöserer journalistischer Print-Titel sind stolz: „Spiegel online“ verzeichnete im März 2006 rund 333 Millionen Abrufe von Seiten mit „redaktionellem Content“, „stern.de“ 128 Millionen, „Focus online“ 122 Millionen Abrufe. Auch die Zahlen für die auf Sportberichterstattung fokussierten Angebote im Web – die sich hiermit feierlich als rehabilitiert betrachten dürfen als journalistische und eben nicht nur Unterhaltungs-Dienste – können sich sehen lassen: „Sport 1“ verzeichnete im März 2006 158 Millionen Zugriffe, „kicker online“ 87 Millionen.

Der Vollständigkeit halber sei darauf hingewiesen, dass die obige Darstellung auf einer Auswahl von Titeln beruht. Die ersten beiden Ränge (vor „bild.de“) belegen, nach aufgerufenen Seiten mit „redaktionellem Content“, bei der IVW das „T-Online-Contentangebot“ (733 Millionen Zugriffe) und AOL (468 Millionen). Ich tat und tue mich nach wie vor jedoch schwer damit, diese Angebote als originär „journalistische“ zu begreifen; allzu untergeordnet erscheint in diesen Angeboten das aktuelle Nachrichtenangebot, gleichsam erdrückt von „Service“-Angeboten des Betreibers, „Shopping“-Sektionen und anderen Zusatzofferten. Ohnehin ist davon auszugehen, dass diese beiden Angebote ihre hohen Zugriffsraten weniger einer gesteigerten Publikumsnachfrage als vielmehr vorrangig der Tatsache verdanken, dass sie den Kunden der Provider *T-Online* beziehungsweise eben AOL ungefragt als Standard-Startseiten zu Beginn einer jeden Online-Sitzung aufgezwungen werden (das lässt sich zwar ändern, wird aber mutmaßlich nicht von vielen Nutzern getan). Ohnehin sollen die oben genannten Zahlen nur als Orientierungsrahmen dienen. Fakt scheint danach zu sein: Journalistische Angebote im Web werden nachgefragt.

¹⁶ Vgl. van Eimeren/Frees 2006, S. 407

¹⁷ Vgl. Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern 2006

Und allen schlechten Nachreden zum Trotz, das Internet sei ein tendenziell „oberflächliches“, buntes, intellektuell kurzatmiges Medium, findet journalistische Vermittlung im Web nach wie vor ganz überwiegend in der vermeintlich schwerstverdaulichen Form statt: verschriftlicht¹⁸. Nun bürgt Textmenge noch längst nicht für inhaltliche Qualität, dies sei zugestanden. Dennoch: Abgesehen vom unvermeidlichen und meist aufgeregt-farbig-animierten Super-Size-Werbeposter am Seitenkopf strahlt etwa „Spiegel online“ eine sachliche, weil vor allem textdominierte Aura aus, von der sich selbst das Mutterblatt in weiten Teilen längst entfernt hat; ungebildete Beiträge sind in „Spiegel online“ durchaus „aufmacherfähig“; eine ungebildete „Titelstory“ dürfte es dagegen in der Printausgabe schon lange nicht mehr gegeben haben. Noch augenfälliger ist die Schriftdominanz bei „Focus online“: Der Web-„Ableser“ jenes Blattes, das in den 1990er Jahren unter anderem wegen seiner Farb- und Abbildungslastigkeit eine Menge Häme und Ablehnung erfahren musste¹⁹, gibt sich im Web ausgesprochen textlastig-konservativ. Von wegen „Fast-Food-Journalismus“: Viele Webangebote, auch „Focus Online“, weisen optisch nach wie vor eine viel größere Nähe zum Printjournalismus auf als zum Fernsehen. Diese Feststellung ist insofern nicht ganz unerheblich, als gerade der Print-„Focus“ insbesondere wegen seines vermeintlich exzessiven Gebrauchs von Informationsgrafiken gelegentlich in einen despektierlich konnotierten Ruch „gedruckten Fernsehens“ gerückt wurde und wird²⁰ – Videoclips jedoch, also fernsehverwandte Formate, sind unter „focus.de“ zwar abrufbar, stellen aber bis heute rein quantitativ einen absoluten Bruchteil des Gesamtinformationsangebots. Letzteres ist und bleibt: textlastig.

Quintessenz: Buchstaben, Zeichen, Schriftsymbole sind auch im Web kardinale Träger journalistischer Information, und sie werden es sicherlich noch einige Zeit bleiben. Und um diese Informationsträger geht es in der vorliegenden Arbeit: um Schriftgestaltung im World Wide Web. Um Web-Typografie.

0.2 Ziel und Aufbau dieser Arbeit

Kernziel der vorliegenden Arbeit ist die Untersuchung des Rezeptionsprozesses verschriftlichter Information im World Wide Web. Das zentrale Interesse gilt dabei der HTML-basierten Schriftanordnung, jener spezifischen Konstruktion von Bildschirmtypografie also, die unmittelbar – also ohne erforderliche Aktivierung von Zusatzprogrammen – von Webbrowsern geleistet wird und bis heute das Rückgrat der absolut überwiegenden Anzahl von Webseiten bildet. Von geringerem Interesse ist in der vorliegenden Arbeit mithin die typografische Ausgestaltung beispielsweise von online verfügbaren PDF-Dokumenten, die ohne den „Acrobat Reader“ oder verwandte Interpretationshilfen nicht zu öffnen sind. Genauso wenig interessiert auch die Schriftgestaltung von Webseiten, deren Betrachtung die Installation eines *Flas-*

¹⁸ Vgl. Müller von Blumencron 2004, S. 680

¹⁹ Vgl. beispielhaft Hacke 1994, S. 3; vgl. ferner Kornelius/Milz 1992, S. 4ff.

²⁰ Vgl. Rediske 1994, S. 11; vgl. ferner Bolesch 1995, S. 3

hPlayers voraussetzt. Es geht in dieser Arbeit also um die „originäre“ Browser-Typografie – und, insbesondere, um Art und Qualität von deren Rezeption am Monitor. Das Augenmerk gilt dabei schwerpunktmäßig Fließtexten, journalistischem „Brottext“; Akzidenztypografie wie die von Überschriften, Zwischentiteln oder auch textlichen Navigations-Buttons findet nur am Rande Beachtung. Soweit möglich, strebe ich dabei an, den Rezeptionsprozess in Webbrowsern nicht nur beschreibend und experimentell zu durchdringen, sondern diese Untersuchungen zusätzlich in praxisrelevante Handreichungen zu übersetzen und weiterzuentwickeln.

Dazu werde ich einleitend, nach Klärung einiger zentraler Begrifflichkeiten, die Genese der HTML-basierten Typografie in den nunmehr 16 Jahren nachzeichnen, die verstrichen sind, seit der erste Webbrowser *WorldWideWeb* in Bern das Licht der Welt erblickte. Hierbei wird deutlich werden, dass und warum die Browser-Typografie eine spezifische, vor allem aber einer eigenständigen Betrachtung würdige Teildisziplin der bildschirmbasierten Typografie bildet. Ein Überblick über die zeitgenössischen kognitionspsychologischen Erklärmodelle des menschlichen Rezeptions-, insbesondere jedoch des Leseprozesses leitet über zur Diskussion der zunächst allgemein gefassten Frage, ob und inwiefern Lesen am Bildschirm sich substantiell – qualitativ und im Lesetempo – unterscheidet vom hergebrachten Lesen von Papier.

Im nächsten Schritt geht es ans Detail. Es werden die zentralen typografischen Faktoren vorgestellt, die nach herrschender Meinung nicht nur das schlicht ästhetische visuelle Erscheinungsbild einer Schriftanordnung prägen, sondern auch und insbesondere die Lesbarkeit dieser Anordnung durch Rezipienten. Diese typografischen Faktoren werden in ihrer Bedeutsamkeit und Wirkmacht einzeln sowie in wechselseitiger Interaktion betrachtet; da der Schriftsatz für Browser, stärker als in der printbasierten Gestaltung, engen Wechselwirkungen mit den technischen Voraussetzungen und insbesondere den Beschränkungen und Unwägbarkeiten des Programmtyps „Browser“ unterliegt, wird auch auf diese Umsetzungsspezifika eingegangen. Zudem werden Empfehlungen und Vorgaben der einschlägigen Literatur zur lesegerechten Ausgestaltung typografischer Faktoren auf HTML-basierten Webseiten wiedergegeben und wechselseitig abgeglichen. Sofern empirische Befunde zur isolierten oder interagierenden Rezeption dieser typografischen Einzelfaktoren vorliegen, werden diese – nach meinen Recherchen erstmals in dieser ausführlichen Form – gleichfalls vorgestellt und zusammenfassend bewertet.

Den Abschluss dieser Arbeit bildet die Herleitung, Vorstellung und die Diskussion der Resultate eines von mir selbst konzipierten, programmierten und durchgeführten Online-Experiments, in dem ich dank der Mitwirkung von über 1400 Teilnehmenden Wirkung und Einfluss von vier zentralen typografischen Faktoren in verschiedensten wechselseitigen Kombinationen und Ausprägungen auf die Lesegeschwindigkeit in Webbrowsern überprüfte.

Methodisch bewegt sich diese Arbeit mithin auf zwei Ebenen. Während der einleitende theoretische Teil inklusive der Herleitung des Designs des von mir durchgeführten Experiments vorrangig auf der Grundlage einer kritisch-reflektierenden Quellenanalyse ruht, bildet der abschließende Ergebnisbericht meinen individuellen Beitrag zur bislang sehr von Studien aus dem angelsächsischen Raum dominierten empirischen Erkenntnislage zur Rezeption verschriftlichter Information in Webbrowsern.

0.3 Der Begriff der Typografie

Wer „die Typografie“ eines zeitgenössischen Druckwerks oder Bildschirmauftritts gutheißt oder kritisiert, nimmt dabei in aller Regel Bezug auf das Aussehen, auf die visuelle Qualität, auf die subjektiv erfahrene oder vermeintlich objektivierbare Güte der jeweiligen optischen Umsetzung. „Typografie“ wird heute begriffen als ästhetisches Produkt, als Ausfluss einer (gebrauchs-)künstlerischen Tätigkeit. Diese Begriffsbelegung ist zweifellos nicht falsch; jedoch ist sie mindestens genauso ahistorisch wie ungenau. Nicht zuletzt deshalb bedarf der Terminus „Typografie“ begrifflich wie inhaltlich einer feineren Ausdifferenzierung, um im weiteren Verlaufe dieser Arbeit als Untersuchungsgegenstand fassbar zu bleiben.

Etymologisch leitet sich der Terminus „Typografie“ (beziehungsweise „Typographie“ als veraltete Schreibweise, die in dieser Arbeit lediglich in direkten Zitaten nicht durch die modernere Form ersetzt wird) als Kunstwort aus der Kombination des griechischen „typos“ („Gepränge, Schlag“) mit „gráphein“ („schreiben“) her²¹. Bereits an dieser Stelle ist die Diskrepanz zur Eingangsfeststellung evident: Der Begriff „Typografie“ beschreibt ursprünglich nicht eine visuelle Kreation oder Erfahrung, sondern schlicht eine Tätigkeit. In einer sehr klassischen, heute arg technikorientiert anmutenden Definition ist Typografie denn auch die Gesamtheit der „Arbeitsschritte Stempelschnitt, Schriftguss, Satz, Einfärben und Abdruck“²² und als Setzung erkennbar aus Zeiten des Bleisatzes überkommen.

Ist Typografie mithin gar nicht ein primär visuelles Phänomen, sondern vielmehr eine Tätigkeit, eine laborative Verrichtung? Gulbins und Kahrmann zumindest laden den Begriff ebenfalls auf im Sinne einer Prozessbeschreibung und weniger als Produktdefinition, wenn sie schreiben, Typografie sei „das Arbeiten mit Schrift und freiem Raum“²³. Auch Köhler fokussiert tendenziell eher auf Tat denn Ergebnis, wenn er als Typografie „die künstlerische Gestaltung eines Druckwerks mittels Schrift, Bild, Linien und Flächen sowie Farben und Papier“²⁴ beschreibt. Rehe nähert sich dem Begriff ebenfalls weniger produktbezogen als vielmehr über die Beschreibung von Rolle und Aufgabe des Agierenden: des Typografen nämlich in einem Massenmedien-System. Auch er betont die künstlerisch-gestalterische Herausforderung

²¹ Vgl. Die Zeit: Das Lexikon. Mit dem besten aus der Zeit. In 20 Bänden. Bd. 15, S. 166

²² Wehde 2000, S. 3

²³ Gulbins/Kahrmann 1992, S. 3

²⁴ Köhler 2002, S. 6

als Kern der Tätigkeit, wenn er umreißt, dem Typografen obliege „die Auswahl der Schrift, der Schriftgröße und die typografische Anordnung der zu übermittelnden Nachricht“²⁵. Die Auswahl, nicht die drucktechnische oder auch bildschirmbasierte Realisierung des Konzepts! Spätestens in Rehes Beschreibung hat sich die Typografie also von der handwerklichen Arbeit an Druckerpresse oder Programm-Code zur gestalterischen Meta-Tätigkeit entwickelt. Der Typograf gibt vor, was der Drucker zu drucken, der Ingenieur zu programmieren hat. Der Typograf ist damit künstlerischer Leiter eines visuellen Vorhabens; die nur mehr Umsetzenden des Projekts aber sind in diesem Sinne keine Typografen mehr.

Immerhin ist nicht zu leugnen, dass in den eben beschriebenen, wiewohl eher prozess- als resultatorientierten Beschreibungen das intellektuelle Ziel typografischen Wirkens bereits merklich durchscheint: das gestaltete Endprodukt. Typografie – und damit ist die Anknüpfung gefunden zu den einleitenden Worten dieses Kapitels – ist eben nicht nur Tat, sondern auch Ziel. Es gilt also als wichtige Zwischen-Erkenntnis an dieser Stelle festzuhalten: Typografie ist in begrifflicher Hinsicht zweierlei, nämlich gleichermaßen „Gegenstand und (...) Tätigkeit“²⁶. Der Begriff der Typografie ist also in dieser Hinsicht doppelt besetzt²⁷. Doch ist er es nicht nur in dieser Hinsicht.

Denn ein weiterer Aspekt bedarf der Klärung; der nämlich, was im Detail überhaupt Befassungsgegenstand der Typografie sei. Wo etwa beispielsweise Rehes Beschreibung der Typografen-Tätigkeit („Auswahl der Schrift“) genau wie die von Gulbins und Kahrman („das Arbeiten mit Schrift“) sehr fixiert, ja verengt bleibt auf das Universum der Buchstaben und Satzzeichen, auf „die rein formale Gestaltung von Text“²⁸, wirkt der Typografie-Begriff andernorts weit umfassender. Willberg formuliert es überraschend: „Man sollte nicht sein ganzes Leben nur mit Buchstaben verbringen – es gibt ja auch noch die Typografie!“²⁹ In diesem erweiterten Verständnis geht Typografie also über die isolierte Befassung mit Zeichen, mit Schriftarten, -schnitten und Zeilenbreiten hinaus³⁰, bezieht insbesondere das Gestaltungselement „Bild“ mit ein³¹, jedoch zum Beispiel auch die Papier- und Formatwahl³².

Das Dilemma ist offenkundig: Von der künstlerischen Schriftwahl und -ausgestaltung im engeren Sinne bis hin zur Gesamtgestaltung eines Werkstücks inklusive Bestimmung von Bildinhalten und Papierbeschaffenheit reicht die Bedeutungsbreite des Begriffs „Typografie“. Mehr noch: Ist denn nicht auch ein Typograf, wer jene divergierenden Schriftzeichen-Sätze überhaupt erst künstlerisch schöpft, aus denen zu wählen ist im Zuge eines Gestaltungsprozesses? Schriftdesigner wie Adrian Frutiger (der Vater der Schriftart *Univers*), Max Miedinger

²⁵ Rehe 1981, S. 14

²⁶ Luidl 1989, S. 14

²⁷ Vgl. Bergner 1990, S. 12

²⁸ Köhler 2002, S. 7

²⁹ Willberg/Forssmann 2005, S. 1f.

³⁰ Vgl. Frutiger 2001

³¹ Vgl. Bayer 2003, S. 43f.

³² Vgl. Aicher 1989, S. 204ff.; vgl. ferner Köhler 2002, S. 6

(der Zeichner der *Helvetica*) oder Matthew Carter (er kreierte die *Verdana*), die sich der hohen Kunst des kleinteiligen Alphabet-Zeichnens widmen, befassen sich zweifellos mit anderen – genauer gesagt: detaillierteren – Aspekten der Informationsvisualisierung als Mediengestalterinnen und Mediengestalter, die werktäglich die Platzierung von Tageszeitungsartikeln im Regionalblatt vornehmen; und jene zweifellos kreativen Menschen tun dies wiederum in aller Regel nach gestalterischen Grund-Bestimmungen jenes Menschen, der im Range eines Art Directors das Gesamtaussehen des jeweiligen Mediums vorgibt und überwacht, ohne im Tagesgeschäft persönlich Hand anzulegen an jede Einzelseite. Typografen im herrschenden Sinne sind jedoch eindeutig all diese Menschen. Eine beachtliche Spannweite!

Hochuli hat in den 1980er Jahren aus dieser Erkenntnis heraus die heute weitgehend konsensuale begriffliche Unterteilung in die Bereiche „Makrotypografie“ und „Mikrotypografie“ vorgeschlagen³³, die – das sei betont – nunmehr eindeutig dem visuellen Wesen des gestalteten Produkts den Interessensvorrang einräumt gegenüber der vorgeschalteten „Tätigkeit“ Typografie. Dabei bezeichnet die Makrotypografie die Gesamtgestaltung eines Werks nicht nur hinsichtlich Schriftwahl und -ausgestaltung, sondern auch hinsichtlich Bildidentifikation, Platzierung und Proportionssetzung. Auf mikrotypografischer Ebene dagegen werden Schriftcharakter, Buchstabenform und -gegenform sowie die Abstände zwischen Buchstaben, Wörtern, Zeilen und anderen grafischen Elementen erfasst³⁴.

Eine sogar noch stärkere Feingliederung des Typografie-Begriffs schlägt Stöckl vor³⁵, die freilich zum Teil an Wehde orientiert scheint³⁶, ohne dass Stöckl seine Kollegin ausdrücklich als Inspirationsquelle benennt. Er regt allerdings eine Aufteilung typografischer Ressourcen in vier Bereiche an (Wehde benennt nur die ersten drei der unten aufgeführten), die hierarchisch aufeinander aufbauen – jede jeweils nächsthöhere Dimension vereinigt also in Stöckls Sinne die darunter rangierenden in sich:

- Mikrotypografie bezeichnet dabei als kleinste Einheit die Schriftzeichen als einzelne in Schriftart, Schriftgröße, Schriftschnitt und Schriftfarbe,
- Mesotypografie ist die erweiterte Gestaltung des Schriftbildes durch Einbeziehung der Dimensionen Zeichenabstand, Wortabstand, Zeilenabstand, Grauwert, Ausrichtung und Schriftmischung,
- Makrotypografie bezeichnet die fortgeschrittene Organisation von Text und Textteilen mithilfe von Variationen von Absätzen, Einrückungen, Initialen, typografischen Hervorhebungen, Orientierungspassagen (zum Beispiel Tabellen, Marginalien oder Fußnoten) sowie der Text-Bild-Montage (im Allgemeinen ist diese „Makrotypografie“ damit das, was gemeinhin als „Layout“ bezeichnet wird),
- Paratypografie – diesen Begriff führt Stöckl zusätzlich zu Wehde ein – bezieht sich schließ-

³³ Vgl. Hochuli 1987, S. 7

³⁴ Vgl. Kunz 1989, S. 94f.

³⁵ Vgl. Stöckl 2004, S. 22f.

³⁶ Vgl. Wehde 2000, S. 120ff.

lich auf die „Materialität der Dokumentgestaltung“, die beeinflussbar ist via Papierqualität und Signier- beziehungsweise Druckverfahren.

Ich werde in dieser Arbeit den Begriff „Typografie“ in einem eher engen Verständnis verwenden und inhaltlich diskutieren, das definitorisch im weitesten Sinne der „Mesotypografie“ im Sinne Stöckls entspricht. Gegenstand dieser Arbeit ist also das Schriftzeichen in seiner visuell-funktionalen Ausformung zu Textgesamten, jedoch nicht dessen räumliche Verortung und Eingliederung in einem Print- oder Bildschirm-Gestaltungsgesamt. Gegenstand der vorliegenden Arbeit ist damit dezidiert nicht die potenzielle Einbindung von Text in ein kreatives Ensemble aus Bildern, Menüleisten und anderen denkbaren Gestaltungs-Elementen. Untersucht wird aus demselben Grund beispielsweise auch nicht Effekt und Sinn der „Umrahmung“ von Textgesamten durch Weißraum³⁷, genauso wenig wie die Tauglichkeit und Wirkung von Spaltensatz im Web³⁸ sowie die Frage, ob Texte am Monitor über eine Bildschirm-länge hinaus reichen dürften³⁹ – es sind dies eben eindeutig makrotypografisch orientierte Fragestellungen, deren Behandlung den erstrebten Umfang dieser Arbeit eindeutig überdehnt hätten.

Der Vollständigkeit halber sei abschließend erwähnt, dass ich in meinen Recherchen verschiedentlich auf meist wohlwollende Referenzierungen einer auf den ersten Blick in der Tat ausgesprochen ausdifferenzierten – und wiederum produkt- und nicht prozessorientierten – Kategorisierung des Typografie-Begriffs stieß, die auf Willberg und Forssmann und deren Standardwerk mit dem anspruchsvollen Titel „Lesetypografie“ zurückgeht. Diese Kategorisierung sei hier zunächst im Original zitiert:

„– Orientierungstypografie muss auf dem Bahnhof oder im Fahrplan den richtigen Weg weisen,
– Werbetypografie soll den Blick einfangen, dazu sind alle Mittel recht,
– Designtypografie will erneuernd, fortschrittlich und risikobereit neue Wege beschreiten,
– Zeitungstypografie soll so schnell wie möglich zur Sache führen,
– Magazintypografie zum Schnuppern und Blättern verleiten,
– Dekorationstypografie soll schön sein, auf Lesbarkeit kommt es hier nicht an,
– Formulare sollen durchschaubar sein,
– und ob das Kleingedruckte in Vertragsformularen lesbar sein oder untergehen soll, mag offen bleiben“⁴⁰.

Als „Königsdisziplin“ stellen Willberg und Forssmann diesen acht typografischen „Formen“ (wie die Autoren sie nennen) die „Buchtypografie“ gegenüber. Bei allem launigen Duktus der Definition ist freilich festzuhalten, dass die Autoren in ihrer „Kategorisierung“ bemer-

³⁷ Vgl. beispielhaft Chaparro et al. 2004; vgl. ferner Bernard et al. 2000

³⁸ Vgl. beispielhaft Baker 2005

³⁹ Vgl. z.B. Puscher 2001, S. 42

⁴⁰ Willberg/Forssmann 2005, S. 14

kenswert unbesorgt zwischen Medien und Designzielen grenzwandeln (kennen Magazine etwa keine Dekorationstypografie? Sollen Formulare nicht gut lesbar sein?) und sich allein dadurch meines Erachtens auf einem theoretischen Niveau bewegen, das zumindest im Hinblick auf die Ziele dieser Arbeit kaum einen Erkenntnis-Fortschritt zu erbringen vermag. Jenseits der Feststellung, dass das Herz der Autoren für Bücher und deren Gestaltung schlägt, entfaltete dieses „Standardwerk“ für mich und mein hier beschriebenes Vorhaben keinerlei Nutzen.

0.4 Usability und Typografie

Ein Begriff hält die Welt der Software-Ersteller und Webdesigner in Atem: Usability. Bedienbarkeit also, Handhabbarkeit, Benutzbarkeit. Nun ist diese vage Bezeichnung eindeutig älter als die Jahrtausendwende, auch wenn Puscher das „Streben danach, Software einfacher benutzbar zu machen“⁴¹, just in seinem Werk aus dem Jahre 2000 „endlich da“ wähte⁴² – der Begriff „Usability“ war nachweislich bereits in den 1970er Jahren in der Hardware-Branche geläufige Vokabel⁴³. Dass sich das Augenmerk damals bei Marktführer *IBM* und seinen Konkurrenten eher auf klassisch ergonomische Fragen richtete wie beispielsweise das Tastatur-Design, ändert nichts an der Betagtheit des Begriffs. Und spätestens mit dem Aufkommen grafischer Benutzeroberflächen wie dem „Operating System“ (OS) der Firma *Apple Macintosh* in den frühen 1980er Jahren und später der Oberfläche *Windows* von *Microsoft* wandelte und erweiterte sich der Begriff⁴⁴, bezeichnete zunehmend die „Benutzbarkeit“ nicht mehr nur von Maus und Keyboard, der „Hardware“ also, sondern auch dessen, was auf dem Bildschirm geschah. Dass dies mehr als notwendig war (und scheinbar immer noch ist), stellte Faulkner noch im Jahre 2000 fest, als sie vermutete, „dass 75 Prozent heutiger Benutzer von Computern mit Maschinen kämpfen, die schwer zu beeinflussen sind und die sich scheinbar nicht die Funktionen entlocken lassen, die die Nutzer ihnen gerne abverlangen würden“⁴⁵.

Usability stellt also den Benutzer in den Mittelpunkt der Entwicklung von Benutzer-Oberflächen⁴⁶. Dies sollte eine Selbstverständlichkeit sein, ist es aber anscheinend viel zu selten – „vor allem wegen der Schlüsselrolle der Entwickler, die solche Systeme entwerfen, und die der Technologie im Designprozess zu viel Gewicht beimessen und zu wenig Aufmerksamkeit auf die ‚menschlichen Faktoren‘ verwenden“⁴⁷.

Dass bereits Mitte der 90er Jahre auch die Gestaltung von Webseiten verstärkt ins Blickfeld der Usability-Forschung geriet, überrascht kaum; schließlich ist es keineswegs ein abseitiger

⁴¹ George 1995, S. 1 (Übersetzung M.L.)

⁴² Vgl. Puscher 2001, S. 1ff.

⁴³ Vgl. George 1995, S. 7f.

⁴⁴ Vgl. Champeon/Fox 1999, S. 44f.

⁴⁵ Faulkner 2000, S. 5 (Übersetzung M.L.)

⁴⁶ Vgl. Lazar 2005, S. 3ff.

⁴⁷ Vgl. Chen 2002, S. 53 (Übersetzung M.L.)

Ansatz, Web-Angebote als kleine, in sich geschlossene Software-Lösungen oder „Interface Designs“⁴⁸ zu begreifen, die dem Nutzer gezielte Steuerungen abverlangen, die potenziell zu Zielen führen. Zwar betonen beispielsweise Spool et al., Software-Usability sei definitiv nicht mit der Usability von Websites vergleichbar, da im ersten Falle der Bedienungserfolg, im zweiten jedoch die Qualität des Seiteninhalts im Fokus des Nutzerinteresses stehe⁴⁹ – eine Mehrheitsmeinung aber scheint diese Unterstellung nicht zu sein. Fakt ist: Es entstand bereits Mitte der 1990er Jahre der Begriff der „Web-Usability“⁵⁰ oder auch der „Web-Ergonomie“⁵¹. Es ist zu untersuchen, welche Rolle in diesen Konzepten der Typografie zukommt oder zukommen kann als kodifiziertem Vehikel der Informationsvermittlung.

Nielsen formuliert seine Vorstellungen von Web-Usability eindeutig: „Ich glaube (...), dass das Hauptziel der Internet-Projekte darin liegen sollte, es dem Kunden so weit wie möglich zu erleichtern, nützliche Aufgaben in und mit diesem Medium zu realisieren“⁵². Mit dieser Setzung ist Nielsen bereits sehr nahe an der Basis-Definition der europäischen Norm EN ISO 9241: „Usability ist das Ausmaß, in dem bestimmte Benutzer in einem bestimmten Kontext ihre Ziele effektiv, effizient und zufrieden stellend erreichen“⁵³. Für Spool et al. gilt eine eher noch kompaktere Definition: „Je mehr eine Seite Menschen hilft, die Informationen zu finden, die sie suchen, desto benutzbarer (im Original eben: ‚usable‘, M.L.) ist sie“⁵⁴.

Alle drei Beschreibungen lassen aufhorchen. Denn: Ist hochgradig verwirklichte „Usability“ im obigen Sinne nicht – zumindest zu einem merklichen Anteil – schlicht die gleichsam „digitale“ Ausprägung dessen, was die Journalistik als „medienadäquate und zielgruppengerechte Informationsvermittlung“ begreift und in der Journalistenausbildung einfordert⁵⁵? Findet sich hier in anderem begrifflichem Gewand die berühmte „Leserorientierung“ wieder, die „Leserfreundlichkeit“, das gern postulierte aktive Eingehen auf „Leserwünsche“? Auf jene Rezipientenanforderungen also, die sich durchaus auch auf formale Gestaltungskriterien wie Angebotsstruktur und nicht zuletzt optische Phänomene beziehen können⁵⁶? Ist das Konzept der „Usability“ mithin auch, womöglich leicht modifiziert, anwendbar auf originär journalistisch ausgerichtete Angebote im Web?

Auch „journalistische Usability“ setzt, unterstellt man zunächst einmal die inhaltlich-begriffliche Verwandtschaft von „Usability“ und journalistischer „Vermittlungsadäquatheit“, Nutzer in den Stand, Ziele „effektiv, effizient und zufrieden stellend“ zu erreichen. Widerspruch? Im Prinzip wird niemand einen solchen geltend machen. Problematisch an der Normierung ist freilich, dass sie inhaltlich, bei näherer Betrachtung, auf dem Niveau einer Sonntagsrede ran-

⁴⁸ Vgl. Champeon/Fox 1999, S. 11ff.

⁴⁹ Vgl. Spool et al. 1999, S. 13f.

⁵⁰ Vgl. Lazar 2005, S. 3

⁵¹ Vgl. Balzert 2004, S. 11ff.

⁵² Nielsen 2000a, S. 11

⁵³ Vgl. EN ISO 9241-11

⁵⁴ Spool et al. 1999, S. 4 (Übersetzung M.L.)

⁵⁵ Vgl. Blöbaum 2000, S. 278f.

⁵⁶ Vgl. Rager et al. 1994, S. 58ff.

giert: Niemand wird schließlich das Avisierte in seinem Wesensgehalt in Frage stellen; wer würde schon eine schlecht zu bedienende Webseite zum Ziel erheben? Selbst nachweislich beratungsresistente Journalisten kaschieren ihr Festhalten an Überkommenen schließlich bevorzugt mit dem abschließenden Bescheid: „Der Leser will das so“.

Doch auf jede Sonntagsrede folgt ein Montag, an dem es eben nicht mehr nur ums vage Allgemeinziel geht, sondern um die Wege zu Erreichung desselben. Und hier argumentiert insbesondere Nielsen nun gar nicht mehr uneingeschränkt konsensfähig. Er verlangt nämlich (wie andere Usability-Verfechter auch⁵⁷) im Prinzip die völlige Unterordnung des optischen Designs unter die Ziele der Benutzbarkeit, erkennbar dem legendären Dogma des Bauhauses folgend, wonach die Form unbedingt der Funktion zu folgen habe⁵⁸: „Es gibt im Grunde genommen zwei Entwurfsansätze: das künstlerische Ideal (...) und das Ideal des Ingenieurs, der danach strebt, Problemlösungen für den Kunden zu schaffen. Dieses Buch steht unverrückbar auf der Seite der Ingenieursicht“⁵⁹. Dass insbesondere Nielsen sich mit dieser Philosophie nicht nur Freunde unter kreativen visuellen Gestaltern macht und gemacht hat, soll an dieser Stelle nicht ausufernd gezeigt werden. Jedoch schimmert der Wesenskern dieses Konflikts zwischen Gestaltern und Usability-„Päpsten“ („Usability experts are from Mars, graphic designers are from Venus“⁶⁰) meines Erachtens bereits erkennbar durch, wenn im Folgenden erörtert wird, welche Rolle der Typografie im Konzept der Web-Usability zukommt.

Im Sinne der „Usability“ – wie ich sie aus den obigen Herleitungen heraus interpretiere – hat Typografie auf Webseiten schlicht beizutragen zum Ziel, Benutzern das Rezipieren von Textaufgaben am Bildschirm im physischen Sinne eines möglichst raschen, körperlich wenig ermüdenden und nachhaltigen Aufnehmens zu gestatten. Schrift ist im Usability-Konzept also nur technokratischer Funktionsträger, als passiver Behälter und ausschließlich „sekundäres visuelles Double des Wortes“⁶¹; mehr nicht. Vor allem nicht Träger von Emotion, sofern dies die Lesbarkeit negativ beeinflusst, auch nicht Element der Innovation, der Aufmerksamkeits-erheischung oder schlicht Mittel der optischen Individualisierung.

Wohlgemerkt: Ich stelle damit nicht den Sinn von guter „Usability“ und ihrer Erforschung in Frage. Was ich allerdings anzweifle, ist der Rang von „Usability“ – in der erwähnten Begrifflichkeit – als praktisch alleiniger Messlatte der Qualität einer Webseite, als letztinstanzliches und praktisch ausschließliches Kriterium der Güte eines Angebots im Web, das im Zweifel konkurrierende Faktoren visueller und ästhetischer Natur zu disqualifizieren in der Lage sei. Insbesondere, da Usability als umfassendes Konzept kaum der Tatsache gerecht wird, dass Webseiten durchaus nicht durch die Bank das Ziel der gezielten Informations-Übergabe kennen – viele Angebote zielen eben auf optische Überraschung, Brechen von Konventionen,

⁵⁷ Vgl. z.B. Le Peuple/Scane 2003, S. 14ff.; vgl. auch Constantine/Lockwood 1999, S. 3ff.

⁵⁸ Vgl. Droste 1998, S. 52ff.

⁵⁹ Nielsen 2000a, S. 11

⁶⁰ Cloninger 2000

⁶¹ Bonsiepe 1996, S. 80

oder auch „nur“ den Transport eines Images. Doch auch für informationszentrierte, also nicht zuletzt journalistische Angebote im Web kann die schiere Benutzerfreundlichkeit meines Erachtens nicht singuläres Kriterium sein; vor allem, wenn kommerzieller Erfolg erstrebt wird. Denn Behauptung (und damit auch: Abgrenzung) gegenüber (kommerziellen) Mitbewerbern erfordert anerkanntermaßen auch die Kreation von Individualität im Sinne eines Corporate Images – eine Individualität, die unter anderem über optische, mithin eben auch typografische Individualisierung, mithilfe eines Corporate Designs also, herstellbar ist. Hier stößt das Usability-Konzept in der radikalen Lesart Nielsens eindeutig an Grenzen⁶².

Nun ließe sich durchaus argumentieren, dass zumindest der angesprochenen „Usability“-Definition der Norm EN ISO 9241 mittels des Postulats der „Zufrieden-Stellung“ des Nutzers durchaus eine Berücksichtigung der visuellen Ästhetik immanent sei. Dies scheint allerdings etwas kurz gegriffen, denn die Zufriedenheit des Nutzers wird in der Norm eindeutig in Bezug gesetzt zur „Aufgaben-Lösung“; darin auch einen ästhetisch-gestalterischen Anspruch zu erkennen, wirkt doch arg konstruiert⁶³. Vielversprechender erscheint da eher ein relativ neuer Ansatz, der die Usability um den Begriff der „Joy of Use“ erweitert.

Der Begriff „Joy of Use“ hat sich in dieser Form noch nicht als eindeutiger Terminus herausgebildet in der Diskussion; oft ist auch die Rede von beispielsweise „hedonic quality“, „feeling“, „pleasurability“ oder „emotional inference“⁶⁴. Doch welcher Begriff auch immer favorisiert wird, gemeinsam ist den Ansätzen, dass sie den Versuch unternehmen, die visuelle Emotionalität, die Attraktivität – in der eben Typografie eine wichtige Rolle spielt – teils als komplementären Aspekt von Usability zu begreifen, teilweise gar als integrales Element. Norman beispielsweise vertritt vehement den Ansatz, dass „affektive Signale“ den Gesamtprozess der Informationsverarbeitung kontext- und situationsbezogen mit bedingen: „Verwenden Sie ein ansprechendes Design, eines, das gut aussieht und – nun ja – sexy wirkt, und das Verhalten (des Nutzers, M.L.) dürfte eleganter, einfacher und besser geraten. Attraktives funktioniert besser“⁶⁵.

Eine ähnliche Stoßrichtung verfolgen Hallnäs und Redström. Sie differenzieren zwischen „use“ (rationaler Anwendung) und „presence“ (emotionaler Komponente) eines Bildschirmauftritts und erkennen eine derart starke Wechselwirkung zwischen beiden Aspekten, dass sie abschließend fordern: „Change focus from design for efficient use to design for meaningful presence“⁶⁶. Auch Hassenzahl et al. fordern ausdrücklich ein „erweitertes Usability-Konzept“, das die Attraktivität eines Angebots als Designziel integriert: „Eine Software dürfte als anspre-

⁶² Vgl. Champeon/Fox 1999, S. 78

⁶³ Vgl. u. a. Eibl 2003, S. 165ff.

⁶⁴ Vgl. Reeps 2004, S. 22ff.

⁶⁵ Norman 2002, S. 37 (Übersetzung M.L.)

⁶⁶ Hallnäs/Redström 2002, S. 111

chend empfunden werden, wenn es sowohl nutzerfreundlich als auch (gestalterisch) interessant ist – in der Konsequenz wird der Nutzer es genießen, diese Software zu verwenden“⁶⁷.

Auch Jordan versteht „pleasure“ als Wesenselement eines gelungenen Produktangebots (er bezieht sich dabei im Übrigen nicht ausschließlich auf das „Produkt“ Bildschirm-Angebot), formuliert allerdings eine eindeutige Hierarchie der Nutzerbedürfnisse, in denen „Funktionalität“ und „Benutzbarkeit“ („functionality“ und „usability“) als erstrebenswerte Produkteigenschaften der „Freude“ eindeutig übergeordnet sind. Anders als die oben zitierten Autoren betrachtet Jordan also das Affektiv-Emotionale nicht als integralen Bestandteil des „Usability“-Begriffs, wohl aber als essenziell ergänzenden Faktor eines ganzheitlichen Angebotsbegriffs⁶⁸.

Ich halte insbesondere den Ansatz Jordans für einen sehr tauglichen. Es erscheint mir letztlich doch etwas hoch gegriffen, Design, Attraktivität, Emotionalität als unabdingbar-substanziellen Bestandteil von „Usability“ zu begreifen; diese Aspekte jedoch als bedeutsame Wirkfaktoren komplementärer Natur zur „Benutzbarkeit“ zu begreifen und in die Entwicklungsprozesse insbesondere von Webseiten nicht nur potenziell, sondern apodiktisch einzubeziehen, ergibt Sinn. Auf diesen Aspekt wird im Folgenden noch häufiger zurückzukommen sein; die Diskussion des „Joy of Use“-Ansatzes allerdings soll hier nicht detaillierter weiterverfolgt werden (eine sehr gute Übersicht findet sich insbesondere bei Blythe et al.⁶⁹, einen sorgfältigen Überblick verschafft auch Reeps⁷⁰).

0.5 Typografie als Gegenstand der Journalistik

Das vorangegangene Kapitel hat deutlich werden lassen, dass mit Begriff und Konzept der „Usability“ im Bereich der bildschirmbasierten Informationsaufbereitung ein ebenso schlicht umrissener wie – vielleicht gerade aufgrund der Simplizität seiner inhaltlichen Aufladung – sehr umstrittener und gern diskutierter Befassungsgegenstand existiert. Es scheint immerhin, dass dieser Begriff der „Usability“ nicht nur als Postulat an die Praxis Tauglichkeit entfaltet, sondern als theoretisches Konzept auch einen sehr fruchtbaren Ausgangspunkt für wissenschaftliche Debatten bildet des Inhalts, mit welchen Mitteln die „Benutzbarkeit“ eines medialen Angebots optimierbar sei. Der „Joy of Use“-Ansatz, der sich in all seinen konzeptionellen Facetten dem „Usability“-Begriff von der gestalterisch-visuellen Position aus zu nähern, ihn teilweise zu begleiten, teilweise integrativ zu vervollkommen versucht, ist nur ein – im Rahmen dieser Arbeit jedoch bedeutsames – Beispiel für die Elastizität und interdisziplinäre Bearbeitbarkeit des Konzepts „Usability“.

⁶⁷ Hassenzahl et al. 2000, S. 202 (Übersetzung M.L.)

⁶⁸ Vgl. Jordan 2002, S. 6ff.

⁶⁹ Blythe et al. 2003

⁷⁰ Vgl. Reeps 2004

Ein der „Usability“ vergleichbares Konzept fehlt der Journalistik. Es existiert hier kein insbesondere gleichermaßen schlichtes wie inhaltlich annäherungsfähiges Konstrukt, das der „Usability“ vergleichbar wäre, also durchaus knapp im Postulat ausfallen dürfte, aber vor allem: als Beschreibung grundsätzlicher Konsens wäre. Zwar sind, wie gezeigt, durchaus inhaltliche Kongruenzen des bildschirmbasierten „Usability“-Konzepts feststellbar zu in der Journalistik gerne verwandten Begrifflichkeiten wie „Leserfreundlichkeit“, „Zielgruppenadäquatheit“ oder auch „Vermittlungsqualität“. Und, wohlgemerkt: Untätigkeit in Erforschung und Lehre rezipientengerechter Informationsaufbereitung unterstelle ich jener Wissenschaft, in der ich schließlich selbst ausgebildet wurde, in keiner Weise. Was ich jedoch unterstelle, ist, dass der Gesamtprozess der journalistischen „Vermittlung“ bislang keine konsensuale Beschreibung gefunden hat in der Journalistik. Es ist insbesondere ungeklärt, welche Determinanten in welcher Ausprägung überhaupt Qualität und Wirkung einer Informationsaufbereitung auf Kommunikator- und Rezipientenseite bedingen.

Für die vorliegende Arbeit ist dieses Defizit durchaus bedeutsam. Während nämlich im Usability-Konzept die gestalterische, also auch typografische Aufbereitung eines Angebots durchaus als integraler Bestandteil des Vermittlungsprozesses begriffen wird, ist die Rolle und Bedeutung der Informations-Visualisierung im Allgemeinen, der Makro- und Mikrotypografie im Besonderen innerhalb des Prozesses der journalistischen Vermittlung im Lehr- und Forschungskanon der wissenschaftlichen Journalistik bis heute nicht zufriedenstellend umrissen und beschrieben.

Diese Tatsache überrascht. Immerhin ist zumindest im tagesaktuellen Print-Journalismus das Geschäft der Produktgestaltung überwiegend längst in die Zuständigkeit des Akteurs „Journalist“ übergegangen, wenn auch eher im Bereich der makro- als der mikrotypografischen Realisierung; der Schriftsetzer als „Zwischeninstanz“ im Produktionsprozess von Redaktion zu Druckerei ist weitgehend obsolet geworden durch die Institution des gestaltenden „Produktionsredakteurs“⁷¹.

Nun mag das Layouten (also: das visuelle Verwirklichen) einer Tageszeitungs-Seite vielleicht noch durchgehen als im Wesen nicht originär journalistische Leistung, sondern als eher nachgeschaltete, quasi produktverwaltende Tätigkeit. Ich sehe dies keineswegs so, billige einer derartigen Haltung aber noch ein Restmaß an Stichhaltigkeit zu. Jedoch: Hat sich nicht mit der Darstellungsform „Informationsgrafik“ spätestens seit den frühen 1990er Jahren auch eine eindeutig journalistische Methode der Informationsaufbereitung etabliert, die Inhalt und optische Form im Optimalfalle nicht nur zur Deckung bringt, sondern sogar unauflöslich integriert⁷²? Und wer wollte die potenzielle nachrichtliche Eigenqualität jener nicht-verbale journalistischen Disziplin in Zweifel ziehen, die sich einer noch längeren Tradition erfreut als die Infografik: die fotografische Abbildung?

⁷¹ Vgl. Altmeppen et al. 2002, S. 353

⁷² Vgl. Jansen/Scharfe 1999; vgl. ferner Sprissler 1999; vgl. ferner Meyer 1997; vgl. ferner Liebig 1999

Es zeigt sich mithin zunächst, dass Visualisierung im journalistischen Vermittlungsprozess zumindest drei unterscheidungswürdige Ebenen kennt:

- Die mittelbare Visualisierung – wie ich sie bezeichnen will – ist dabei die Übersetzung originär sprachlicher Mitteilung in kodifizierte Zeichen – in Buchstaben und Satzzeichen also, die per se zunächst nicht mehr sind als optische Aufzeichnungs-Vehikel einer ursprünglich und auch nach der Visualisierung immer noch wesentlich sprachlichen Aufbereitung.
- Die Originär-Visualisierung hingegen – auch diese Bezeichnung schlage ich hier lediglich vor und gebe sie zur Diskussion frei – entfaltet ihre informationelle Eigenqualität nicht als Verbal-Übersetzung, sondern wesentlich aus sich selbst heraus. Zwar lassen sich Informationsgrafiken und Fotografien im Zweifel zumindest teilweise auch verbal beschreiben; diese Beschreibung wäre aber eine prozessual nachgeschaltete.
- Die organisierende Visualisierung schließlich begreife ich als die räumliche wechselseitige Anordnung von visualisierten Informations-Einheiten auf Papier oder Bildschirm; es ist mithin die bekannte Tätigkeit des „Layoutens“.

Diese eben vorgestellte Systematik zur Einbettung der Typografie in die Journalistik als Disziplin ist zweifellos diskutierenswert und vertiefungsfähig. Immerhin aber erlaubt sie vorläufig eine recht konsistente Einordnung des Betrachtungsgegenstands der vorliegenden Arbeit in den Prozess journalistischer Inhaltsvermittlung auf visualisierender Basis, die zumindest die wissenschaftliche Literatur der Journalistik in meinem Verständnis bislang nicht anzubieten vermag.

Weischenberg etwa widmet sich dem Problemfeld der Visualisierung journalistischer Leistung in seiner zweibändigen „Journalistik“ eindeutig kaum im Sinne eines Vermittlungsproblems denn höchstens als „technologischem Imperativ“, der im Gefolge digitalisierter Satz- und Drucktechnik insbesondere die Arbeitsplatzbeschreibung des journalistisch Tätigen beeinflusse und verändere⁷³. Sein Kapitel „Sprache und Verständlichkeit“⁷⁴ findet jedenfalls in beiden Bänden kein Pendant, in dem das Phänomen „Sprache“ auch in seiner typografischen Aufzeichnungsmethodik qualitativ diskutiert würde. Überhaupt befasst sich Weischenberg so gut wie gar nicht mit der Visualität journalistischer Ergebnisse, seien diese „nur“ fixierende Ausflüsse einer originär verbalen Vermittlungsleistung (wie im Textsatz) oder aber bildliche Informationen von eindeutig optischer wie nachrichtlicher Eigenqualität (wie eben beispielsweise Fotografien oder Informationsgrafiken).

Auch Neverla et al. lassen in der Themenwahl ihrer umfangreichen Antologie „Grundlagentexte zur Journalistik“⁷⁵ höchstens über Altmeppen et al. die technischen Umwälzungen im Redaktionsalltag⁷⁶ anklingen, die das Berufsbild des Journalisten in wachsendem Maße mit

⁷³ Vgl. Weischenberg 1995, S. 15ff.

⁷⁴ Vgl. Weischenberg 1995, S. 179ff.

⁷⁵ Vgl. Neverla et al. 2002

⁷⁶ Vgl. Altmeppen et al. 2002, S. 353

technischen, unter anderem satztechnischen Aufgaben konfrontiere. Eine wie auch immer geartete Eigenqualität der Visualisierung journalistischer Tätigkeit wird jedoch in dem umfangreichen Werk nicht thematisiert.

Durchaus Erwähnung und Anerkennung findet der „optische Journalismus“ hingegen in jener inzwischen gut anderthalb Jahrzehnte alten, jedoch seitdem nur sehr phasenweise fortgeführten wissenschaftlichen Diskussion, ob und wie die Güte journalistischer Ergebnisse zu bewerten, empirisch zu operationalisieren und zu messen sei: in der Debatte um die journalistische Qualität nämlich. Rager regte 1994 an, die Güte der Vermittlung (neben der Aktualität, der Richtigkeit und der Relevanz⁷⁷ eines Beitrags) als eine von eben vier Qualitäts-„Dimensionen“ journalistischer Resultate zu begreifen (später diskutierte Rager auch die Dimension „Ethik“ als womöglich ergänzenden fünften Aspekt⁷⁸). „Vermittlung“ definiert Rager dabei so: „Gegenseitige Bezüge herstellen zwischen KommunikatorInnen und Publikum, (...) Kommunikations-Beziehungen aufnehmen“⁷⁹ – und dies sei, wohlgermerkt unter anderem und nicht ausschließlich, verwirklicht über „redaktionelle Vorgaben über Design und Illustration“. Unter anderem Wyss griff diesen Gedanken in seinen Erörterungen 2002 auf, ohne ihn jedoch substantiell weiterzuentwickeln⁸⁰. Wiewohl aber die Qualitäts-Debatte immer noch nicht entscheidend über den Status der Faktoren-Beschreibung und deren wechselseitiger Gewichtung hinausgelangt ist⁸¹, ist Ragers Einbeziehung der visuellen Komponente in die journalismuswissenschaftliche Diskussion bislang zumindest ohne aktiven Widerspruch geblieben.

Zudem ist festzustellen, dass die Visualisierung von Information in der weniger wissenschaftlichen, sondern eher praxisorientierten Literatur neueren Erscheinungsdatums mehrheitlich Anerkennung gefunden hat als integraler Aspekt der journalistischen Tätigkeit. Claudia Mast räumt dem Kapitel „Layout von Presseprodukten“ immerhin 20 Seiten ein im von ihr herausgegebenen „ABC des Journalismus“. Sie begründet dies einleitend damit, dass eine journalistisch gelungene Arbeit nicht zur Geltung komme, wenn Rezipienten „von einer unverständlichen, verwirrenden oder gar langweiligen Präsentation abgeschreckt werden. Journalisten müssen sich auch um die grafische, akustische und visuelle Aufbereitung ihrer Produkte kümmern“⁸². Zwar bleibt die Autorin die Konkretisierung schuldig, ob diese Aufbereitungs-Tätigkeit vom Journalisten nur zu kontrollieren oder selbst durchzuführen sei – im einschlägigen Kapitel finden sich allerdings auch von Mast selbst zusammengetragene, insgesamt eher kurz gehaltene Ausführungen zur Makro- und Mikrotypografie, was die Vermutung nahelegt, dass die Autorin zumindest die Gestaltung von Print-Produkten als Teil des Berufsbildes

⁷⁷ Vgl. Rager 1994, S. 190f.

⁷⁸ Vgl. Rager 2000

⁷⁹ Rager 1994, S. 202

⁸⁰ Vgl. Wyss 2002, S. 139ff.

⁸¹ Vgl. Hassemer/Rager 2006, S. 19

⁸² Vgl. Mast 2004, S. 359

zeitgenössischer Journalisten begreift. Die visuelle, gar die programmiertechnische Gestaltung von Online-Publikationen findet hingegen keine spezifische Berücksichtigung bei Mast.

Auch Pürer et al. reservieren im Ratgeber „Praktischer Journalismus“ 23 Seiten für den Beitrag „Zeitungsgestaltung“ von Norbert Küpper⁸³, in dem dieser ebenfalls sehr spartanisch das Thema Mikrotypografie streift. Schneider und Raue befassen sich im „Handbuch des Journalismus“ im Hauptkapitel „Wie man den Leser gewinnt“ über knapp sechs Seiten hinweg mit dem Thema Layout – mit der Begründung, es sei durchaus journalistische Aufgabe, „diejenigen optischen (...) Mittel anzuwenden, die geeignet sind, den Blätterer zu einem Text überhaupt hinzuführen: Foto und Bildunterschrift, Überschrift und Vorspann – und das Layout überhaupt, die schlüssige und gefällige Gliederung der Seite“⁸⁴. Die Visualisierung hat im Verständnis von Schneider und Raue also vorrangig vorgeschaltete, dienende, lockende Funktion, ist Mittel zum letztlichen Zweck: Rezipienten zur Text-Lektüre zu verleiten. Aber immerhin: Die Optik einer journalistischen Leistung wird auch hier verstanden als Teil der journalistischen Tätigkeit.

Walther von La Roche erwähnt ebenfalls, Redakteure hätten neben der Frage: „Stimmt der Inhalt?“ auch stets die Frage: „Stimmt die Form?“ zu beantworten⁸⁵. Da er jedoch nur „die allen Medien gemeinsamen Probleme“⁸⁶ besprechen wolle, beschränke er sich in dieser Hinsicht aufs Notwendige – er tut es auf einer halben Seite, die dem Thema „Print-Layout“ gewidmet ist. Auch in der Publikation „Die tägliche Neu-Erscheinung“⁸⁷ räumen die Herausgebenden in der Diskussion der Entwicklungspotenziale des Mediums „Tageszeitung“ Hingst ausreichend Raum ein, die makro- und typografischen Aspekte einer zukunftsgerichteten Produktgestaltung zu erörtern⁸⁸. Auch Ruß-Mohl beschäftigt sich in seinem „Hand- und Lehrbuch“ mit dem „Präsentieren“ als „sichtbare Spitze‘ des Eisbergs Journalismus. (...) Dabei kommt es nicht nur auf Ästhetik und optische Auflockerung an, sondern auch auf Orientierungshilfe und Leseanreize“⁸⁹.

Zumindest zur Gestaltung von Zeitungen und Zeitschriften liegen inzwischen auch mehrere monothematische Publikationen vor. Zu nennen ist hier insbesondere das nicht mehr ganz tauforsche, aber immer noch höchst praxistaugliche Grundlagenwerk „Contemporary Newspaper Design“⁹⁰ des weltweit wohl renommiertesten US-amerikanischen Zeitungsdesigners (und gelernten Journalisten⁹¹) Mario Garcia. Auch Brielmaier und Wolf nähern sich dem Thema unbestritten von journalistischem Standpunkt aus⁹², genau wie Blum und Bu-

⁸³ Vgl. Küpper 2004, S. 257ff.

⁸⁴ Schneider/Raue 1996, S. 152

⁸⁵ Vgl. von La Roche 2003, S. 16

⁸⁶ von La Roche 2003, S. 18

⁸⁷ Vgl. Rager et al. 1992

⁸⁸ Vgl. Hingst 1992, S. 95ff.

⁸⁹ Ruß-Mohl 2003, S. 170

⁹⁰ Vgl. Garcia 1987

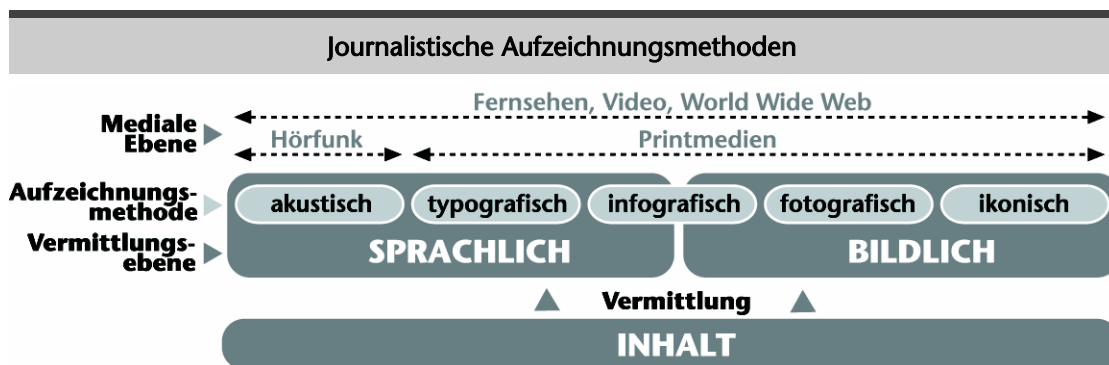
⁹¹ Vgl. Garcia 1998, S. 16

⁹² Vgl. Brielmaier/Wolf 2000

cher⁹³, beide Autorenpaare jedoch eher aus makrotypografischer Perspektive. Die eher „fachfremde“ Sicht des Grafikers formuliert Rehe in seinem Klassiker „Typografie und Design für Zeitungen“⁹⁴. Die meisten dieser Werke sind freilich eindeutig als Praxishandreichungen konzipiert und leisten entsprechend keine konsistente Integration des Gestaltungsbegriffs in die wissenschaftliche Journalistik. Eher finden sich Bedeutsamkeitseinordnungen von schlichter Qualität wie dieser: „Auch der beste Text kommt beim Leser besser an, wenn er gut präsentiert wird“⁹⁵. Ein derartiger Allgemeinplatz freilich legt eine wissenschaftliche Befassung mit dem Phänomen „Gestaltung“ im Journalismus allerhöchstens nahe; eine qualifizierte, gar wissenschaftliche Einordnung der Visualität von Information als Wirkmacht im Zusammenspiel und Bedeutsamkeits-Abgleich insbesondere mit der verbalen Aufbereitung einer Mitteilung leisten Behauptungen wie diese aber eindeutig nicht.

Insgesamt ist festzuhalten: Die Rolle der „Optik“ als Wirkfaktor in der journalistischen Vermittlung scheint zumindest in der praxisorientierteren Literatur inzwischen außer Frage zu stehen. Auch in den Lehrkanons einschlägiger, mehr oder minder praxisorientierter journalismuswissenschaftlicher Studiengänge scheint die Mediengestaltung als Vermittlungsfaktor zunehmend Berücksichtigung zu finden – was diese wiederum als Betrachtungsgegenstand der Journalistik qualifiziert, definiert man mit Weischenberg behelfsweise zur Journalistik, „was in den Studiengängen gleichen Namens gelehrt wird“⁹⁶.

In genau diesem Verständnis also wird in der vorliegenden Arbeit die Mikrotypografie journalistischer Produkte als Teilaspekt einer wissenschaftlich relevanten Vermittlungsproblematik begriffen. Als Aufzeichnungsmethode im Begriffs- und Handlungs-Komplex der journalistischen Vermittlung, die sich auf einer originär sprachlichen Ebene wie folgt einordnen lässt:



Eine Systematik von Aufzeichnungsmethoden im journalistischen Vermittlungsprozess.

ABBILDUNG 0.1

Die Existenz eines substanziellen journalistischen Aussagepotenzials („Inhalt“) wird dabei apodiktisch unterstellt; dieses Aussagegesamt ist aufzubereiten, zu vermitteln, kurz: rezipier-

⁹³ Vgl. Blum/Bucher 1998

⁹⁴ Vgl. Rehe 1986

⁹⁵ Meissner 1992, Buchrücken

⁹⁶ Weischenberg 1992, S. 22

bar zu gestalten. Diese Vermittlungsleistung kann dabei verwirklicht werden unter Einsatz und wechselseitiger Kombination einer oder mehrerer von insgesamt fünf journalistischen „Aufzeichnungsmethoden“, die ich identifiziere: der akustischen nämlich, der typografischen, der infografischen, der fotografischen (womit auch die Bewegt-Fotografie, der Film also, bezeichnet ist) oder der ikonischen (womit zeichnerische Umsetzungen beispielsweise in Form eines Logos oder einer Vignette, jedoch auch karikaturistische Darstellungen gemeint sind).

Akustische und typografische Aufzeichnung verstehe ich dabei als Methoden, die einer übergeordneten „sprachlichen Vermittlungsebene“ zuzuordnen sind, fotografische und ikonische Aufzeichnungen schlage ich dagegen der „bildlichen Vermittlungsebene“ zu. Die Infografik schließlich als journalistische Aufbereitungsgattung mit in aller Regel sowohl typografischem wie bildlichem Anteil und damit aufbereitungstechnischem Mischcharakter begreife ich entsprechend als sowohl der sprachlichen als auch der bildlichen Vermittlungsebene zugehörig.

Wie Abbildung 0.1 zusätzlich vermittelt, kennt das Genre „Hörfunk“ lediglich eine Aufzeichnungsmethode: die akustische nämlich – und damit wiederum die einzige der fünf identifizierten, deren Einsatz klassischen Printmedien versagt bleibt. Fernsehen und Internet dagegen stehen als Plattformen potenziell zur Verfügung für Aufzeichnungen aller beschriebenen Gattungen.

Ob und inwiefern die Qualität einer Aufzeichnung dabei Wirkung und Einfluss entfaltet auf den Vermittlungsprozess als Ganzes, ist die entscheidende, die spannende Frage. Entwertet etwa eine unterbelichtete Fotografie deren Informationswirkung und damit den Vermittlungserfolg insgesamt? Wird schlecht eingesprochenen Hörfunk-Beiträgen – unabhängig vom Inhalt – mindere Seriosität zugebilligt von Rezipientenseite? Und schließlich und nicht zuletzt: Wirkt Typografie qualitativ zurück auf den Text, den sie übersetzt? Konterkariert eine schlecht gewählte Zeilenbreite den Vermittlungserfolg? Befördert umgekehrt die Auswahl einer angemessenen Schriftart Erinnerung und Nachhaltigkeit einer Mitteilung?

Es sind Fragen wie diese, deren Beantwortung sich die vorliegende Arbeit zu nähern sucht.

0.6 Bildschirm- und Webbrowser-Typografie

Die vorliegende Arbeit befasst sich mit einer höchstens 16 Jahre jungen Teildisziplin der Mesotypografie: der Typografie in Webbrowsern – oder kurz: der „Browser-Typografie“ oder eben auch „Web-Typografie“. Dabei bildet die Browser-Typografie eine Teildisziplin der wiederum auch nicht auf eine überlange Tradition referenzierenden Bildschirm- oder Monitor-Typografie. Dieses Kapitel hat vorrangig die Klärung eben dieser Begrifflichkeiten und ihre inhaltliche wechselseitige Abgrenzung zum Gegenstand.

0.6.1 Typografie auf dem Monitor

Ist Schriftsatz am Bildschirm, ist Lesen am Monitor überhaupt eine Verrichtung von so neuer Qualität, dass Bildschirm-Typografie den Status einer neuen Disziplin zu beanspruchen hätte? Was qualifiziert Typografie am Monitor im Allgemeinen, Typografie in Webbrowsern im Besonderen überhaupt als Gegenstand des Forschungsinteresses? Zunächst seien zwei – meiner Ansicht nach – entscheidende Aspekte benannt, die Typografie am Bildschirm qualitativ vom klassischen Lesen von Papier abgrenzen:

- die Entstofflichung des Schriftzeichens am Monitor,
- die damit verbundene Identitäts-Entknüpfung von Speicher- und Darstellungsmedium: die „Flüchtigkeit“ der Bildschirm-Typografie.

0.6.1.1 Entstofflichung der Typografie: das Medium Monitor

Zunächst: Gestaltung, mithin auch Typografie am Bildschirm, ist immateriell. Typografie am Bildschirm hat sich von der passiven Lichtabhängigkeit gelöst, die bis in das vergangene Jahrhundert hinein Bedingung im Grunde jeder visuellen Kommunikation war.

Jede optische Kreation, mithin auch jede Schrift bedurfte bis vor gut 100 Jahren der Beleuchtung, einer sie bestrahlenden Lichtquelle, um rezipierbar zu sein. Ohne bestrahlendes Licht, sei es das natürliche der Sonne oder künstliches, war und ist jedes Buch unlesbar, jede bedruckte Textile unsichtbar, keine Steingravur rezipierbar⁹⁷. Das Aufkommen der Leuchtreklame um die Jahrhundertwende der wilhelminischen Epoche relativierte dieses Apodiktum zunächst teilweise, doch die eigentliche Innovation war das Medium, das sich als „Selbststrahler“ vollends von der Notwendigkeit löste, beleuchtet zu werden, um zu funktionieren: der Monitor. Ab der Mitte des 20. Jahrhunderts eroberten die Fernseh-Bildschirme als erstes jener Massenmedium überhaupt die Haushalte, das für den Konsum weder Tageslicht noch Kerzenschein oder Lampenbestrahlung voraussetzte.

Mehr noch: Bildschirme generieren nicht nur Licht, sind also sogar selbst potenzielle Beleuchtungsquelle – es ist das durch sie generierte Licht an sich, das Farbeindrücke im Auge des Betrachters überhaupt erst initiiert. Das Licht ist nicht mehr nur Voraussetzung, sondern nunmehr essenzieller Träger visueller Information.

Damit hat der Bildschirm einen Wandel in der visuellen Kommunikation begründet, der nicht gering geschätzt werden kann: Typografie am Monitor hat ihre Eigen-Stofflichkeit eingebüßt. Wo einst das Vorhandensein von Materien wie Stein, Papyrus, Wollgewebe Bedingung für das Übermitteln von Zeichen war, die entweder in besagte Materie graviert, geritzt oder mithilfe von Pigmentstoffen aufgetragen wurden, ist der Buchstabe „A“ am Monitor

⁹⁷ Vgl. Gibson 1982, S. 49ff.

eine Licht-Schimäre. Unbenommen: Auch dieses „A“ verdankt seine Existenz ganz ursprünglich etwas Stofflichem, einer Platine etwa, die digitale Daten prozessiert, und einer Lochmaske, die das digital Erhaltene auf den Monitor projiziert – das „A“ selbst aber ist, anders als sein mit Druckerschwärze generiertes Ebenbild in der Tageszeitung, eben definitiv nicht-stofflich und damit in höherem Maße vergänglich selbst als das papierne Pendant. Es gibt weder Gravurmaterial noch Bedruckstoff. Text am Bildschirm ist also, wie es Karow entwaffnend einfach zuspitzt, „elektronisch gedruckt, d.h. flüchtig und nicht zum Anfassen“⁹⁸. Wer schreibt, der bleibt? Nicht mehr am Bildschirm.

Das muss zunächst kein Nachteil sein. Im Gegenteil: Sieht man einmal von den Höhlenwänden der Steinzeit ab als „Leinwand“ der Malerei unserer Vorfahren, hatte und hat jeder der klassischen „Bedruckstoffe“ ebenso seinen Preis wie die Pigmente, die darauf aufgebracht werden. Der schwer bestimmbare monetäre Gegenwert schriftlicher oder bildlicher Information erhöht sich also in Nicht-Bildschirmmedien stets um den Preis für Bedruckstoff und Druckstoff. Ist es nicht immer wieder der „Papierpreis“, der Verlagen als Rechtfertigung für die Erhöhung der monatlichen Abonnements-Abbuchungen dient?

Der Monitor als moderner „Bedruckstoff“ dagegen ist stetig „wiederbeschreibbar“. Dazu ist – anders als beim Papier – nicht einmal ein kostenintensiver Recycling-Vorgang vonnöten. Jeder Computerbildschirm vermag im Verlaufe seiner Lebensdauer solche Unmengen an Inhalten zu projizieren, dass deren Ausdrücke wohl abertausende von Seiten füllen würden. Umgelegt auf seinen Anschaffungspreis dürfte sich ein vielgenutzter Bildschirm daher rasch amortisieren – rascher vermutlich, als viele Web-Nutzer vermuten würden, bevor sie die potenziellen Farbpatronen- und Papierkosten für den unterstellten Ausdruck aller von ihnen innerhalb der Lebensdauer ihres Bildschirms betrachteten Webseiten mit dem Kaufpreis des Monitors an sich abgleichen würden.

Die negativen Auswirkungen der tendenziellen Entstofflichung von Typografie durch digitale, selbststrahlende Medien sind freilich gleichfalls abschätzbar, wenn auch nicht monetär exakt quantifizierbar wie im vorigen Vergleich. Denn wo Schrift auf Stein, Papier oder anderem Textil eine unmittelbare Symbiose eingeht mit dem Speichermedium, dem Bedruck- oder Gravurmaterial also, ist Bildschirmtypografie eben mittelbarer Natur: Speichermedium, interpretierendes Programm und Ausgabemedium sind separiert. Was nichts anderes bedeutet als dies: Wenn eines der drei Medien nicht zur Verfügung steht, sind die jeweils anderen beiden wertlos.

Bildschirmtypografie knüpft also ihre Existenz nicht nur an das Überleben des Speichermediums (zum Beispiel der Computer-„Festplatte“), sondern auch an das Fortbestehen des die Zeichen dekodierenden Programms (zum Beispiel *Microsoft Word*) und an die Existenz eines Bildschirms, der die Informationen des jeweiligen Programms in Schriftzeichen zu übersetzen

⁹⁸ Karow 1992, S. 41

vermag. Diese Mittelbarkeit von Bildschirm-Typografie zeitigt bereits erste Konsequenzen: Die 5,25-Zoll-Diskette, auf der viele Menschen noch Anfang der 90er Jahre des abgelaufenen Jahrhunderts durchaus Konservierwürdiges speicherten, findet schon seit Jahren kein passendes Laufwerk mehr in zeitgenössischen Rechnern; die kleinere Variante, die 3,5-Zoll-Diskette, sieht einem absehbar vergleichbaren Schicksal entgegen. Ich selbst hatte Anfang des Jahres 2005 im Zuge der Recherchen zu dieser Arbeit meinen Kommilitonen Gerrit Faust kontaktiert mit der Bitte, mir seine 1991 entstandene – aus meiner Erinnerung heraus hervorragende – Studienarbeit zum Thema „Zeitungstypografie“ zur Verfügung zu stellen. Gerrit Faust sagte mir seine Unterstützung telefonisch zu – um mir wenige Tage später (per E-Mail!) zur Kenntnis zu bringen, dass er selbst über keine Papierausgabe der Arbeit mehr verfüge, und die Rekonstruktion der Datei, aus der der Ausdruck des Werks generiert worden sei, Probleme mit sich bringe; diese Datei nämlich habe er an einem nicht mehr ohne weiteres verfügbaren Computer der Firma „Schneider“ in einem heute nicht mehr ohne weiteres zugänglichen Textverarbeitungsprogramm verfasst.

0.6.1.2 Bildschirmmedien: Monitor-Konsum oder Ausdruck?

Das Problem ist evident. Archäologen des 22. Jahrhunderts werden ihre absehbare Mühe und Not damit haben, digitale Dokumente der absolvierten Jahrtausendwende zu sichten; womöglich wird sich eine ganze Armada an Wissenschaftlern ausschließlich mit der Frage beschäftigen, wie eventuell konservierten Speichermedien wie Festplatte, CD-ROM oder Diskette die darauf enthaltenen Informationen zu entlocken seien. Eine Rekonstruktionsaufgabe wird dies sein, angesichts derer das „Zusammen-Puzzeln“ der erhaltenen Fetzen der Qumram-Rollen nahezu primitiv wirken wird. Denn: Wer hebt schon Webseiten auf?

0.6.1.3 Die Flüchtigkeit des www oder: Wer hebt schon Webseiten auf?

An dieser Stelle kommt ein entscheidender Aspekt ins Spiel: Wird ein digitales Dokument vornehmlich erstellt mit dem Ziel, es nach Fertigstellung auf Papier zu drucken, es also zu „verstofflichen“, und in dieser Form an den Rezipienten zu bringen? Ist also die digitale Datei nur Mittel zum Endzweck: ein „analoges“ Produkt zu erzeugen, einen „Ausdruck“? Die vorliegende Arbeit zumindest wurde durchaus mit Schwerpunkt auf dieser Prämisse verfasst – nicht zuletzt aufgrund der Vorgaben der Promotionsordnung, an die diese Zeilen gebunden sind.

Angebote im World Wide Web, kaum minder aber auch ein Großteil von auf CD-ROM gebrannten bildschirmbasierten Inhalten, gehorchen jedoch anderen Gesetzen. Diese nämlich unterstellen in weit höherem, wenn nicht ausschließlichem Maße eine Rezeption der vorgehaltenen Inhalte am Monitor– was insbesondere bei „interaktiv“ konstruierten Inhalten auch

gar nicht anders denkbar ist. Wer auf einer „Startseite“ – sei es im Web oder auf einer CD-ROM – allein zwischen vier anzahlig gleichgewichtig bestückten Kategorien wählen kann, davon letztlich eine anklickt und im Gefolge fünf weitere Subkategorien im Angebot hat, von denen er wiederum nur eine aktiviert, hat de facto bereits 19 Seiten unbeachtet, damit aber auch potenziell unkonserviert gelassen. Auf CD-ROMs ist dies grundsätzlich zunächst kein Problem: Denn die 19 unbeachteten Seiten sind jederzeit reproduzierbar, zumindest so lange ein Betriebssystem verwendet wird, das die Informationen zu lesen imstande ist.

Anders im Web. Web-Angebote existieren eine Zeitlang, werden dann „relauncht“, erneuert, vermeintlich oder tatsächlich verbessert. Der jeweils vorangegangene Status Quo jedoch, der ersetzt wurde, verschwindet in aller Regel, ohne umfassend und vor allem: nicht-digital erfasst und archiviert zu sein.

Dieser Aspekt der Konservierbarkeit gewinnt tatsächlich besondere Bedeutung in Bezug auf Webseiten. So sind die papiernen Druckausgaben des Nachrichtenmagazins „Der Spiegel“ nicht nur im Hausarchiv des Hamburger Verlags mit etwas Mühe durchaus vollständig einsehbar, sondern auch in gut sortierten Hochschulbibliotheken. Aber ist sichergestellt, dass die frühen Ausgaben von „Spiegel online“, dem Web-Angebot des Magazins also, wiederholbar konserviert sind?

„Am 25. Oktober 1994 erschien die erste Ausgabe von Spiegel-Online im <http://www.spiegel.de>. Damit ist es das erste Nachrichtenmagazin im Internet, genau einen Tag älter als die Online-Version des 'TIME'-Magazins. (...) Das Online-Magazin wurde von den Spiegel-Redakteuren Uly Foerster und Gerd Meißner ins Leben gerufen. (...) Am Anfang arbeiteten zwei Teilzeitkräfte für das Online-Angebot. Schon 1996 wurde Aktualität beim Spiegel groß geschrieben. So heißt es in einem Memo von Uly Foerster: 'Es muss immer aktuell sein. Ich bestehe auf wöchentlichen Updates. Mindestens!'⁹⁹ Wer also 1996 die URL www.spiegel.de aufrief, bekam nicht zuletzt aus diesem Aktualitätspostulat heraus etwas anderes dargeboten als im Jahre 2006. Inhaltlich wie optisch. Und es ist fraglich, wie lange beispielsweise noch die Domain <http://www.spiegel.de/static/spon1996/online/ruek.html> abrufbar sein wird, jene gefühlige Reminiszenz an die Pioniertage des Online-Ablegers des „Hamburger Nachrichtenmagazins“. Wird sie gelöscht, ist der „Spiegel online“ der ersten Jahre verloren – beziehungsweise: er existiert dann höchstens noch auf den Festplatten ausgewählter Betreiber der frühen Jahre. Allgemein zugänglich aber wird er – im Gegensatz zu den „Spiegel“-Sammelbänden in einschlägigen Bibliotheken – absehbar nicht mehr sein. Ein Webserver, der abgeschaltet wird, ist für die Gemeinschaft der Online-Nutzer praktisch verloren, selbst, wenn noch irgendwo eine Offline-Sicherungskopie hinterläge.

In diesem Sinne ist das Web als Medium sogar noch flüchtiger als eben CD-ROM-basierte Lernsoftware: Denn die CD-ROM als Datenträger erfährt gemeinhin rein quantitativ eine

⁹⁹ O.A. 2004

weit größere Verbreitung, als dies für Webinhalte der Fall wäre; diese liegen in aller Regel auf einem einzigen Webserver, einer singulären Festplatte. Zumindest kommerziell vertriebene CD-ROMs erfahren eben durch den Vertrieb eine quasi dezentrale Archivierung, die den meisten Webseiten versagt bleibt.

0.6.1.4 Von der Bildschirm- zur Web-Typografie

Typografie im Web gewinnt ihre besondere Qualität und Einzigartigkeit also durch ihr meisteiliges Verharren im Zustand der Monitor-Virtualität. Zweifellos werden viele Webseiten auch ausgedruckt; die Mehrheit der Seiten jedoch, insbesondere jene dynamischen, die sich potenziell minütlich in Inhalt und Erscheinungsbild ändern können, bleiben unmaterialisiert: Shaikh fand 2004 in einer Befragung unter 330 US-Nutzern (221 Frauen, 109 Männer), dass diese insbesondere Nachrichten zu 68 Prozent ausschließlich am Bildschirm lesen und in der Regel keinen Ausdruck davon anfertigen – lediglich 12 Prozent der Befragten erklärten, sie würden Nachrichten online meistens erst grob überfliegen, um sie dann auszudrucken und genauer zu lesen¹⁰⁰.

Die Buchstaben am Bildschirm stehen damit definitiv nicht mehr im vornehmlichen Range einer Druckvorschau; die Lettern sind hier nicht nur vorläufig virtuell wie in einem druckfertigen *Word*-Dokument, das der papiernen Materialisierung harrt – die Virtualität ist Normalzustand in der Web-Typografie, das eigentliche „Gestaltungsziel“¹⁰¹. Diese Eigenschaft freilich hebt die Typografie von Webseiten allein immer noch nicht in den Rang eines betrachtungswürdigen Gegenstands. Was Webseiten wirklich einzigartig und zum würdigen Gegenstand einer eigenständigen Untersuchung im Rahmen der vorliegenden Arbeit geraten lässt, ist ihr ureigener, immaterieller Bedruckstoff, eine Programmsorte besonderer Art: der Webbrowser.

Es ist daher diese Programmgattung „Browser“, die die „Bildschirm-Typografie“ als übergeordneten Befassungsgegenstand auf das zentrale Betrachtungsphänomen der vorliegenden Arbeit hin spezifiziert. Web-Typografie ist, im Sinne der vorliegenden Arbeit, damit jener Teilbereich der Bildschirm-Typografie, der sich mit der visuellen Verschriftlichung von Information in Webbrowsern auseinandersetzt. Dass – und warum – diese Teildisziplin der Bildschirm-Typografie durchaus einer gesonderten Betrachtung und Diskussion bedarf, wird im Folgenden näher erläutert.

0.6.2 Das Konzept des Webbrowsers

Es gilt zu differenzieren: Der Grundgedanke, am eigenen Computer erzeugte Daten via Telefonkabel an andere Rechner zu übertragen – oder fremde Daten auf demselben Wege auf

¹⁰⁰ Vgl. Shaikh 2004

¹⁰¹ Vgl. u.a. Bayer 2003, S. 12

die eigene Festplatte zu bringen –, reicht in der Tat bis in die sechziger Jahre des vergangenen 20. Jahrhunderts zurück. In diesem Sinne ist das „Internet“ als Idee und Konzept so neu zweifellos nicht. Doch das Konzept der telefonisch korrespondierenden Rechner allein ist nicht ausreichender Erklärungsfaktor für den zeitgenössischen Erfolg des Mediums Internet und insbesondere des World Wide Web: Die Art und Weise, wie die empfangenen Daten auf dem eigenen Rechner visuell reproduziert werden, ist nicht minder von Bedeutung. Und hier tritt Idee und Konzept des Browsers hinzu, jener Programmsorte, ohne die der beeindruckende Aufstieg des Internets vom akademischen Diskussionsmedium zum Massenmedium heutiger Tage wohl in dieser Form nicht stattgefunden hätte¹⁰².

Web-Browser (auch: Webbrowser, oder kurz: Browser, von englisch to browse = stöbern), deren typografische Ausgestaltungspotenziale und -besonderheiten den Gegenstand dieser Arbeit bilden, sind – meistens kostenlos abgegebene – Clients, spezielle Computerprogramme also, zur Darstellung von Webseiten¹⁰³. Niederst und Freedman fassen es entwerfend einfach: „Wäre das Internet ein Fernsehsender, dann wäre der Browser das Fernsehgerät“¹⁰⁴. Von diesen Browsern gibt es einige, wenn auch längst nicht so viele wie Hersteller von Fernsehapparaten; die gebräuchlichsten und daher wohl bekanntesten sind der *Internet Explorer* der US-amerikanischen Firma *Microsoft*, der *Navigator* der Firma *Netscape* und das Projekt *Mozilla*, aus dem der zunehmend populäre Browser *Firefox* hervorgegangen ist.

Worin lag das neue, das innovative Element des Webbrowsers? Zunächst: Praktisch alle Browser halten eine mehr oder minder elaborierte grafische Benutzer-Oberfläche (englisch „graphic user interface“ – abgekürzt GUI¹⁰⁵) aus Icons, Menüelementen und Tasten vor, wo bis in die späten achtziger Jahre nur weiße Buchstaben auf meist schwarzen oder grünem Grund sichtbar waren. Browser erlauben das Navigieren zwischen verschiedenen Webseiten per Mausklick, wo ehemals kryptische, zeilenweise Befehlseingaben notwendig waren, auf die folgend sich – bestenfalls – Kaskaden von weiteren Buchstaben-Zeilen über den Bildschirm ergossen. Kurz: Browser waren für das Internet das, was das Betriebssystem *Windows* mit seiner Maus-Steuerung für den DOS-geplagten, über die Tastatur navigierenden Normalnutzer am *IBM-PC* war¹⁰⁶: eine bahnbrechende Arbeitserleichterung, der entscheidende Schritt weg von der Insider-Semantik des Eingabebefehls hin zur intuitiven visuellen Navigation – und damit ein entscheidender Schritt zur Massentauglichkeit.

Doch die bloße Schöpfung des Programmtyps „Webbrowser“ allein ist immer noch nicht Erklärung genug für den atemberaubenden Aufstieg des Internets. Denn was fast noch wichtiger ist: Das Anfertigen von betrachtbaren Dokumenten für die Webbrowser, von Webseiten also, wurde im Konzept des Webbrowsers denkbar einfach – um Längen einfacher jedenfalls

¹⁰² Vgl. Balzert 2004, S. 6

¹⁰³ Vgl. Balzert 2004, S. 6

¹⁰⁴ Niederst/Freedman 1996, S. 6

¹⁰⁵ Vgl. Bonsiepe 1996, S. 52

¹⁰⁶ Vgl. Champeon/Fox 1999, S. 23f.

als das Programmieren von Bildschirminhalten auf der Basis komplexer Sprachen wie C, Pascal oder Fortran. Wer das Internet nicht nur konsumieren, sondern auch selbst inhaltlich um eigene Beiträge bereichern wollte, erhielt nun ein potentes Instrumentarium dafür an die Hand. Denn im Vergleich mit den komplexen Programmiersprachen der späten 1980er, frühen 1990er Jahre nahm und nimmt sich die „Sprache des Web“ kinderleicht aus: HTML.

0.6.2.1 HTML

Den ersten Webbrowser – er hörte auf den klangvollen Namen *WorldWideWeb*, später benannten die Schöpfer ihn um in *Nexus* – entwickelte 1990 der Engländer Tim Berners-Lee gemeinsam mit seinem Partner Mike Sendall am Europäischen Labor für Partikelphysik in Bern (CERN)¹⁰⁷. Dieser Browser war noch rein textbasiert, interpretierte aber bereits die Markierungen von HTML, der „Hypertext Markup Language“, die ebenfalls Berners-Lee entwickelt hatte¹⁰⁸. Dass er mit HTML eine Art „Weltsprache“ geschaffen hatte – zumindest die Saat gesät hatte zu einer solchen – mag Berners-Lee seinerzeit gar nicht bewusst gewesen sein. Doch er tat es.

Denn nach wie vor basiert die ganz überwiegende Anzahl der Webseiten, die heute im World Wide Web aufrufbar sind, im Grundsatz auf der Markierungs- oder auch Auszeichnungssprache¹⁰⁹ HTML. Entgegen häufiger Behauptungen¹¹⁰ ist HTML keine Seitenbeschreibungssprache (wie etwa PostScript, das sich auf die Ausgabe auf Papier oder anderen Bedruckstoff bezieht) und eben auch keine Programmiersprache¹¹¹ (wie C++ oder Pascal). In diesem Sinne werden HTML-Seiten auch nicht programmiert, sondern geschrieben¹¹².

Es schmälert den Ruhm Berners-Lees gewiss nicht, wenn an dieser Stelle darauf hingewiesen wird, dass HTML im Grundkonzept keine grundstürzende Neuschöpfung war, sondern eher ein gezielt „entschlackter“ und damit einfacher handhabbarer „Dialekt“ der „Urmutter“ aller allgemeinen Auszeichnungssprachen: der „Generalized Markup Language“ (GML), die nach zahlreichen Vorstudien in den frühen siebziger Jahren des 20. Jahrhunderts vom amerikanischen *IBM*-Mitarbeiter Charles Goldfarb¹¹³ und seinen Kollegen Edward Mosher und Ray Lorie entwickelt, später erweitert und unter der Bezeichnung „Standard Generalizes Markup Language“ (SGML) Mitte der achtziger Jahre in den Rang der internationalen Norm ISO 8879 erhoben wurde¹¹⁴.

¹⁰⁷ Vgl. Zehnder 1998, S. 71

¹⁰⁸ Vgl. Abbate 2000, S. 214f.

¹⁰⁹ Vgl. Kübler/Struppek 1996, S. 1

¹¹⁰ Vgl. beispielsweise Kent 1999, S. 178

¹¹¹ Vgl. Lamprecht 2002, S. 57

¹¹² Vgl. Lubkowitz 2005, S. 125

¹¹³ Vgl. Kobert 2004, S. 36

¹¹⁴ Vgl. Veen 2001, S. 20f.

Was zeichnet eine Auszeichnungssprache wie SGML im Allgemeinen, HTML im Besonderen aus? Laborenz nennt zunächst als Grundprinzip einer Auszeichnungssprache, dass sie ein jedes Dokument so beschreibt, „dass es auf vielen verschiedenen Ausgabegeräten gemäß seiner Bedeutung wiedergegeben werden kann“¹¹⁵. Das bedeutet: Der Code wird im Modell der Auszeichnungs- oder Markierungssprache bestenfalls auf seinen schieren Inhalt, darüber hinaus höchstens auf seine elementarsten strukturellen Attribute reduziert – und damit der Bindung an Betriebssysteme, Satzprogramme und vorhandene Schriftarten nicht gänzlich, aber sehr weitgehend entledigt.

Die Potenziale dieser Grund-Idee sind reich. Denn auf Grundlage des „Markierungskonzepts“ konnten und können Texte aus der „Gefangenschaft“ des einen Textverarbeitungsprogramms erlöst werden, in dem sie originär erstellt wurden und in dem – wie beispielsweise auch im Programm *Word* – Textinhalt und typografische Darstellung in aller Regel eng verwoben sind. Die Reduktion auf grobe Markierungsanweisungen erlaubt es also, Texte unterschiedlichster Herkunft, erstellt auf den unterschiedlichsten Rechnern und in verschiedenen Textverarbeitungsprogrammen, auf Grundlage einiger Basal-Auszeichnungen, zu sammeln, zu archivieren und vor allem systemunabhängig zugreifbar zu gestalten¹¹⁶ – Grundlage und Voraussetzung nicht zuletzt für den Aufbau und die Pflege von Archiven und Datenbanken. Und letztlich ist das World Wide Web nichts anderes als dies: Eine riesige Datenbank, auf die jeder zugreifen kann, der einen Rechner besitzt, einen Telefonanschluss – und eben einen HTML interpretierenden Browser als Programm auf dem Rechner. Dieser Programmtyp „Browser“ aber war von Berners-Lee von Beginn an als „Open-Source“-Projekt konzipiert¹¹⁷ – das heißt: Berners-Lee legte die Konstruktionsprinzipien seines Browsers offen, anstatt dafür eine Lizenz anzumelden, bot seinen *Nexus* zudem zum kostenlosen Download an. Im Prinzip konnte und kann damit jeder Befähigte „surfen“ oder gar seinen eigenen Browser entwickeln, und vor allem: die Teilnahme am World Wide Web war, zumindest was die Software-Seite angeht, potenziell mit keinerlei Kosten verbunden.

0.6.2.2 Das Prinzip der logischen Auszeichnung

Seine Einfachheit – und damit die erstrebte Programm- und Plattformunabhängigkeit – verdankte die ursprüngliche HTML-Version insbesondere dem Prinzip der „logischen Textauszeichnung“, das Berners-Lee konsequent anwandte¹¹⁸. Grundidee der logischen Auszeichnung ist, dass Textkomponenten, wenn überhaupt, gemäß ihrer inhaltlichen Charakteristik, ihrer „semantischen Bedeutung“¹¹⁹ also, definiert werden, jedoch eindeutig nicht bezüglich

¹¹⁵ Laborenz 2005, S. 20

¹¹⁶ Vgl. Zehnder 1998, S. 17f.

¹¹⁷ Vgl. Hillstrom 2005, S. 109

¹¹⁸ Laborenz 2005, S. 20

¹¹⁹ Maurice/Rex 2006, S. 15

ihrer typografischen Physis¹²⁰. Eine Anweisung in einer Auszeichnungssprache lautet also – sinngemäß, nicht in der tatsächlichen Auszeichnung – kurz und knapp: „Die folgende Textpassage ist eine Überschrift“ – und eben nicht: „Die folgenden Wörter sollen erscheinen in der Schriftart ‚Arial‘, fetter Schnitt, 36 Punkt groß, Zeilenabstand 38 Punkt, zentriert“. Ein Basiskonzept von enormer Tragweite: Hier werden Inhalt und formale visuelle Ausgestaltung sauber voneinander getrennt¹²¹, das Dokument auf seinen Wesenskern „eingedampft“.

Freilich war mit dieser Grundkonzeption von HTML als Auszeichnungssprache bereits der Grundstein gelegt für ein bis heute latentes Dilemma: Denn in dem Maße, in dem das Internet wuchs und damit zwangsläufig auch kommerzielle Interessen weckte, wuchs auch die Nachfrage der Seitenproduzenten und ihrer Auftraggeber nach Potenzialen, ansprechende und vor allem: auf allen Browsern äquivalente Optik auf den Bildschirm zu bringen, das visuelle Erscheinungsbild ihrer Angebote also kontrollieren zu können. Wie gezeigt, prallten hier zwei Kulturen aufeinander: Das Konzept der Auszeichnungssprache, nach dem Passagen lediglich grob „charakterisiert“ werden (beispielsweise als „Überschrift erster Ordnung“), ist eben per se inkompatibel mit den Ansprüchen von Designern auf größtmögliche Kontrolle über die Seitenoptik¹²², was eben eben den Bedarf nach verlässlicher Typografie nicht zuletzt einschließt¹²³. Es ist daher fast ein Euphemismus zu behaupten, an Gestaltung habe Berners-Lee „wohl höchstens am Rande“¹²⁴ gedacht. Wahrer ist wohl dies: „Bei der Entwicklung von HTML spielten (...) ästhetische Überlegungen keine Rolle“¹²⁵. Dass HTML „typografie-fremd“¹²⁶ geriet, war Absicht, nicht Versehen.

Ein Blick auf den übersichtlichen Befehlssatz der ersten HTML-Version untermauert diese These. Wiewohl zu diesem Zeitpunkt bereits einige Jahre alt, trägt die HTML-Urversion von Berners-Lee das offizielle Veröffentlichungsdatum 3.11.1992 – und kaum eine moderne Website wäre mit den Auszeichnungs-Instrumenten dieses Urstandards verwirklichtbar gewesen: HTML erlaubte in seiner ersten Variante ausschließlich die Darstellung von Text. Dieser war markierbar als Überschrift (Berners-Lee sah sechs Ordnungen davon vor), ferner konnten Passagen beschrieben werden als Liste oder als Absatz. Im wesentlichen war’s das.

0.6.2.3 Der „Sündenfall“: physische Textauszeichnung

Längst war Anfang der 90er Jahre Tim Berners-Lees *Nexus* nicht mehr der einzige Webbrowser. Großer Beliebtheit – wenn auch nach wie vor in vergleichsweise geschlossenen akademi-

¹²⁰ Vgl. Schäffer 2001, S. 204

¹²¹ Vgl. Herrmann 2005, S. 229

¹²² Vgl. Peck 2000, S. 139

¹²³ Vgl. Weinman 2000, S. 254

¹²⁴ Vgl. Lubkowitz 2005, S. 50

¹²⁵ Lamprecht 2002, S. 37

¹²⁶ Vgl. Bollwage 2001, S. 103

schen Zirkeln – erfreute sich vor allem der Browser *Mosaic*¹²⁷ (NCSA = „National Center for Supercomputing Applications“, Sitz in Illinois, USA¹²⁸), den der Student Marc Andreessen entwickelt hatte¹²⁹. Dieser Marc Andreessen sollte es sein, der nicht allein, doch vehement mit seiner 1993 gegründeten Firma *Netscape*¹³⁰ bis in die späten 1990er Jahre energisch vor allem die optische Weiterentwicklung der Browser-Umgebung vorantrieb, die Elemente „Bild“ und „Farbe“ zu einem integralen Bestandteil des Web machte; erst *Netscape* machte das Web wirklich massenattraktiv: „Es war vergleichbar mit dem Moment, da Radio zu Fernsehen wurde – und beinahe genauso dramatisch“¹³¹. Diese Popularisierung freilich gelang Andreessen und seinen Mitstreitern unter anderem, doch nicht zuletzt durch einen HTML-„Stilbruch“, den Web-Puristen bis heute als Sündenfall schlechthin begreifen¹³², andere jedoch als die überfällige Korrektur eines „Geburtsfehlers“ des Web¹³³: die Erweiterung des HTML-Codes um „physische“ Auszeichnungsbefehle.

Denn wo Berners-Lees HTML eine Überschrift zweiter Ordnung definierte, konnte, je nach Browser, in der Anfangszeit des Web etwas visuell sehr unterschiedliches herauskommen¹³⁴. In einem Browser war diese Überschrift „nur“ gefettet, in der Schriftgröße aber identisch mit dem Fließtext; im anderen war zwar die Schriftgröße heraufgesetzt, der Schnitt aber mager. Klassische logische Text-Auszeichnung eben: grob definiert gemäß inhaltlicher Charakteristik.

Wie passte da plötzlich der Befehl (englisch „tag“) namens „“ hinein, für „bold“, also „fett“, der in der 1992er-*Mosaic*-Version vorgesehen war? Oder „<i>“ für „italic“, also Kursivsatz? Dies waren keine logischen Auszeichnungen mehr, sondern physische. Das W3C – gleichsam der „Dachverband“ des Web, auf dessen Funktion und Rolle ich weiter unten noch zu sprechen kommen werde – sträubte sich eine Zeitlang gegen diese Neuerungen, musste aber auch einsehen, dass Andreessen und seine Mitstreiter längst „physische“ Fakten geschaffen hatten, hinter die man nicht mehr zurückkam¹³⁵.

Es folgte die zweite W3C-offizielle HTML-Version, die am 30.4.1993 veröffentlicht wurde. Die eigentliche, nachgerade revolutionäre Neuerung dieser „Version 2“ (die diese Versionsnummer offiziell gar nicht trug, sondern schlicht wieder „HTML“ hieß) war jedoch die Möglichkeit, Elemente zu platzieren, die heute teilweise ganze Webseiten alleine ausmachen. Andreessen höchstpersönlich hatte seinem *Mosaic*-Browser den revolutionären Befehl „“ beigebracht, der nun die höheren Weihen eines offiziellen HTML-Befehls erlangte: Es gelangten die ersten Bilder ins Web.

¹²⁷ Vgl. Kobert 2004, S. 27

¹²⁸ Vgl. Abbate 2000, S. 216

¹²⁹ Vgl. Champeon/Fox 1999, S. 48

¹³⁰ Vgl. Zehnder 1998, S. 73f.

¹³¹ Hillstrom 2005, S. 104

¹³² Vgl. Weinman/Lentz 1998, S. 173

¹³³ Vgl. Pring 2000, S. 18

¹³⁴ Vgl. Hillstrom 2005, S. 25f.

¹³⁵ Vgl. Maurice/Rex 2006, S. 16

0.6.2.4 Der „zahnlose Tiger“? Das W3C

Bereits um das Jahr 1994 herum zeichnete sich ab, dass es das eine, das „semantisch“ richtige und zentral definierte HTML absehbar nicht würde geben können im generisch dezentralen, daher auch gerne als „anarchisch“ bezeichneten World Wide Web. Zwar war in eben jenem Jahr, 1994, das „World Wide Web Consortium“ (W3C) durch Tim Berners-Lee gegründet worden, ein Gremium zur Standardisierung von Techniken im World Wide Web, zu denen natürlich auch und vor allem HTML gehörte und gehört. Tim Berners-Lee ist übrigens bis heute Vorsitzender des Konsortiums; er ist damit freilich „Chairman“ einer Organisation, die, anders als beispielsweise die „International Organization for Standardization“ (Internationale Organisation für Normung, ISO), ausdrücklich keine zwischenstaatlich anerkannte Organisation ist¹³⁶. Die vom W3C herausgegebenen Standards gehen daher formal auch nicht über den Charakter von Empfehlungen hinaus¹³⁷ (und sie heißen daher auch nur „Recommendations“¹³⁸), und auch tatsächlich ist das W3C zwar kein „zahnloser Tiger“, aber mindestens genauso wenig eine omnipotente Vereinigung oder gar eine Web-„Regierung“ – trotz der formellen Mitgliedschaft von über 370 teilweise sehr renommierten Unternehmen wie *Microsoft*, *SAP* oder *T-Online*¹³⁹. Zwar genießt das W3C bis heute hohes Ansehen, auch werden seine Empfehlungen durchaus weithin als „hochinstanzliche“ Setzungen respektiert. Seit sich jedoch mit dem Anbieten von Browsern Markterfolg erzielen lässt, ist der W3C-Standard von HTML zwar eine, aber längst nicht die alleinige Richtschnur von Webdesignern. Denn vor allem zwei Unternehmen hielten sich lange Zeit kaum bis gar nicht an die W3C-Empfehlungen, beziehungsweise preschten mit unabgestimmten Neuerungen immer wieder vor: *Netscape* und *Microsoft*.

0.6.2.5 *Netscape*, *Microsoft* und der „Browser-War“ der 1990er Jahre

Version 1.0 des *Navigators*, der Nachfolger der *Mosaic*-Browser, erschien im Oktober 1994¹⁴⁰. Rasch wurde deutlich: Marc Andreessen würde auch mit seiner frisch gegründeten Firma *Netscape* seinen innovativen Kurs der Vorjahre beibehalten. Auch *Netscape* begnügte sich nicht damit, dem *Navigator* die Darstellung von Webseiten nach W3C-Standard „beizubringen“. Die Firma kreierte eigene HTML-Befehle, vor allem „für Designer geeignete Tags“¹⁴¹, die notwendigerweise nicht HTML-Standard im Sinne des W3C waren und dementsprechend auch nur im *Navigator* dargestellt wurden. *Netscapes* Quasi-Monopol von teilweise über 90 % Weltmarktanteil gestaltete diesen Status Quo insgesamt recht erträglich; wer Webseiten mithilfe von *Netscape*-Befehlen kreierte, kreierte auf relativ solider Basis. Mit dem

¹³⁶ Vgl. Zehnder 1998, S. 5

¹³⁷ Vgl. Tolksdorf 2003, S. 2

¹³⁸ Vgl. Maurice/Rex 2006, S. 17

¹³⁹ Vgl. Hellbusch 2005, S. 33

¹⁴⁰ Vgl. Hafner/Lyon 2000, S. 313

¹⁴¹ Weinman/Lentz 1998, S. 173

Internet Explorer freilich, jenem Browser, mit dessen Markteinführung *Microsoft* 1995 gerade noch den Sprung auf den fast abgefahrenen Zug „World Wide Web“ bewerkstelligte, begann die Ära des HTML-Chaos'. Der „Browser-Krieg“¹⁴². Denn *Microsoft* folgte dem Beispiel des damals noch übermächtigen Konkurrenten.

„Beide Hersteller fügten ihren Produkten immer abenteuerliche Fähigkeiten hinzu, um jeweils den anderen Browser auszustechen“¹⁴³. Mit dem Zwang zur raschen Innovation wurde HTML damit freilich nicht nur immer diffuser, spaltete sich nach 1995 nicht nur auf in ein faktisches *Netscape*- und ein *Internet Explorer*-HTML (Kübler und Struppek sprechen in diesem Zusammenhang von „HTML-Dialekten“¹⁴⁴); die Weiterentwicklung der Auszeichnungssprache drohte in der Dynamik der Fortentwicklung außer Kontrolle zu geraten.

Wohlgemerkt: *Netscape* und *Microsoft* haben sich mit diversen ihrer Alleingänge zweifellos um die Weiterentwicklung von HTML sehr Veient gemacht¹⁴⁵, denn viele ihrer Befehls-Kreationen wurden später, quasi rückwirkend vom W3C zum HTML-Standard erhoben. Dennoch: Jeder Befehl, der von *Microsoft* oder *Netscape* ins HTML „hineingeflickt“¹⁴⁶ wurde, degradierete das W3C vom Setzungs-Organ zum Getriebenen. Das Gremium „ging dazu über, die Features in den bereits ausgelieferten Browsern zu sammeln und zu standardisieren, statt seine Zeit mit dem Nachdenken über zukünftige Standards zu verschwenden. Das Konsortium ging also vom Proklamieren zum Konsolidieren über“¹⁴⁷.

So erlaubte der *Netscape Navigator* beispielsweise bereits in Version 1.1, die im April 1995 in endgültiger Fassung erschien, die Definition und Darstellung von Tabellen; *Microsoft* zog im Oktober 1995 mit dem *Internet Explorer* 2.0 nach. Jedoch erst 21 Monate nach dem Marktzutritt von *Netscape* 1.1, 15 Monate nach dem des *Internet Explorers* 2.0, im Januar 1997 nämlich, wurden Tabellen offizieller W3C-Standard in der HTML-Version 3.2. Es zeugt im Übrigen auch von der damaligen Marktmacht der Firma *Netscape*, dass eine offizielle HTML-Version 3.0 des W3C nie erschien, da *Netscape* im August 1996 mit dem *Navigator* 3.0 bereits Fakten geschaffen hatte, auf die das Consortium wiederum nur noch reagieren konnte. Auf die W3C-HTML-Version 2.0 folgte also kurioserweise besagte Version 3.2. Dass zweifelhafte Innovationen wie der *Netscape*-Tag „<blink>“, der unerträglich blinkenden Text auf den Bildschirm brachte, sich nicht durchsetzten als Standard, sei an dieser Stelle nicht nur erwähnt, sondern ausdrücklich begrüßt.

Freilich legten *Netscape* wie *Microsoft* mit ihrem „Browser-Krieg“ den Grundstein dafür, dass heute noch das Schreiben selbst eines W3C-konformen, „sauberen“ HTML-Codes die frustrierende Konsequenz zeitigt, dass kaum ein Browser die besagte Seite so anzeigt, wie es der

¹⁴² Vgl. Seibert/Hoffmann 2006, S. 20; vgl. ferner Zehnder 1998, S. 74ff.

¹⁴³ Stocksmeier 2002, S. 23

¹⁴⁴ Vgl. Kübler/Struppek 1996, S. 2

¹⁴⁵ Vgl. Lubkowitz 2005, S. 62

¹⁴⁶ Kesper et al. 2004, S. 22

¹⁴⁷ Veen 2001, S. 27

einer Konkurrenzfirma tut¹⁴⁸. Abgemildert dauert der „Krieg“ bis heute an, mit anderen Protagonisten freilich, weil der Browser *Navigator* inzwischen marktanteilmäßig in die Bedeutungslosigkeit abgerutscht ist (dazu später mehr); auch die Einführung der „Cascading Style Sheets“ (siehe unten) hat leider nicht dazu beigetragen, dass browser-übergreifende Verlässlichkeit in die Darstellung von Webseiten eingekehrt ist¹⁴⁹. So gilt der Ratschlag von Kommer und Mersin bedauerlicherweise unverändert: „(Sie) sollten sich von dem Gedanken verabschieden, dass das Layout Ihrer Seiten in allen Browsern gleich aussehen muss. DAS WIRD OHNEHIN NIE DER FALL SEIN“¹⁵⁰.

0.6.3 Der steinige Weg: Typografie in Webbrowsern

Die verschlungenen Entwicklungspfade der Auszeichnungssprache HTML, die Kleinschärmützel und teilweise abstrusen, oft glücklicherweise ins Leere laufenden Befehlsschöpfungen der maßgeblichen Browserhersteller *Microsoft* und *Netscape* in der Mitte der 90er Jahre sollen hier nicht weiter im Detail rekonstruiert werden.

Ohnehin: Ist die Befassung mit den Fallstricken und Widrigkeiten der Herstellung von browser-basierter Typografie, einem klassischen Kommunikator-Problem also, im Zusammenhang mit dem auf Rezipienten orientierten Kernziel der vorliegenden Arbeit überhaupt von Relevanz – jenem Ziel nämlich, die Natur und Eigenart der Rezeption von Web-Texten auf Nutzerseite zu untersuchen? Ich meine: ja.

Zugegeben: Dass eine Majorität von Web-Nutzern tiefere Gedanken an die typografischen Sorgen der Webdesigner verschwendet, darf getrost als unwahrscheinlich gelten. Die Kreation von Web-Typografie ist höchstwahrscheinlich nicht zentrales Interesse derer, die morgens im Büro, statt unmittelbar zu Werke zu gehen, zunächst einmal die Schlagzeilen bei „Spiegel online“ überfliegen, jener, die aktuelle Bundesliga-Tabelle auf „kicker.de“ aufrufen, um für die Kaffeepausen-Diskussion mit statistischem Basiswissen versorgt zu sein – oder sich, dann doch ganz im Sinne der Geschäftsführung, im web-basierten Intranet über betriebliche Geschehnisse auf dem Laufenden zu halten. Als höchst wahrscheinlich ist vielmehr einzuschätzen, dass diese Nutzer – bewusst oder unbewusst – eine zielführende, angemessene und attraktive Typografie erwarten auf den Webseiten, die sie aufrufen – und dass sie Seiten, die eine adäquate (das heißt eben auch: optisch ansprechende) Schriftgestaltung vermissen lassen, zumindest tendenziell eher verlassen als solche, die qualitativ mehr zu bieten haben auf diesem Gebiet: „Nicht nur vom Inhalt werden Leser bei der Stange gehalten, sondern auch durch eine lesefreundliche Darbietung des Lesestoffs, was die Leser im Allgemeinen nicht bewusst wahrnehmen, aber deutlich empfinden“¹⁵¹.

¹⁴⁸ Vgl. Peck 2000, S. 36f.

¹⁴⁹ Vgl. Seibert/Hoffmann 2006, S. 23f.

¹⁵⁰ Kommer/Mersin 2002, S. 188

¹⁵¹ Bollwage 2001, S. 8

Das Problem: Anders als in der printbasierten Typografie nützte auch der bestwilligen Gestalterin von Webseiten das Wissen um die visuelle Beschaffenheit einer „optimal lesbaren“ Web-Schriftenanordnung wenig. Denn selbst wenn jene optimale Schriftenanordnung bekannt wäre – das ist nicht der Fall –, könnte diese Gestalterin, anders als der print-spezialisierte Kollege, immer noch nicht verlässlich zur Umsetzung schreiten. Der Königsweg zur technischen Realisierung einer gut rezipierbaren Schriftenanordnung im Web ist ebenso wenig identifiziert wie diese Schriftenanordnung selbst. Web-Typografie erfordert bis heute nicht nur gestalterische Kompetenz, sondern auch fundiertes Wissen um die Limitiertheit technischer Verwirklichung von optisch Wünschenswertem im Webbrowser. Was nützt beispielsweise der gelungenste Entwurf einer Webseite, der die ausschließliche Verwendung der Schriftart *Stone* vorsieht? Ein derartiger Entwurf ist technisch unrealisierbar fürs Web – und damit schlicht und ergreifend wertlos. Die folgende Darstellung der von mir identifizierten vier basalen Entwicklungsstufen der webbasierten, kommunikatorseitigen Typografie-Manipulation beleuchtet das „Sorgenkind“¹⁵² Web-Typografie daher auch in diesem technischen Sinne. Ich unterscheide dabei:

- Phase 1 (1989 – 1992): Ausschließlich logische Textauszeichnung
- Phase 2 (1993): Die ersten physischen Textauszeichnungen
- Phase 3 (1994 – 1998): Schriftvorschlag und Spaltensatz durch Tabellen-Layouts
- Phase 4 (seit 1998): Cascading Style Sheets (CSS)

0.6.3.1 Phase 1: Ausschließlich logische Textauszeichnung

Das Grundkonzept der logischen Textauszeichnung wurde bereits dargestellt: Texte werden ihrem inhaltlichen Charakter gemäß definiert, nicht in ihrer physischen Gestalt. Logische Auszeichnung markiert also beispielsweise eine Haupt-Überschrift oder Listenelemente als solche, überlässt dem Browser aber die letztliche visuelle Ausgestaltung dieser Elemente. HTML war rein logisch in der ersten Version, die ab 1989 existierte, obwohl die HTML-Urversion offiziell auf das Jahr 1992 datiert.

0.6.3.2 Phase 2: Die ersten physischen Textauszeichnungen

Mit der zweiten HTML-Version (nicht HTML 2.0 genannt, sondern immer noch schlicht „HTML“) vom April 1993 zogen die ersten „physischen“ Textauszeichnungen in den Befehlschatz der Auszeichnungssprache ein. Textpassagen konnten nun – immer vorausgesetzt, man betrachtete die Seiten mithilfe eines Browsers, der die neueste HTML-Version zu interpretieren in der Lage war – fett, kursiv, dicktengleich (also im „Fernschreiber“-Stil), unter-

¹⁵² Vgl. Lankau 2000, S. 366

und durchgestrichen dargestellt werden. Eine erste Abkehr vom logischen Auszeichnungsprinzip!

Nicht minder bedeutsam war in dieser Version, dass nunmehr auch Bilder auf Webseiten platziert werden konnten (anfänglich waren das praktisch ausschließlich GIFs, Bilder in einem Format, das die US-amerikanische Firma *CompuServe* entwickelt hatte). Aus streng mikrotypografischer Sicht ist dieses Faktum sicherlich nur am Rande von Bedeutung, aber für die Popularisierung des Web hat der ermöglichte Bildeinsatz in Browsern zweifellos eine ungemein treibende Dynamik entfaltet.

0.6.3.3 Phase 3: Schriftvorschlag und Spaltensatz durch Tabellen-Layouts

Der April 1995 kam, *Netscape* 1.0 war gerade einmal sechs Monate auf dem Markt, da markierte die Nachfolge-Version bereits nicht weniger als den vorläufigen Schlusspunkt einer Innovation, die aus gestalterischer Sicht zur Revolution geriet. Mit *Netscape* 1.1 begann das eigentliche Zeitalter des Webdesigns.

Denn diese Version bescherte den Webdesignern nicht nur, aber vor allem das Tag „“. Dieses erlaubte den Gestaltern von Webseiten immerhin, eine oder mehrere Schriftarten „vorschlagen“, in der die betreffende Seite typografisch aufbereitet werden sollte (dass die vorgeschlagene Schrift dabei auf dem Rechner des Betrachters installiert sein musste, war nicht nur damals bedauerlich, sondern ist bis heute lästige Einschränkung). Ein Befehl ist dies, der nun endgültig mit dem Grundprinzip von HTML als „logischer“ Auszeichnungssprache brach: „Was sagt über den Text aus, der von diesem Befehl eingeschlossen wird? Auf jeden Fall nichts über die Inhalte“¹⁵³. Münz macht in diesem Befehl gar „die schlimmste ‚Verunreinigung‘ von HTML als reiner Struktursprache“¹⁵⁴ aus.

Ferner war mit der neuen *Netscape*-Version die Möglichkeit da, Schriftgrößen zumindest relativ zu bestimmen; ein bestimmter Textbereich konnte also beispielsweise um „eine Stufe“ größer definiert werden als ein nachfolgender oder voranstehender auf ein und derselben Seite. Insgesamt bot *Netscape* 1.1 dabei sechs Schriftgrößen-Stufen an: „3“ war jeweils die vom Nutzer voreingestellte Größe, die übrigen orientierten sich relativ daran nach oben oder unten. Absolute Festsetzungen der Schriftgröße, beispielsweise in den Maßeinheiten „Punkt“ oder „Pixel“ waren also nicht möglich, wohl aber gestalterisch durchdachte, um Fettung oder Kursivstellung bereicherte gezielte Hierarchisierungen.

Fast schon zum Randaspekt geriet angesichts dieser – aus zeitgenössischer Sicht – atemberaubenden Neuerungen die Tatsache, dass mit der *Navigator*-Version 1.1 auch die Farbe einer Schrift bestimmbar war – und dass Bilder, die in den Vorgängerversionen stets eine verti-

¹⁵³ Veen 2001, S. 29

¹⁵⁴ Münz 2005

kale Unterbrechung im Textfluss gebildet hatten, jetzt in den Textfluss integrierbar waren: Die Buchstaben waren also um die Abbildung „herumführbar“ geworden. Makrotypografisch ein erkennbarer Fortschritt.

Blieben zwei wesentliche typografische Merkmale, die in der Vorgängerversion nicht bestimmbar gewesen waren: Zeilenbreite und Zeilenabstand. Zumindest in Sachen Zeilenbreite sollte die neue *Navigator*-Version gleichfalls Revolutionäres vorhalten.

Denn bislang fehlte Webdesignern das Instrumentarium zur Herstellung dessen, was Grundlage jedes klassischen Buchdrucks, jedes Zeitungs- oder Zeitschriften-Layouts ist: die Möglichkeit des Spaltensatzes. Die Möglichkeit also, informationstragende Blöcke neben- statt linear-scrollbedürftig untereinander zu platzieren, wie es nach wie vor Usus war im World Wide Web.

Das Tag „<table>“, ebenfalls 1995 neu eingeführt mit *Netscape 1.1*, eröffnete dieses Potenzial: Nun konnten zwei-, drei, vierspaltige Tabellen angelegt werden, die sich über die gesamte oder auch nur einen Teil der Seitenbreite erstreckten und in ihren Zellen Navigation, Hauptinhalt und Zusatzelemente beherbergen konnten. Indem man diesen Tabellen die Rahmenbreite „0“ zuwies, sie also als „blind“ charakterisierte, waren die nebeneinander liegenden Tabellenspalten zu Textspalten umfunktioniert, quasi typografisch zweckentfremdet¹⁵⁵. Die bis heute klassische „linke Navigationsleiste“ war nun möglich und auch rasch geboren! Und mit ihr auch die Möglichkeit, die Breite einer Spalte, mithin die Breite des in dieser Spalte platzierten Lauftextes, zu begrenzen.

Netscape 1.1 war eine Revolution. Eine Neuerung, die so gut wie alle Mittel zur Verfügung stellte, die bis heute zur Erstellung einer soliden Webseite erforderlich sind. Es überrascht kaum, dass wenig Ernstzunehmendes nachkam. Nach 1995 versuchten sich die Marktführer *Netscape* und *Microsoft* in diversen Erweiterungen des HTML-Befehlssatzes, die samt und sonders entweder überflüssig waren oder die Unübersichtlichkeit der konkurrierenden HTML-Versionen nur noch erhöhten; meist war beides der Fall.

0.6.3.4 Phase 4: Cascading Style Sheets (CSS)

Die Revolution frisst ihre Kinder. Das Tag „“, das sicherlich dem Berufsbild „Webdesigner“ nicht allein, aber maßgeblich die Chance zum Durchbruch gewährte, gilt heute W3C-offiziell als „veraltet“¹⁵⁶. Ein Lamento soll an dieser Stelle nicht intoniert werden, denn dem Tag folgte Besseres. Dass er aber Veienstvoll verstarb, sei zumindest erwähnt. Denn ohne ihn gäbe es heute wahrscheinlich auch keine Cascading Style Sheets (sinngemäß „kaskadierende Formatvorlagen“, kurz CSS).

¹⁵⁵ Vgl. McKelvey 1999, S. 28f.

¹⁵⁶ Vgl. Schäffer 2001, S. 190; vgl. ferner Wenz/Hauser 2004, S. 23; vgl. ferner Rottermund 2002, S. 41; vgl. ferner Peck 2000, S. 163

Das W3C stellten den ersten CSS-Entwurf 1994 vor und verabschiedete die erste offizielle CSS-Version („CSS Level 1“) im Jahre 1996¹⁵⁷. In gewissem Sinne eroberte sich das W3C an diesem Tage definitiv einen Gutteil seiner im „Browser-Krieg“ verloren gegangenen Souveränität und Setzungsmacht gegenüber den kommerziellen Anbietern (vor allem *Netscape* und *Microsoft*) zurück¹⁵⁸, denn das Konzept war und ist allzu schlüssig und unter Web-Gestaltern anerkannt als potentes Instrumentarium, als dass die kommerziellen Anbieter es schlicht hätten ignorieren können. Das *Netscape*-Konkurrenzprodukt *JavaScript Stylesheets* jedenfalls, das ab 1997 vom *Netscape Navigator* interpretiert wurde, spielt heute keine Rolle mehr. Die Version 3 des *Internet Explorers* interpretierte bereits diverse Stylesheets des W3C-Entwurfs.

Stylesheets sind vergleichbar den Formatvorlagen in print-orientierten Programmen wie beispielsweise *InDesign* oder *Word*¹⁵⁹: Alle (oder zumindest die wichtigsten) Eigenschaften einer bestimmten, selbst definierten Textsorte werden dabei einmalig definiert und dann immer wiederkehrend auf bestimmte Textpassagen „angewendet“¹⁶⁰. Die Stil- oder auch Formatfestlegung findet nach dem CSS-Konzept im Kopfbereich einer HTML-Datei statt oder – und das ist heute die Regel – in einer separaten Datei, die auf die Dateiendung „.css“ hört¹⁶¹ (prinzipiell sind CSS-Anweisungen auch auf einzelne Textpassagen direkt im HTML-Quellcode anwendbar, aber dieses Prinzip ist ebenso unökonomisch wie ungebräuchlich¹⁶²).

„CSS Level 1“ beschränkte sich weitestgehend auf Textgestaltung und Farben. 1997 kamen, als vorläufige kleinere Erweiterungen der CSS-Urversion, erste Layoutanweisungs-Befehle hinzu. „CSS Level 2“ erschien im Mai 1998, eine deutlich erweiterte Implementierung, mittels derer nun auch geschlossene Text-Bild-Ensembles (sogenannte „Boxen“) als Ganzheiten beschrieben werden konnten (oder auch nur in ihrer physischen Ausdehnung, Rahmen- und Farbgestaltung). Diese zweite Version¹⁶³ gilt im Wesentlichen bis heute als relevant. Im Folgenden wird sie – dem Thema dieser Arbeit folgend – vornehmlich in ihren mikrotypografischen Potenzialen beleuchtet.

Ein weiteres – prinzipiell konstituierendes, im Rahmen dieser Arbeit allerdings weniger interessantes – Moment im Konzept der „Cascading Style Sheets“ ist das Prinzip der „Vererbung“. So ist es beispielsweise möglich, eine bestimmte typografische Anordnung als basale Formatvorlage zu definieren und dieser Formatvorlage wiederum weitere Sub-Formatvorlagen hierarchisch unterzuordnen. Diese Sub-Formatvorlagen leiten sich dabei grundsätzlich aus der „Mutterbeschreibung“ her, indem sie deren Anweisungen übernehmen, sind aber bedarfsweise erweiterbar um weitere Detail-Definitionen. Es können also im CSS-Konzept ab-

¹⁵⁷ Vgl. Maurice/Rex 2006, S. 16

¹⁵⁸ Vgl. Champeon/Fox 1999, S. 188ff.

¹⁵⁹ Vgl. Dengler/Volland 2000, S. 126; vgl. ferner Kobert 2004, S. 475

¹⁶⁰ Weinman 2000, S. 302

¹⁶¹ Vgl. McKelvey 1999, S. 37

¹⁶² Vgl. Böhringer et al. 2003, S. 885ff.

¹⁶³ Vgl. World Wide Web Consortium (2006)

gestufte, daher konzeptionell als „kaskadierend“ bezeichnete Ketten von wechselseitig unter- und übergeordneten und sich aufeinander beziehenden Stilanweisungen konstruiert werden.

Worin lagen die elementaren Fortschritte, die die CSS dem Webdesign eröffneten? Zunächst ermöglicht das Konzept der „Cascading Style Sheets“ potenziell, HTML-Codes deutlich zu „entschlacken“, überkommene Befehlsketten der HTML-Frühzeit durch kürzere, schlankere Formtzuweisungen zu ersetzen. Dies ließ die Quelltexte von Webseiten nicht nur in ihrem quantitativen Umfang schrumpfen – und damit auch in der Ladezeit! –, sondern auch deutlich übersichtlicher geraten und mithin besser verwaltbar.

Ein Beispiel mag dies illustrieren: Wer im Jahre 1995, vor Erscheinen der ersten CSS-Version also, die Schlagzeile „Kohl und Jelzin besuchen Speyerer Dom“ auf einer Webseite in der Schriftart *Arial* zu definieren wünschte, in einer vage definierten Überschriften-Größe und gefettet-kursiv sowie in einem dunklen Grauton und zentriert ausgerichtet, schrieb vermutlich einen Quellcode, der dem folgenden recht ähnlich war:

```
<center><font face="Arial, Helvetica, Verdana" size="5" color="grey"><b><i>Kohl und Jelzin  
besuchen Speyerer Dom</i></b></font></center>
```

Das Problem: Wenn weiter unten auf derselben Webseite eine weitere Schlagzeile in exakt derselben typografischen Anordnung gewünscht war – zum Beispiel: „Lafontaine übernimmt SPD-Parteivorsitz “ –, musste der gesamte Befehlssatz wiederholt werden:

```
<center><font face="Arial, Helvetica, Verdana" size="5" color="grey"><b><i>Lafontaine  
übernimmt SPD-Parteivorsitz </i></b></font></center>
```

Im Konzept der Cascading Style Sheets ist dagegen nur eine einmalige Definition der typografischen Basal-Anordnung nach beispielsweise folgendem Muster erforderlich:

```
.schlagzeile { font-family:Arial,Helvetica,Verdana; font-size:18pt; font-weight:bold; font-  
style:italic; color:grey; text-align:center }
```

... und es genügt in HTML dann der Befehlsaufruf „“, um alle dahinter befindlichen Textbestandteile in der gewünschten Formatierung darzustellen (vorausgesetzt, der Betrachter hat CSS eingeschaltet und die Schriftart Verdana auf seinem Rechner installiert. Davon ist freilich zumindest zum Zeitpunkt der Erstellung dieser Arbeit in 95 % aller Fälle auszugehen). Die beiden obigen Beispiel-Schlagzeilen hätten sich im Quelltext der eigentlichen HTML-Seite im CSS-Konzept also nur noch wie folgt und eindeutig ökonomischer wiedergefunden:

```
<span class="schlagzeile">Kohl und Jelzin besuchen Speyerer Dom</span>  
<span class="schlagzeile">Lafontaine übernimmt SPD-Parteivorsitz </span>
```

Das Konzept der Cascading Style Sheets verschlankte damit in seiner konzeptionellen Stoßrichtung potenziell nicht nur die „Monstercodes“¹⁶⁴ und damit die Ladezeiten einzelner Webseiten. Es eröffnete auch die Möglichkeit, auf der Grundlage nur einer einzigen Formatierungs-Datei typografische Stringenz auf Dutzenden, ja Hunderten oder gar Tausenden von Einzel-Webseiten eines integrierten Auftritts zu gewährleisten¹⁶⁵ – und mithin auch im Web ein kohärentes visuelles Erscheinungsbild, im Sinne des Marketing ein Corporate Design auszuprägen, das sich bestimmend auch in der typografischen Grundkonzeption manifestiert¹⁶⁶. Ein Corporate Design, das im CSS-Konzept eben auch auf einen Schlag mittels weniger Korrekturen an einer einzigen Formatanweisungen in seiner typografischen Basisanlage beliebig und bequem zu reformieren war¹⁶⁷.

Und mehr noch. Die CSS reanimierten nämlich zumindest teilweise auch einen Grundgedanken, der, wie bereits dargestellt, den allerersten HTML-Versionen zugrunde gelegen hatte: den der Trennung von Inhalt und typografischer Form¹⁶⁸. Allerdings, und diese Relativierung ist genauso wichtig, war es eindeutig nicht Ziel der „Cascading Style Sheets“, das HTML-Rad darüber hinaus so weit zurückzudrehen, dass physische Anweisungen aus dem HTML-/CSS-Befehlskanon verschwinden sollten – ganz im Gegenteil. Die physische Textauszeichnung wurde in den „Cascading Style Sheets“ endlich nicht nur widerwillig geduldet als unwillkommene, aber unverhinderbare „Verschmutzung“ des vermeintlich reinen, logischen HTML-Urkonzepts von Berners-Lee, sondern akzeptiert und sogar adaptiert als integrales Element der Gestaltung von Webseiten. Mit diesem vermeintlichen Kotau hatte sich das W3C allerdings endlich wieder an die Spitze der Bewegung gesetzt: Die Weiterentwicklung von HTML und eben CSS lag wieder vornehmlich in den Händen des Konsortiums.

Auch der Ausbildung und Fortentwicklung des Berufsbildes des „Online-Journalisten“ oder auch „-redakteurs“ dürfte die Setzung der „Cascading Style Sheets“ förderlich gewesen sein. Denn wo die allerersten Ersteller von Webseiten in den frühen 1990er Jahren als quasi „Eier legende Wollmilchsauen“ nicht nur ihre Texte selbst verfasst hatten, sondern auch ihre Bilder eingescannt, formatiert und „nebenher“ in HTML auch noch das visuelle Erscheinungsbild der Seite eingerichtet hatten, traten in den späteren 90er Jahren erste Online-Redakteure auf den Plan, die das Handwerk des journalistischen Recherchierens und Schreibens wohl beherrschten, auch die Potenziale des neuen Mediums wohl kannten, nicht aber das Handwerk der HTML-Seitendefinition. Das heißt: Diese Menschen traten an, zum Gelingen eines Angebots beizutragen, das sie nicht oder nur unzureichend technisch-gestalterisch realisieren konnten. Die Separation von Inhalt und Form, im Printgewerbe gang und gäbe, war also

¹⁶⁴ Vgl. Seibert/Hoffmann 2006, S. 20

¹⁶⁵ Vgl. Lynch/Horton 1999, S. 94; vgl. ferner Nielsen 2000a, S. 77

¹⁶⁶ Vgl. z.B. Bollwage 2001, S. 94 ff.; vgl. ferner Rivers 2003, S. 14f.; vgl. ferner Abdullah/Hübner 2002, S. 31ff.

¹⁶⁷ Wenz/Hauser 2004, S. 45

¹⁶⁸ Vgl. Lubkowitz 2005, S. 263

auch aus diesem Grunde im Web dringlicher geworden – und wurde mit den „Cascading Style Sheets“ vielversprechend in Angriff genommen.

Bleibt vorläufig noch ein wichtiger Aspekt, der an dieser Stelle erwähnt werden muss. In mikrotypografischer Hinsicht nämlich haben die Cascading Style Sheets dem Webdesign schlussendlich eines des letzten wirklich wichtigen Felder erschlossen: das des Zeilenabstands. Denn dieser typografische Faktor war, allen *Netscape*-Alleingängen zum Trotz, bis 1996 praktisch unkontrollierbar gewesen in seiner optischen Physis auf Webseiten. Damit sind – zumindest theoretisch – seit inzwischen zehn Jahren prinzipiell alle wirklich zentralen typografischen Merkmale eines Textes auch auf Webseiten definierbar. Prinzipiell, wohlge-merkt.

0.6.3.5 Die Unwägbarkeiten bestehen fort

Die „Cascading Style Sheets“ haben den Gestalterinnen und Gestaltern von Webseiten zweifelsohne ein Mehr an Kontrolle darüber verschafft, was die potenziellen Betrachter auf ihrem Bildschirm zu sehen bekommen. Dennoch gilt im Wesentlichen fort, was Webdesign seit der Vorstellung des ersten Webbrowsers genauso spannend wie nervenzermürend macht: „(Das) Design bleibt (...) ein ‚Angebot‘ an den Leser einer Webseite, das dieser auch ablehnen kann und darf, indem er beispielsweise Bilder und Stylesheets ausschaltet“¹⁶⁹. Oder, dies sei als Beispiel ergänzt, indem der Nutzer die vom Designer als Schriftgröße definierten „11 Punkt“ schlicht als zu klein erkennt – und in seinem Webbrowser daraufhin schlicht unter dem Menüpunkt „Schriftgröße“ von „mittel“ auf „größer“ schaltet. Dass, als Folge dieser nutzerseitigen Manipulation, aus einer zweizeiligen Überschrift gerne mal eine protzig-vierzeilige wird, mag den Gestalter noch so sehr fuchsen: Der User sitzt in dieser Hinsicht immer noch am längeren Hebel. Es sei unbestritten, dass man dies „großartig und extrem nützlich“ finden kann „für jemanden, dessen Sehvermögen eingeschränkt ist“¹⁷⁰. Es ist und bleibt aber so, dass Webdesign auch im Zeitalter der Style Sheets ein Geschäft mit erheblichen Unwägbarkeiten geblieben ist, dass Web-Typografie weiterhin mit Näherungen und Spekulationen arbeiten muss, die Print-Gestalter zu Recht als bloße Zumutung empfinden würden. In diesem Sinne liegt Pring also nicht ganz richtig: „Die Hierarchie, die der Web-Autor (mithilfe von CSS, M.L.) einmal für seine Seiten festgelegt hat, setzt sich mit Nachdruck gegenüber etwa der vom User gewünschten Hierarchie durch“¹⁷¹.

Hinzu tritt, dass „CSS Level 2“ heute zwar im Prinzip von praktisch allen Browserherstellern als Richtschnur akzeptiert wird, nachdem Ende der 1990er Jahre durchaus ein gewisses Maß an hinhaltendem Widerstand bei den Marktführern auszumachen war¹⁷²: „Die Browserher-

¹⁶⁹ Maurice/Rex 2006, S. 264

¹⁷⁰ Crederholm 2006, S. 16

¹⁷¹ Pring 2000, S. 28

¹⁷² Pring 2000, S. 28

steller haben sich bei der Implementierung von Cascading Style Sheets (...) sehr viel Zeit gelassen. (...) Die älteren Programmversionen (...), das heißt *Internet Explorer* bis Version 5.5 einschließlich und *Netscape Communicator* bis Version 6.2 einschließlich, interpretieren eine Reihe von CSS-Features gar nicht oder fehlerhaft¹⁷³. Zumindest die angesprochenen Versionen des *Internet Explorers* dürften durchaus noch auf einigen zeitgenössischen Rechnern installiert und in Gebrauch sein.

Nun wird sich diese Einschränkung im Verlaufe der kommenden Jahre zweifellos „browserdemografisch“ auflösen; die veralteten Versionen werden sukzessive von den PCs verschwinden und aktuelleren Platz machen. Ein weiteres Problem allerdings scheint zur Verzweiflung von Webdesignern schier unausrottbar: Denn wiewohl die aktuellen Versionen der marktbeherrschenden Browser grundsätzlich – bis auf wenige Ausnahmen¹⁷⁴ – praktisch alle CSS-Befehle „kennen“, so interpretieren sie sie doch gelegentlich völlig unterschiedlich aus¹⁷⁵. Dies betrifft zwar den Satz der originär mikro- und mesotypografiebezogenen Befehle deutlich weniger als beispielsweise die CSS-Definition von Binnenabständen; ärgerlich ist es nichtsdestoweniger.

Anders als im „Browser-Krieg“ der 1990er Jahre sind also nicht mehr konkurrierende, inkompatible Befehlssätze das Hindernis – sondern nunmehr die abweichenden Visualisierungsprozesse auf der Grundlage eines semantisch eigentlich uniformierten Befehlskanons. Besonders tut sich hier wiederum Marktführer *Microsoft* hervor, dessen *Internet Explorer* viele W3C-Konventionen, gelinde gesagt, „eigenwillig“ interpretiert.

HTML- und CSS-basierte Seiten sind damit auch im Jahre 2006, allen Fortschritten zum Trotz, nach wie vor „die wohl feindlichste Umgebung für Typografie, die man sich vorstellen kann“¹⁷⁶. Vielleicht ist es etwas übertrieben, mit Pring das Gelingen von Schriftgestaltung in Webbrowsern schlicht zur „Glückssache“¹⁷⁷ zu erklären oder gar zu unterstellen: „Optisches Design mit HTML (...) ist wie der Versuch, ein Porträt mit einem Farbbroller zu malen“¹⁷⁸. Fakt aber ist und bleibt: Web-Typografie ist eine spezielle Herausforderung, die in ihren Fallstricken und Besonderheiten auch im Abgleich mit Bildschirmdokumenten anderer Programm-basierung – zum Beispiel PDF-Dokumenten – einer gesonderten Betrachtung bedarf. Denn es ist und bleibt ein „Balanceakt zwischen Möglichem und Zweckmäßigem“¹⁷⁹.

¹⁷³ Kesper et al. 2004, S. 20; vgl. ferner Champeon/Fox 1999, S. 11ff.

¹⁷⁴ Vgl. Kommer/Mersin 2002, S. 187

¹⁷⁵ Vgl. Lubkowitz 2005, S. 264f.

¹⁷⁶ Dengler/Volland 2000, S. 50

¹⁷⁷ Pring 2000, S. 14

¹⁷⁸ Vein 2001, S. 37

¹⁷⁹ Herrmann 2005, S. 215

0.6.4 Die „Marktanteile“ der Browser

Dass ausgerechnet *Microsoft* in der Interpretation von CSS-Spezifikationen – ob willkürlich, aus Missinterpretationen oder Unvermögen heraus – wieder einmal Sonderwege beschreitet, ist durchaus bedeutsam. Denn augenscheinlich nutzen zwar nach wie vor die meisten User weltweit den *Internet Explorer*, um Webseiten zu betrachten – es wäre allerdings grundfalsch, daraus abzuleiten, dass Gestalter und Programmierer lediglich auf die *Microsoft*-Fehler in der grafischen Interpretation von CSS-Anweisungen Rücksicht nehmen zu bräuchten, um taugliche Angebote zu erstellen. Denn die Konkurrenz erobert stetig Marktanteile.

Nun sind die exakten Marktanteile der Browser definitiv nicht eruierbar; diese zu ermitteln, setzte eine Vollerhebung unter Webnutzern voraus, die allein organisatorisch schlicht nicht zu verwirklichen wäre. Es gibt aber durchaus Hinweise. Anerkannt sind insbesondere die Daten der Firma *WebHits*, eines Anbieters von Web-Statistiken und Software zur statistischen Auswertung von Websites. *WebHits* stellte für den August 2006 folgende Browser-Rangliste vor, die laut Firmenangabe aus den Daten von 8100 Abonnenten des Dienstleisters aus aller Welt gewonnen wurden¹⁸⁰ (die Daten anderer Anbieter¹⁸¹ bestätigen die von *WebHits* erhobenen Relationen im Wesentlichen):

Tabelle 0.1: Marktanteile der Browser	
Browser	Marktanteil
<i>Microsoft Internet Explorer</i>	71,0 %
<i>Mozilla Firefox</i>	23,1 %
<i>Apple Safari</i>	2,7 %
<i>Netscape Navigator</i>	1,6 %
<i>Opera</i>	1,4 %
<i>Sonstige</i>	0,2 %
Quelle: http://www.webhits.de	

Bemerkenswert ist insbesondere der Aufstieg des Browsers *Mozilla Firefox*, der erst Anfang 2002 auf den Markt trat (und im Übrigen auf dem Basis-Quellcode des Ex-Marktführers *Netscape* beruht, der mit 16 % Anteil inzwischen sogar hinter den *Apple*-Browser *Safari* zurückgefallen ist). Fast ein

Viertel aller Nutzer geht mittlerweile mit *Firefox* ins Web. Von einem Quasi-Monopol von *Microsoft* kann also definitiv nicht mehr die Rede sein; ein Webseiten-Entwurf, der im *Internet Explorer* korrekt dargestellt wird, sollte mithin stets auch auf seine Interpretation zumindest im *Firefox* überprüft werden.

¹⁸⁰ Webhits Internet Design GmbH 2006

¹⁸¹ Vgl. unter anderem W3 Schools 2006

1 Sprachfixierung und Leseprozess

1.1 Die Kulturleistung der Sprachfixierung

Die Feststellung mag banal erscheinen, grundlegend ist sie nichtsdestoweniger: Schrift ist fixierte gesprochene Sprache. Welch ungeheure Kulturleistung: Um sich verbal auszutauschen, bedarf der schriftkundige Mensch nicht mehr des persönlichen Tête-à-têtes mit einem Adressaten – vorausgesetzt natürlich, dessen Schriftkundigkeit reicht aus, den fixierten Zeichen-Code des Mitteilenden zu entschlüsseln –, nicht einmal die Stimme braucht der Sendende noch zu erheben, um Ideen mitzuteilen. Doch nicht nur dies: Gedanken sind aufzeichnbar, mithin bewahrbar durch Sprachfixierung – und damit konservierbar über den Tod des Sendenden hinaus, in sicherlich höherem Maße, als dies jede mündliche Überlieferung vermag. Die Spekulation sei gestattet: Dass das Märchen von „Hänsel und Gretel“ ohne die schriftliche Aufzeichnung der mündlichen Überlieferung durch die Brüder Grimm heute seinen Bekanntheitsgrad genießen würde, ist mehr als zweifelhaft.

Dieses Kapitel soll in zweierlei Hinsicht auf das Ziel der vorliegenden Arbeit hinführen. Zum einen wird Wesen und Charakter jenes Basissystems der Sprachfixierung vorgestellt, auf dessen kodifizierter Grundlage verschriftlichte Information in unserem Kulturkreis organisiert ist: das alphabetische System der Buchstaben, der Phonogramme. Im zweiten Teil dieses Kapitels werden Modelle der Kognitionspsychologie diskutiert, die die rezipierende und sinnentnehmende Verarbeitung eben jener Schriftzeichen durch Lesekundige zu erklären suchen. Aus diesen Erkenntnissen versuche ich, Rückschlüsse für „gute“ Typografie auch am Bildschirm zu gewinnen.

1.2 Die Systeme der Sprachfixierung

Sprachfixierung setzt einen akzeptierten Kanon definierter Zeichen voraus, die mit konsensuellem Bedeutungsgehalt ausgestattet sind¹⁸². Die Erforschung von Zeichensystemen – und zwar längst nicht nur jener sprachrepräsentierender Natur – ist das Geschäft der „Semiotik“, die meist synonym begriffen und bezeichnet wird mit dem Terminus „Semiologie“¹⁸³. Und wiewohl die Systematisierung und Erklärung der Beschaffenheit und Funktion von Schriftzeichen nur einen kleinen Ausschnitt der Gegenstände dieser wissenschaftlichen Disziplin bildet¹⁸⁴, kann selbst dieser Ausschnitt im Folgenden nicht mehr als nur angerissen werden. Einige interessante und für die weitere Diskussion dieser Arbeit hilfreiche Erkenntnisse werden sich nichtsdestoweniger herauskristallisieren.

¹⁸² Vgl. Gross 1994, S. 46

¹⁸³ Vgl. Müller 2003, S. 158f.

¹⁸⁴ Vgl. Eco 1972, S. 15ff.

Frutiger unterscheidet zunächst zwei grundlegend differierende Systeme der schriftlichen Sprachfixierung: das „alphabetische“ nämlich und das „bildhaft gebliebene“¹⁸⁵ (das andernorts auch als „ideografisches“ bezeichnet wird¹⁸⁶). Im Grunde genommen ist diese grobe Kategorisierung irreführend, zumindest aber unscharf: Unser „A“ ist schließlich zweifellos ebenso ein Bildzeichen, ein „Piktogramm“, wie es das chinesische Schriftzeichen für „Pferd“ ist. Die basalen Unterschiede zwischen beiden Fixierungssystemen erschließen sich erst auf den zweiten Blick – und hier wiederum auf zweierlei Ebenen.

Wobei auch das erste dieser beiden Fein-Differenzierungsmerkmale zwischen beiden Systemen zugegebenermaßen ein immer noch recht „schwammiges“ ist. So grenzt Frutiger „alphabetische“ (Ende des 19. Jahrhunderts sprach man diesbezüglich offenbar von „alfabetarischen“¹⁸⁷) und „bildhaft gebliebene“ Fixierungsmethoden voneinander ab durch den Grad, in dem sich die Piktogramme von ihren rein bildhaften Ursprüngen gelöst und den Charakter von Zeichen autonomer Qualität erlangt haben. Zu den „alphabetischen“ Systemen ist dabei zweifellos das lateinische Alphabet zu zählen, unter dessen Verwendung auch die vorliegende Arbeit verfasst wurde, ebenso sind die kyrillische und die griechische Schrift unter dieser Kategorie zu führen: „Die Alphabetschrift lässt sich nicht mehr an den Dingen ablesen; das Pferd heißt nicht Pferd, weil es so aussieht“¹⁸⁸. Als Prototyp der „bildhaft gebliebenen“ Systeme der Sprachfixierung gilt dagegen die chinesische Bildschrift, auch wenn diese mit ihren insgesamt 54 Grund-„Deutzeichen“¹⁸⁹ natürlich auch Simplifizierungen unterzogen wurde im Verlaufe der Jahre – ein Code-System ist dies immerhin, das für ein gutes Fünftel der Weltbevölkerung als verbindlich gilt.

Warum aber die zunächst umständlich wirkende Titulation als „bildhaft geblieben“? Weil – Frutiger zufolge – davon auszugehen ist, dass auch die bekannten Alphabete samt und sonders bildhaften Ursprungs sind. Auch das abstrakte Zeichen „A“ unseres lateinischen Alphabets ist keine abstrakte Eigenschöpfung. Es hat sich, wie die chinesischen Bildzeichen, aus einer bildhaften Methode der Sprachfixierung entwickelt. Abbildung 1.1 zeichnet die Entwicklung des Bildzeichens für Stier (griechisch „Aleph“) vom recht ausgefeilten Abbild zum eigenständigen, extrem abstrahierten Qualizeichen nach. Wie dieselbe Abbildung zeigt, hat auch das chinesische Bildzeichen für



¹⁸⁵ Vgl. Frutiger 2001, S. 113ff.

¹⁸⁶ Vgl. Bergner 1990, S. 17

¹⁸⁷ Vgl. Wuttke 1872, S. 14

¹⁸⁸ Gross 1994, S. 47

¹⁸⁹ Vgl. Bergner 1990, S. 17

„Pferd“ durchaus ähnliche Entwicklungen durchgemacht; der Grad der systematischen Simplifizierung – der nicht allein, aber durchaus festzumachen ist am Vergleich des Zeitraumes, der zur Niederschrift der Ikonen aufzubringen ist – ist allerdings weit geringer als beim „A“.

Es sollte an dieser Stelle nicht unerwähnt bleiben, dass diese generische Theorie nicht nur Befürworter kennt. Leroi-Gourhan begreift beispielsweise die Abstrahierung von Piktogrammen hin zu Zeichen der Schriftfixierung ausdrücklich nicht als bloßen Teil eines linearen Prozesses der historischen Weiterentwicklung von Abbildern, sondern als originär-schöpferischen, intellektuell eigenständigen Akt¹⁹⁰, der mithin die natürliche Metamorphose der Zeichen gebrochen haben müsse; Gross erblickt ebenfalls ein „ideologisches Konstrukt“ in der „Sukzession von der Zeichnung zum Piktogramm, zum Ideogramm und schließlich zum phonetischen Graphem“¹⁹¹. Wobei verwirrt, dass die Autorin an späterer Stelle erklärt, dass die „meisten Alphabetbuchstaben (...) ursprünglich abbildenden Charakter (hatten), (...) also verblasste Ikon-Zeichen (sind)“¹⁹² – der Begriff „verblasst“ impliziert in meinem Verständnis einen gleichsam „natürlichen“ Werdungs-Prozess, den die Autorin ja gerade bestreitet. Doch wie immer auch die Genese von alphabetischen wie „bildhaft gebliebenen“ Systemen historisch vonstatten gegangen sein mag: Für die vorliegende Untersuchung ist nicht im Kern der Prozess der Entstehung, sondern der moderne Charakter von Zeichensystemen von Bedeutung. Dass die Ursprünge beider Systeme letztlich im Bildlichen wurzeln, unterstelle ich als Konsens. Die Umstände ihrer jeweiligen Detail-Genese bis hin zum zeitgenössischen Status mögen diskussionswürdig sein; im Rahmen der vorliegenden Arbeit sind sie eindeutig von nachgeordnetem Interesse.

Denn das manifeste Unterscheidungsmerkmal zwischen den beiden Schriftfixierungs-Systemen macht sich entscheidend an anderer Stelle fest. Unsere lateinische Schrift nämlich basiert – anders als die ideografischen Systeme – auf dem Grundprinzip, dass eine normierte Anzahl definierter Schriftzeichen (Grapheme) jeweils einem mehr oder minder klar definierten hörbaren Laut (Phonem) zugeordnet sind (und umgekehrt) – ein Prinzip, das nach herrschender Meinung um 1500 vor Christus von den Phöniziern in seinen Grundzügen (jedoch vorerst nur auf Konsonanten-Zeichen basierend) konzipiert¹⁹³ und um 800 in Griechenland durch die Erweiterung um Vokal-Zeichen entscheidend fortentwickelt wurde¹⁹⁴. Das „Piktogramm“ wird im „alphabetischen“ System damit zum „Phonogramm“: „Eine Buchstabenschrift (...) zeichnet nicht jedes einzelne Wort als ein bildhaftes Zeichen auf, sie bezieht sich also nicht nur auf das Gesamtschema ‚Wort‘, sondern sie bezieht sich auf die abgebildeten Laute der Sprache (die Buchstaben also) – eine höchst ökonomische Erfindung, die es er-

¹⁹⁰ Vgl. Leroi-Gourhan 1980, S. 249ff.

¹⁹¹ Gross 1994, S. 47

¹⁹² Gross 1994, S. 48

¹⁹³ Vgl. Barthel 1972, S. 140ff.

¹⁹⁴ Vgl. Bollwage 2001, S. 62f.

möglichst, jedes neue Wort nicht durch ein weiteres Zeichen, sondern mithilfe weniger (also der rund 30) Buchstaben aufzubauen. Das entlastet unser Gedächtnis enorm“¹⁹⁵.

Doch nicht nur dies. Im Prinzip ist nämlich in einem alphabetischen Buchstabensystem jeder Mensch, der sich einmal die grundlegenden Gestalt-Klang-Beziehungen dieses Systems angeeignet hat, in der Lage, ein Wort, das er niemals zuvor gehört, geschweige denn selbst ausgesprochen hat, sprechend von einer geschriebenen Vorlage zu reproduzieren. Dies ist zweifellos einer der Vorzüge, die alphabetische Systeme gegenüber „bildhaft gebliebenen“ auszeichnet. Dass die „Notation“ der Sprache in diesem System weit ungenauer ist, als das einleuchtend revolutionäre Grundprinzip vermuten lässt¹⁹⁶ – und beispielsweise auch weniger eindeutig als das musikalische Notensystem¹⁹⁷ –, beweist sich bereits im divergierenden Klang des Buchstaben „e“ in den Wörtern „leiblos“, „Auge“ und „Keil“: die „Notationalität“ unserer Buchstaben ist in Bezug auf ihre Lautentsprechungen alles andere als absolut¹⁹⁸. Die heiß bekämpfte Rechtschreibreform der neunziger und vergangenen Jahre hatte auch das Ziel, hier mehr Eindeutigkeit zu schaffen; was daraus geworden ist, soll an dieser Stelle nicht weiter vertieft werden.

1.3 Der Prozess der Textdekodierung

Es bleibt festzuhalten: In ihrer Grundkonstruktion sind „alphabetische“ und „bildhaft gebliebene“ Sprachfixierungs-Systeme eindeutig unterschiedlicher Basis-Natur. Umso mehr überrascht, dass eine populäre – aber, wie zu zeigen sein wird, in der Wahrnehmungspsychologie längst als offenbar unhaltbar klassifizierte oder zumindest verkürzt beschreibend erkannte – Theorie unterstellt, dass geübte Lesende beider Schriftkulturen im Grunde identische Strategien der Erfassung schriftlicher Mitteilungen anwenden: die „Wortbild-Erfassung“. Da diese These jenseits der wissenschaftlichen Psychologie nach wie vor eine merkliche Popularität genießt (auch in der Journalistik), werde ich sie im Folgenden zunächst – unter dem angesprochenen Vorbehalt wohlgerne – vorstellen. Um sie anschließend mit den herrschenden Thesen zu konfrontieren.

1.3.1 Die Gestaltpsychologie als Grundlage

Gemeinsam ist allen im Folgenden vorgestellten Erklärmodellen die Basis: die Gestaltpsychologie. Diese Theorie wurde ab den 20er Jahren des vergangenen Jahrhunderts maßgeblich unter anderem in Deutschland, insbesondere von einer Forschergruppe um den Berliner

¹⁹⁵ Menzel 2000, S. 15

¹⁹⁶ Vgl. Blumenstock 1997, S. 16f.

¹⁹⁷ Vgl. u.a. Nöth 1985, S. 390ff.

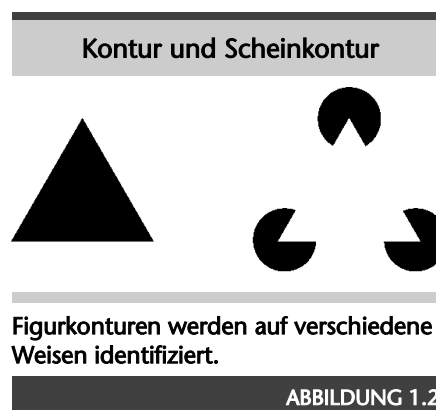
¹⁹⁸ Vgl. Gross 1994, S. 45

Max Wertheimer, entwickelt¹⁹⁹. Sie bildete im Gefolge zahlreiche Schulen aus, so dass streng genommen nicht von einer umfassenden, in sich geschlossenen Theorie gesprochen werden kann²⁰⁰; ein Kanon an Grunderkenntnissen aber lässt sich identifizieren, der heute in ihren Wesensgedanken als weitgehend unumstritten gelten darf²⁰¹.

Als Fundamentaltheorie befasst sich die Gestalttheorie mit der Frage der menschlichen „Wahrnehmungsorganisation“²⁰² – mit dem Problem also, auf welche Weise Menschen die zahllosen Einzelreize, die die Umwelt für sie bereit hält, zu verarbeitbaren Konstrukten vervollständigen. In diesem universalen Sinne entfaltet die Gestaltpsychologie also auch Bedeutung für die Lehre von der Typografie, weshalb sie hier – wenn auch sehr kurz und komprimiert – in ihren basalen Dogmen vorgestellt werden soll.

Grundlage der Gestalttheorie ist die Überzeugung, dass es dem Menschen unmöglich sei, sämtliche Einzelreize seiner Umgebung als individuelle, partikuläre Stimuli zu rezipieren und zu verarbeiten: „Wir würden ein Kaleidoskop unverbundener farbiger Stückchen erleben, die sich vor unseren Augen bewegen und umherwirbeln“²⁰³. Die Gestalttheorie geht daher davon aus, dass Menschen visuelle Einzel-Stimuli zu sinnvollen Formen, organisierten Gesamtheiten (Gestalten) addieren – nach dem Prinzip, dass das Ganze mehr sei als die Summe seiner Einzelteile²⁰⁴. Der Ansatz der Gestaltpsychologie ist also ein konstruktiv-ganzheitlicher²⁰⁵, kein partikularistischer; die Wahrnehmung wird nicht über das Erkennen kleinster visueller Einheiten definiert, sondern als Prozess der Figur-Identifikation. Michael Wertheimer streift bereits die Grenzen der Poetik, wenn er die zentrale Idee wie folgt umreißt: „Auch wenn kein menschlicher Geist in der Lage sein mag, das Ganze der Schöpfung voll zu erfassen, so ist es uns doch möglich, (...) Strukturen zu entdecken“²⁰⁶.

Ein wesentlicher Grundbegriff ist dabei der der „Figur“. Um nämlich eine oder mehrere „Figuren“ innerhalb einer visuellen Darstellung schlüssig identifizieren zu können, unternimmt das betrachtende Individuum in der Erklärung der Gestaltpsychologen eine Separation von Vorder- und Hintergrundbereichen – das Gehirn organisiert also die sogenannte „Figur-Grund-Beziehung“²⁰⁷. Kleine Flä-



¹⁹⁹ Vgl. Goldstein 2002, S. 190f.

²⁰⁰ Vgl. Blankertz/Doubrawa 2005. S. 156ff.

²⁰¹ Vgl. Neel 1974, S. 341f.

²⁰² Vgl. Zimbardo/Gerrig 2004, S. 176

²⁰³ Zimbardo/Gerrig 2004, S. 175

²⁰⁴ Vgl. Neel 1974, S. 342

²⁰⁵ Vgl. Müller 2003, S. 171f.

²⁰⁶ Wertheimer 1991, S. 123f.

²⁰⁷ Vgl. Glaser 1998, S. 240

chen werden dabei tendenziell eher als Figur erkannt, größere als Grund²⁰⁸; sehr oft teilen sich mehrere Figuren einen Grund²⁰⁹.

Figuren werden dabei über ihre Gestalt-Kontur identifiziert, ihren erkannten Umriss. Kompakte Farbflächen wie beispielsweise das schwarz gefärbte Dreieck im linken Bereich von Abbildung 1.2 erschließen sich relativ problemlos; doch auch real nicht vorhandene Konturen ist das menschliche Individuum in der Lage zu erkennen. Man spricht hier von der Konstruktion von „Scheinkonturen“²¹⁰, die, einer kognitiven „Schließungstendenz“ folgend, auch unvollständige Gestalten kognitiv identifizierbar werden lassen²¹¹. Die rechte visuelle Konstruktion in Abbildung 1.2 veranschaulicht das Phänomen der „Scheinkontur“, das dem kognitiven Bedürfnis Rechnung trägt, „Reize als vollständig, ausgewogen und symmetrisch wahrzunehmen, auch wenn Lücken, Unausgewogenheit und Asymmetrien vorliegen“²¹². Dieses Phänomen wird auch beschrieben als das „Gesetz der guten Gestalt“, das Individuen offenbar instinktiv zur Anwendung bringen bei der Identifikation von Figuren.

Dass die Differenzierung von Figur und Grund allerdings nicht immer reibungslos vonstatten geht, beweist unter anderem der berühmte „Rubinsche Becher“²¹³ (auch bezeichnet als „Kippfigur“, vgl. Abbildung 1.3): Sehen wir hier zwei schwarze Gesichtsp Profile – oder eine weiße Vase? Hier wirkt das „Gesetz der guten Gestalt“ gleichsam verhängnisvoll: Denn sowohl Schwarz- als auch Weißfläche erlauben in diesem Bild die Konstruktion harmonisch-symmetrischer Figuren; die Kognition steht mithin vor einem Erkenntnis Konflikt, welcher Bereich als „Grund“ zu verstehen sei, welcher als „Figur“.



Inwiefern kann die Gestalttheorie helfen beim Verständnis des Rezeptionsprozesses beim Lesen, in der Erklärung des Erfassens des visuellen Typs „Schriftzeichen“ also? Zunächst einmal findet sich die basale Erkenntnis, dass auch Schriftzeichen im Sinne der Gestalttheorie als ganzheitliche „Gestalten“ begriffen werden können, die als „Figuren“ erkennbar sind, sofern und weil sie sich von einem wie auch immer beschaffenen Hintergrund abheben²¹⁴. Diese Erkenntnis immerhin erklärt beispielsweise, warum wir sowohl schwarze Buchstaben auf weißem Grund als solche zu rezipieren in der Lage sind als auch weiße Buchstaben auf schwarzem Grund. Eine weit wichtigere Erkenntnis freilich ergibt sich, löst man sich von der Einzelfigur und betritt man die typografisch nächsthöhere Ebene: die Buchstabengruppe. Das Wort.

²⁰⁸ Vgl. Redelius 1998, S. 19

²⁰⁹ Weiland/von Gizycki 2002, S. 35ff.

²¹⁰ Vgl. Goldstein 2002, S. 191

²¹¹ Vgl. Zimbardo/Gerrig 2004, S. 176

²¹² Zimbardo/Gerrig 2004, S. 176

²¹³ Vgl. Lefrancois 1994, S. 101

²¹⁴ Vgl. Friedl 1998, S. 18

Ein weiterer wichtiger Aspekt der Gestalttheorie ist nämlich derjenige der „Gruppen-“ oder auch „Musterbildung“, der Wahrnehmung von Figurenensembles. Danach werden verschiedene, voneinander räumlich unabhängige Figuren unter bestimmten Voraussetzungen als visuelle Gruppen erfasst. Diese Voraussetzungen wurden in verschiedenen differenzierteren „Gesetzen“ beschrieben und erklärt, von denen zwei besonders interessant und anwendbar sind auf die Typografie: das „Gesetz der Ähnlichkeit“ sowie das „Gesetz der Nähe“ (eine ausführlichere Darstellung von Ansätzen der visuellen Mustererkennungen – beispielsweise des „Gesetzes der guten Fortsetzung“ – soll an dieser Stelle unterbleiben²¹⁵, da diese großenteils nur Fingerzeige liefern, die für die Diskussion dieser Arbeit lediglich sehr entfernt von Wert wären).

Das „Gesetz der Ähnlichkeit“ unterstellt dabei, dass konturlich gleichartige oder zumindest ähnliche Einzel-Figuren die Tendenz zur visuellen Gruppierung aufweisen in der kognitiven Verarbeitung. Begreift man beispielsweise die lateinisch-alphabetischen Schriftzeichen als einen Satz visuell durchaus verwandter Figurtypen, so erklärt dieses Gesetz, warum Textgesamtheiten als solche identifiziert, jedoch auch wechselseitig als voneinander abgegrenzt erlebt werden können.

Insbesondere den Prozess der individuellen Worterkennung und -abgrenzung hilft das „Gesetz der Nähe“ zu erklären. Es besagt schlicht, dass räumlich nahe stehende Figuren tendenziell als Einheiten erlebt werden. Wort- und Zeilenzwischenräume als typografisch sinnhaftig trennende Raum-Einheiten erfahren hier also ihre gestalttheoretische Begründung.

1.3.2 Die Theorie von „Wortbild“ und „Wortsilhouette“

Wie lassen sich nun die obigen Erkenntnisse fortentwickeln zu einer tragenden Erklärung jener komplexen Verrichtung des „Lesens“? Eine recht populäre, aber, wie bereits angedeutet, augenscheinlich heute kaum mehr haltbare Theorie setzt dabei nicht bei der isolierten Figur – dem Schriftzeichen – an, sondern bei visuellen Gruppen: Silben und Wörtern als Gesamtheiten, die zum einen eine fixe, wiederkehrende und damit wiedererkennbare Buchstabenkombination gemein haben, zum anderen eine markante Umrissform, eine individuelle Silhouette.

Bei geübten Lesern unseres Kulturkreises tritt – so der Erkläransatz – das Buchstabe-für-Buchstabe-Entziffern (das, wie gezeigt, fundamentale Prinzip „alphabetischer“ Schriften ist) fast vollständig zurück hinter das Erfassen ganzer Buchstabenensembles, oft voller Wörter, die dann eben nicht mehr als Ketten von Einzel-Zeichen wahrgenommen, schon gar nicht wie im frühen Grundschulalter „entziffert“ werden, sondern als geschlossene, bildliche Gesamt-

²¹⁵ Ausführlich vgl. unter anderem Metzger 1976, S. 142ff.; vgl. ferner Guski 1996; vgl. ferner Metzger 1966, S. 62ff.; vgl. ferner Zimbaro/Gerrig 2004, S. 176ff.; vgl. ferner Rock 1987, S. 7ff.; vgl. ferner Goldstein 2002, S. 190ff.

heiten, die ihrerseits wiederum charakteristische Silhouetten ausformen²¹⁶. Luidl erklärt es in seinem Standardwerk „Typografie“ so: „Erfasst werden Wortbilder und mit Wortbildern in unserem Gedächtnis verglichen. Deckt sich das Wortbild vor uns mit dem Wortbild in uns, so haben wir das Wort erkannt, wir haben es gelesen“²¹⁷. Dieser Bild-Schatz besteht laut Adrian Frutiger, dem Schöpfer einiger der großartigsten Schriften der zweiten Hälfte des 20. Jahrhunderts, allerdings nicht nur aus ganzen Wörtern, sondern auch aus Silben, und sie werden ihm zufolge vorgehalten „in jedem Leser wie ein tabellenartiges Grundschema“²¹⁸. Während Frutiger und Luidl also stark auf „Buchstabenensembles“ fokussieren, bezeichnet Rehe die Wort-Silhouette als bedeutsameres Phänomen – im Jahre 1974 bereits, aber durchaus spät genug, um konkurrierende Ansätze wie das „Pandämonium-Modell“ (siehe unten) zu kennen – als entscheidende visuelle Form, „die der Leser wahrnimmt und in seinem Gedächtnis speichert. Hat er diese Wort-Silhouette erst einmal in sich aufgenommen, so wird er das Wort immer wieder erkennen, ohne Buchstaben für Buchstaben entziffern zu müssen“²¹⁹. Doch ob nun eher auf die Silhouette abgehoben wird oder auf das Wort-„Bild“ aus Zeichen: Beide Ansätze sind insofern höchst eng verwandt, als sie in ganzheitlichem Sinne eine Ablösung propagieren von der isolierten Betrachtung von Lettern hin zum Buchstabenensemble als den Leseprozess vorrangig bestimmender Größe.

Im Lichte früher Erkenntnisse der Lesbarkeitsforschung – die teilweise bis in das vorvergangene, das 19. Jahrhundert zurückreichen – war dieser Interpretationsansatz des Leseprozesses durchaus nicht unsinnig: Schließlich gilt bis heute als vollkommen unbestritten, „dass die zur Erkennung eines Wortes notwendige Darbietungszeit sich nicht summativ aus denen für die einzelnen Buchstaben des Wortes ergibt“²²⁰. Der „Wortbild“-Ansatz erlaubt unbestritten eine vernünftige Interpretation dieser zentralen Erkenntnis. Und er hielt sich lange: In weitgehend identischem Muster und mit äquivalenter Quintessenz wie in den vorangehenden Abschnitten argumentierten etwa Rayner und Pollatsek²²¹ sowie Aicher²²² in 1989 im Sinne der „Wortbild“-Theorie, Bergner tat es 1990²²³; Wendt verfocht noch 1992 die These der „Wortbilder“²²⁴, genau wie es Lynch und Horton sogar 1999 taten²²⁵.

Immerhin – dies sei zugestanden – ist aus dem „Wortbild“-Erklärungsansatz heraus unter anderem erklärbar, warum in verschiedenen Studien in Versal-Buchstaben gesetzte (also nur in Großbuchstaben verfasste) Texte als schlechter – zumindest langsamer – lesbar erkannt wurden als Texte, die in orthografisch korrekter Manier aus Groß- und Kleinbuchstaben (Minus-

²¹⁶ Vgl. Wendt 1992, S. 272ff.

²¹⁷ Luidl 1989, S. 92

²¹⁸ Frutiger 2001, S. 198

²¹⁹ Rehe 1981, S. 39

²²⁰ Vgl. Wendt 1992, S. 275

²²¹ Vgl. Rayner/Pollatsek 1989, S. 174

²²² Vgl. Aicher 1989, S. 50ff.

²²³ Vgl. Bergner 1990, S. 20

²²⁴ Vgl. Wendt 1992, S. 272ff.

²²⁵ Vgl. Lynch/Horton 1999, S. 85f.

keln) bestanden²²⁶. Reihungen von Großbuchstaben entstellen schließlich unbestreitbar die bekannten Schriftbilder, aus denen geübte Lesende nach der „Wortbild“-Theorie ihre Wort- und Silben-Silhouetten entnehmen, zu gleichsam „monotonen Rechtecken“, bestehend dann aus der Reihung nahezu gleich großer Buchstaben und in der Gesamtheit resultierend in einem Block aus Horizontal-, Diagonal und Vertikal-Strichen, die als Ganzes keine markante Eigenform mehr zu entfalten in der Lage sind²²⁷ (der renommierte Zeitungsgestalter Norbert Küpper spricht in diesem Zusammenhang auch vom „Lattenzaun-Effekt“²²⁸).

Es existieren sogar durchaus alltagsempirische Fingerzeige, dass die Theorie der „Silhouetterkennung“ stichhaltig sein könnte. Erklärlich wäre in diesem Modell beispielsweise das verbreitete Stocken vieler Lesender bei der Rezeption des Wortes „beinhalten“: Die Kombination der ersten vier Buchstaben dieses Wortes löst das vermeintliche Erkennen des Wortes „Bein“ aus, auf einer detaillierteren Wahrnehmungs-Ebene die fälschliche Identifizierung der Graphem-Kombination „e – i“ als Lautverbindung.

Fazit: Aus vermeintlich abstrakten, damit auch potenziell beliebig reihbaren Buchstaben-Kombinationen im „alphabetischen“ System werden im Verständnis der „Wortbild“-Theorie wiederum gestalterische Gesamt-Konstrukte, die mehr repräsentieren als das gereimte Ganze der Einzelteile – der Lettern –, aus denen sie bestehen. Das heißt: Der geübte Leser erfasst auch im „Graphem – Phonem“-System eben nicht Buchstaben, sondern aus Buchstaben geformte Bilder. Die Analogie zum orientalischen Bildzeichen, Frutigers „bildhaft gebliebenen“ Sprachfixierungssystemen ist durchaus frappierend und wird von Aicher auch ausdrücklich so konstruiert: „In Wahrheit sind wir wieder zu einer Wortschrift zurückgekehrt“²²⁹.

Doch ein erweiterter Blick auf die zeitgenössisch vertretenen Theorien der Zeichenverarbeitung zeigt, dass die „Wortbild“-Theorie augenscheinlich nur noch als „Urban Legend“ existiert, nicht mehr jedoch als in der Fakultät mehrheitlich für richtig erkannte Meinung. So negierten nicht nur McClelland und Rumelhart schon 1980 ausdrücklich die These, die visuelle Wortsilhouette könne merklichen Einfluss nehmen auf die Textrezeption; unter anderem sei bei angenommener Richtigkeit der „Wortbild“-Theorie nicht mehr nachvollziehbar, warum aus Großbuchstaben konstruierte Wörter dann überhaupt noch lesbar sein sollten²³⁰. Auch Neisser erteilte der Unterstellung, „dass Wörter durch Vergleich (...) erkannt werden, wobei diese Schablone das ganze Wort wäre“²³¹, bereits 1974 eine klare Absage. Genauso hält Larson die „Wortbild“-Theorie mindestens für ein „Missverständnis“ beziehungsweise eine Fehlinterpretation gängiger Erkenntnisse und Theorien²³². Zumindest Zweifel an der empirischen

²²⁶ Vgl. Beispielhaft Tinker/Paterson 1928, S. 358ff.

²²⁷ Vgl. Lamprecht 2002, S. 38

²²⁸ Vgl. Küpper 1984, S. 24

²²⁹ Aicher 1989, S.50

²³⁰ Vgl. McClelland/Rumelhart 1980, S. 5

²³¹ Vgl. Neisser 1974, S. 67f.

²³² Vgl. Larson 2004

Haltbarkeit der „Wortbild“-Theorie artikuliert auch Gross: „Wie weit sich die Analogie zwischen Wort und Bild führen lässt, ist noch nicht geklärt“²³³.

Fest steht: Es dominieren heute augenscheinlich solche Modelle die herrschende Meinung der Fachwissenschaft, die den Leseprozess zunächst ausdrücklich von der kleinsten Einheit des Alphabets ausgehend erklären – des Buchstabens beziehungsweise der Satzzeichens nämlich – und erst von dort aus, im Folgeschritt, tendenziell ganzheitlich, das heißt: wort- oder gar satzbezogen. Doch selbst diese Modelle messen der Wort- oder Silbensilhouette als visuellem, bildhaftem Gesamt keine zentrale Bedeutung bei der Textrezeption mehr bei. Im Folgenden werden die Ansätze detaillierter beschrieben.

1.3.3 Die Dekodierung von Einzel-Zeichen

Im wesentlichen zwei unterschiedliche Theorien liegen vor bezüglich der Grundsatz-Frage, wie Einzel-Zeichen als „Kleinst-Bausteine“ des Alphabets im Gehirn verarbeitet und entschlüsselt werden: das „Schablonen-Modell“ und das Modell der „Merkmalsanalyse“. Weitgehend unumstritten scheint in beiden Ansätzen die Annahme, dass visuelle Zeichen, die über die Netzhaut aufgenommen werden, nur sehr kurzfristig in einer Kurzzeitgedächtnisleistung als getreue bildliche „Abbilder“ ins Gehirn gelangen, der eigentliche Dekodierungsprozess aber ausgesprochen unmittelbar einsetzt und sich auf abstrakterer Ebene vollzieht²³⁴.

Ich werde mich im Folgenden auf eine wesentliche Beschreibung der beiden Modelle beschränken. Auf eine erschöpfendere Diskussion der durchaus reichhaltigen empirischen Forschung, die die vorgestellten Modelle unter verschiedensten Fragestellungen auf ihre Tauglichkeit überprüft hat, verzichte ich dagegen.

1.3.3.1 Das Schablonen-Modell

Das Modell des „Schablonenabgleichs“ unterstellt, dass Lesekundige für jedes semiotische Zeichen eines Alphabets im Langzeitgedächtnis eine eigene, bildliche Schablone vorhalten²³⁵ – ganzheitlich formierte, in sich geschlossene und unseren Alphabet-Lettern entsprechende visuelle Muster also, die abgeglichen werden mit den rezipierten Schriftzeichen und bei hinreichender Identität mit diesen in Bezug gesetzt werden. In diesem Modell verfügt also jeder Alphabetisierte gleichsam über ein isoliertes „A“-Muster, ein „d“-Muster und ein „f“-Muster, um Text zu dekodieren. Es wird also „die exakte Übereinstimmung zwischen einem Muster und einem Stimulus bestimmt“²³⁶.

²³³ Gross 1994, S. 9

²³⁴ Weiland/von Gizycki 2002, S. 35ff.

²³⁵ Vgl. Neisser 1974, S. 67f.

²³⁶ Anderson 2001, S. 51

Im Grundsatz, darauf sei ausdrücklich verwiesen, ist das „Schablonenmodell“ keineswegs ein schriftzeichen-spezifisches. Vielmehr basiert das Modell auf der Grundannahme, dass wirklich sämtliche visuell wahrnehmbaren Umwelteinflüsse, eben nicht nur semiotische Zeichen, durch Abgleich mit vertrauten Schablonen verarbeitet werden²³⁷. Was durchaus Zweifel entstehen lässt an der Tauglichkeit des Modells im Ganzen: Schließlich setzt dieser Erkläransatz eine schier unübersehbare Fülle an vorgehaltenen Figurschablonen in jedem Menschen voraus. Redelius weist auf eine weitere Schwäche des Ansatzes hin: „Das Modell würde auch voraussagen, dass ein neuer Buchstabe, der zuvor nie gesehen worden ist, nicht als ein Buchstabe erkannt werden könnte“²³⁸.

Ein erweiterter Ansatz des Schablonenmodells unterstellt daher, dass die kognitiven Vergleichsmuster dem rezipierenden Menschen nicht als gleichsam ungeordnete Sammlung zur Verfügung stehen, sondern auf einer Meta-Ebene verschiedenen Klassen von Umweltreizen zuzuordnen sind; unter anderem eben dem Umweltreiz „Schriftzeichen“. Nach diesem „Prototyp-Modell“ werden Lettern also zunächst dem „Urmodell“ Zeichen zugeordnet und dann auf Grundlage des Schablonenmodells verarbeitet²³⁹. Lindsay und Norman sprechen hier auch von einem „konzeptuell-gesteuerten Prozess“: „Gemeint ist damit, dass der Verarbeitungsvorgang damit beginnt, dass man sich ein Konzept darüber bildet, um was es sich da handeln könnte“²⁴⁰. Freilich: An der Unmenge von Schablonen, die auch dieses Modell der kognitiven Wahrnehmung voraussetzt, ändert sich grundsätzlich nichts.

Doch das Schablonenmodell wirft noch weitere Fragen auf. So zeichnet sich jeder schriftkundige Mensch aus durch eine individuelle Handschrift, und Druckschriften gibt es tausende. Wie ist also zu erklären, dass beispielsweise ein „e“, handschriftlich zu Papier gebracht von zwanzig verschiedenen Menschen in notwendigerweise divergierender Gestalt, in den meisten Fällen doch als das eindeutige Graphem erkannt wird, als das es niedergeschrieben wurde?

Unterstellt man das Schablonenmodell als plausibel, sind zwei Antworten vorstellbar. Die eine: Die auf die Netzhaut projizierte Letter wird in einem dem eigentlichen Schablonenabgleich vorgeschalteten Prozess kognitiv „begradigt“, quasi einer figürlichen „Vorbehandlung“ unterzogen, in der beispielsweise perspektivische Verschrägungen korrigiert und unvollständige Linienformen komplettiert werden²⁴¹.

Eine zweite Antwort könnte lauten: Die Gestalt-Schablonen, die wir beim Erkennen von Schriftlichem zugrunde legen, sind unscharf. In diesem Verständnis existiert also – in visueller Hinsicht – kein „absolutes e“, keine endgültige „e-Schablone“, sondern es gibt nur diverse

²³⁷ Vgl. Neisser 1974, S. 88

²³⁸ Redelius 1998, S. 21

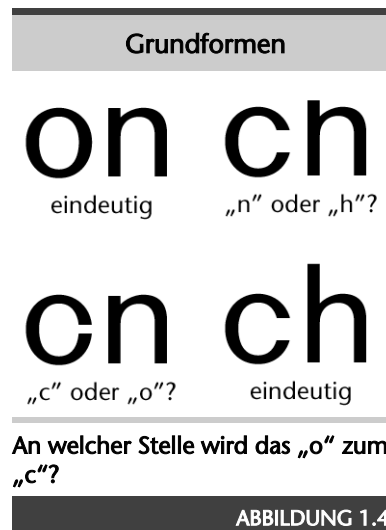
²³⁹ Vgl. Redelius 1998, S. 22

²⁴⁰ Vgl. Lindsay/Norman 1981, S. 7

²⁴¹ Vgl. Lindsay/Norman 1981, S. 3ff.

Form-Derivate einer unbestimmten Basis-Form, die wir in unserem Gehirn als „e“ dekodieren.

Die Silhouetten dieser vagen Basis-Formen müssten dann aber auf eigentümliche, schwerlich vorstellbare Weise unscharf und trenngenau zugleich ausfallen²⁴². Abbildung 1.4 verdeutlicht dies: Der Differenzierungsbereich zwischen „o“ und „c“ einerseits, „h“ und „n“ andererseits ist so eng definiert, dass eine kognitive „Schablone“ dieser Lettern eben nicht unscharf sein dürfte, sondern im Gegenteil hoch exakt formiert sein müsste. Warum sonst sollten wir beispielsweise auch in der Lage sein, in der Regel problemlos die minimalen Wesensunterschiede der Buchstaben „O“ und „Q“ zu identifizieren²⁴³?



Es sind vor allem diese Unsicherheiten, die das Schablonenmodell als Erklärmodell der Zeichenerkennung angreifbar geraten lassen.

1.3.3.2 Das Modell der Merkmalsanalyse

Einen detaillierteren Ansatz verfolgt daher, als Konkurrenz- und offenkundig auch zeitgenössische Mehrheitserklärung, das Modell der „Merkmalsanalyse“. Diesem Modell zufolge werden visuell rezipierte Schriftzeichen nicht mit mehr oder minder formäquivalenten ganzheitlichen Schablonen im Gehirn abgeglichen, sondern in einem vorgeschalteten Schritt auf die kritischen visuellen Merkmale „heruntergebrochen“, aus denen sie sich zusammensetzen; es wird also „jeder Reiz als Kombination elementarer Merkmale angesehen“²⁴⁴. Rezipierte Zeichen werden also lediglich auf markante Merkmalsaspekte und deren wechselseitiges Zusammenspiel untersucht. In diesem Sinne wird der Buchstabe „A“ nicht als ein visuelles Gesamt aufgefasst, sondern als eine Kombination dreier Linien, zweier diagonalen nämlich und einer horizontalen, die sich an definierten Punkten treffen. Es wird also nicht mehr von einer „Schablone“ pro existenter Letter ausgegangen, sondern gleichsam von „Sub-Schablonen“, sogenannten „Detektoren“²⁴⁵, aus denen Schriftzeichen zusammengesetzt sind²⁴⁶: „Durch den Rückgriff auf Merkmale anstelle von umfassenden Mustern verringert sich die Zahl der benötigten Schablonen, da nicht für jedes Muster eine eigene Schablone benötigt wird“²⁴⁷.

²⁴² Vgl. Neisser 1974, S. 67f.

²⁴³ Vgl. Neisser 1974, S. 88

²⁴⁴ Anderson 2001, S. 51

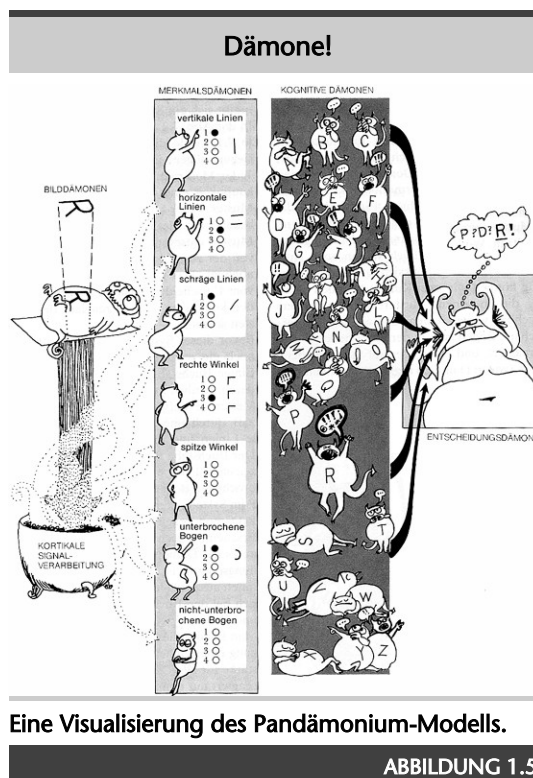
²⁴⁵ Vgl. Lindsay/Norman 1981, S. 4

²⁴⁶ Vgl. Mohr 1984, S. 51ff.

²⁴⁷ Bayer 2003, S. 37

Anzahl und Art der „kritischen Merkmale“, aus denen sich Lettern zusammensetzen können, sind dabei umstritten. Lindsay und Norman identifizieren im von ihnen beschriebenen „Pandämonium“-Modell beispielsweise sieben Teilphänomene von Buchstaben und Ziffern, die je Buchstabe in eindeutigem Mengenverhältnis zueinander stehen und damit diesen Buchstaben eindeutig identifizieren: vertikale Linien, horizontale Linien, schräge Linien, rechte Winkel, spitze Winkel, unterbrochene sowie nicht-unterbrochene Bogen.

Das „Pandämonium“-Modell entbehrt nicht eines humorvollen Beischwing-Tons. Denn es erklärt die kognitive Verarbeitung von Schriftzeichen tatsächlich unter Beschreibung imaginärer „Dämonen“, die im Gehirn die Identifikation eines Zeichens gleichsam arbeitsteilig versehen. So nimmt in diesem Modell der imaginäre „Bilddämon“ die Letter als gestalterisches Gesamt auf und übergibt dieses Gesamt an die „Merkmalsdämonen“, die das Zeichen-Bild, in einem zweiten Schritt, auf Vorhandensein und Menge der sieben oben beschriebenen Merkmale untersuchen. So würde beispielsweise der virtuelle „Dämon“ mit der Zuständigkeit für vertikale Linien derer eine in dem Buchstaben



Eine Visualisierung des Pandämonium-Modells.

ABBILDUNG 1.5

„R“ erkennen, der „Dämon“ mit dem „Fachgebiet“ „unterbrochene Bogen“ ebenfalls ein einmaliges Auftreten, der „Dämon“ mit der Sensorik für „rechte Winkel“ deren drei, schließlich der „Dämon“ mit dem „Spezialgebiet“ „schräge Linien“ wiederum eine.

Diese Informationen gehen im „Pandämonium“-Modell weiter an die „kognitiven Dämonen“, die gleichsam auf Einzelzeichen „spezialisiert“ sind und „Alarm schlagen“, sobald die vorgeschalteten „Merkmalsdämonen“ Merkmalsblöcke identifizieren, die mit „ihrem“ Buchstaben in Verbindung zu bringen sind. Im Falle des „R“ würden also wahrscheinlich die „Dämonen“ mit den Zuständigkeiten für „P“, „R“ und „D“ laut geben, da sich in sämtlichen drei Lettern diverse Eigenschaften der von den „Merkmals-Dämonen“ erkannten Blöcke wieder finden. Jedoch nur das „R“ weist alle von den Merkmals-Dämonen gemeldeten Elemente in exakter Anzahl auf. Der „Entscheidungs dämon“ als höchste Instanz wird also die finale kognitive Entscheidung zugunsten dieses Buchstabens fällen²⁴⁸.

²⁴⁸ Vgl. Lindsay/Norman 1981, S. 203ff.

Neisser gehört zu den ausdrücklichen Befürwortern des „Pandämonium“-Modells. Mehr noch, er billigt ihm eine gleichsam „panvisuelle“ Bedeutung zu, die weit über das Universum typografischer Zeichen hinausreicht: „Im Prinzip kann ein Pandämonium JEDES (Hervorhebung im Original) Muster erkennen, solange eine gewichtete Kombination der getesteten Merkmale dazu dient, Angehörige einer Kategorie von Nichtangehörigen zu unterscheiden“²⁴⁹.

Im Wesentlichen scheint das Erklärmodell der Merkmalsanalyse heute weitgehend Konsens zu sein²⁵⁰ und sogar Adaptionen in der Ingenieurwissenschaft – hier nämlich im Bereich der Konzeption und technischen Realisierung maschineller Musterauslesung – zu finden²⁵¹.

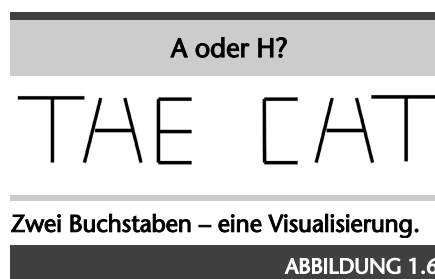
1.3.4 Zeichenrezeption und Kontextwissen

Die oben beschriebenen Erklärmodelle – Schablonenmodell und Merkmals-Erkennung – zur Verarbeitung rezipierter Zeichen gehen mehr oder minder von einem „isolierten“ Zeichenstimulus aus, der auf den Lesenden wirke. Die Fokussierung auf die Einzelfigur erscheint jedoch in dem Sinne allzu monokausal begründet, als sie die Einflussnahme weiterer Faktoren auf den Stimulus- und Dekodierungsprozess weitgehend ausblenden.

So ist davon auszugehen, dass vor allem auch das sogenannte „Kontextwissen“ im Leseprozess eine erhebliche Rolle spielt. Das bedeutet, dass nicht nur die rein semantische Zeichenfolge von Lesenden als Quelle genutzt wird, sondern intellektuelles Vorwissen sowie Teile des bereits Gelesenen den Verarbeitungsprozess mit beeinflussen. Diese These wurde maßgeblich durch McClelland und Rumelhart entwickelt und auf ihre Stimmigkeit getestet²⁵².

Wie dringlich und stichhaltig die Ausweitung des Erklärverfahrens der Zeichen-Dekodierung um einen inhaltlich-kontextualen Aspekt ist, wird bei Betrachtung des Beispiels in Abbildung 1.6 deutlich. Zwei Zeichen (das „H“ im Wort „the“ sowie das „A“ im Wort „cat“) sind hier visuell vollkommen identisch ausgeformt, dabei jedoch dennoch

weitgehend problemlos als jene differierenden Phonogramme identifizierbar, die sie sind. Sowohl „Schablonenmodell“ als auch „Merkmalsanalyse“ stoßen hier als wesentlich einzelzeichenfixierte Erkläransätze an Grenzen. Unterstellt man jedoch eine eben inhaltlich-kontextuelle „Begleitverarbeitung“ der rein sensorischen Zeichendekodierung, so ist erklärbar, warum in Abbildung 1.6 die Information „THE CAT“ entnommen werden kann: Unter anderem wird



²⁴⁹ Neisser 1974, S. 101

²⁵⁰ Vgl. Mohr 1984, S. 53

²⁵¹ Vgl. beispielhaft Walter 2000

²⁵² Vgl. McClelland/Rumelhart 1980

der geübte Lesende rasch feststellen, dass weder das Wort „TAE“ noch „CHT“ in seiner Erfahrung mit bisher Gelesenem einen weiteren Sinn ergibt²⁵³ – es sei denn als in der Tat neue Wortschöpfung, deren Auftreten in kohärenten Alltagstexten aber eindeutig die Ausnahme bildet. Die Psychologie bezeichnet dieses Phänomen als „Wortüberlegenheitseffekt“²⁵⁴ beziehungsweise spricht von einer sogenannten „Top-Down-Verarbeitung“, „weil allgemeines Wissen auf einer hohen Ebene bestimmt, wie Wahrnehmungseinheiten auf einer niedrigen Ebene interpretiert werden“²⁵⁵. Die Verarbeitung von „unten nach oben“ („bottom-up“²⁵⁶), also vom Einzelzeichen zum Wort, dann zum Satz wird hier also um eine quasi vom anderen Ende her Einfluss nehmenden Wirkmacht ergänzt. Zimbardo und Gerrig sprechen in diesem Zusammenhang von einer „Wahrnehmungs-Voreinstellung“ oder einem „perzeptuellen Set“, definiert als „die erhöhte Bereitschaft, einen bestimmten Reiz in einem gegebenen Kontext zu entdecken“²⁵⁷.

Dabei muss es nicht allein das Vorwissen eines Lesenden sein, das die „Top-Bottom-Verarbeitung“ vereinfacht, sondern auch der unmittelbare inhaltliche Kontext²⁵⁸: „So werden beim Lesen Hypothesen hinsichtlich des Textinhaltes (...) gebildet (...). Aufgrund des vorangegangenen, bereits gelesenen Abschnittes können die nächsten Wörter (...) anhand des Sinnzusammenhangs erschlossen werden“²⁵⁹. Die Reihung „Katze – Maus – Vogel – H?nd“ wird hinter dem Fragezeichen wohl ein „u“ vermuten lassen, in der Reihung „Arm – Ellenbogen – H?nd“ dagegen vermutlich eher ein „a“²⁶⁰. Hier wird also nach unmittelbarem semantischem Zusammenhang das fehlende Zeichen ersetzt.

Lindsay und Norman zufolge ist davon auszugehen, dass geübte Lesende bereits nach dem Erfassen einiger weniger Buchstaben eines Wortes eine gedankliche Vervollständigung dieser Buchstaben zu einem bekannten Wort versuchen²⁶¹ – einem Wort, das zudem in seiner Länge der Länge des vorliegenden geschriebenen Wortes näherungsweise entspricht²⁶² (diese grobe Längenbestimmung wird nach herrschender Erkenntnis beständig und rezeptionsbegleitend geleistet, zumindest von routinierten Lesenden).

Womöglich genügt zur Anstiftung dieses Vervollkommnungsprozesses sogar bereits ein einziger Buchstabe – der erste nämlich – eines Wortes, um bei Individuen die Unternehmung „Wortkomplettierung“ auszulösen. Immerhin kursierte vor einigen Monaten im Internet eine Art „Ketten-E-Mail“, die folgenden erstaunlichen Text enthielt: „Gmäeß eneir Sutide eneir elgnihcesn Uvinisterät ist es nchit witihcg, in wlecehr Rneflogheie die Bstachuebn in eneim

²⁵³ Vgl. Neisser 1974, S. 140

²⁵⁴ Vgl. Anderson 2001, S. 64f.

²⁵⁵ Anderson 2001, S. 63

²⁵⁶ Vgl. Ziefle 2002, S. 26

²⁵⁷ Zimbardo/Gerrig 2004, S. 193f.

²⁵⁸ Vgl. McClelland/Rumelhart 1980, S. 7

²⁵⁹ Ziefle 2002, S. 28

²⁶⁰ Vgl. Zimbardo/Gerrig 2004, S. 193f.

²⁶¹ Vgl. Lindsay/Norman 1981, S. 221ff.

²⁶² Vgl. Neisser 1974, S. 140ff.

Wrot snid, das ezniige, was wcthiig ist, ist dsas der estre und der leztte Bstabchue an der tihcegn Pstoiion snid. Der Rset knan ein ttoaelr Bsinöldn sien, tedztorm knan man ihn onhe Pemoblre lseen. Das ist so, wiel wir nciht jeedn Bstachuebn enzelin leesn, snderon das Wrot als gseatems.“

Nun ist der Urheber dieser E-Mail nach meinen Recherchen bislang nicht ermittelt; auch fand ich nirgendwo bislang die „englische Universität“ noch deren vermeintliche „Studie“ konkretisiert, von der im Text die Rede ist. Bemerkenswert ist und bleibt jedoch die Rezipierbarkeit des Textes, in dem kaum ein Buchstabe dort zu finden ist, wo er orthografisch hingehörte – abgesehen von Anfangs- und Endlettern eines jeden Worts. In diesem Sinne scheint es tatsächlich, dass, wie Anderson feststellt, Schriftliches „aus weitaus mehr Merkmalen (besteht), als für eine Unterscheidung des einen Stimulus von einem anderen notwendig wären. Die Wahrnehmung kann bereits erfolgreich voranschreiten, wenn einige der Merkmale erkannt sind; der Kontext ‚füllt‘ die fehlenden Merkmale auf“²⁶³. Im Übrigen, dies sei abschließend bemerkt, ist der obige Verwirrttext des unbekanntes Autors ein weiteres, sehr schlagendes Element gegen die „Wortbild-Theorie“: Denn von wieder erkennbaren Wort-Silhouetten kann hier ja kaum noch die Rede sein.

1.3.5 Sakkaden

Der Leseprozess wurde auf den vorigen Seiten vornehmlich aus kognitionspsychologischer Sicht betrachtet. Neurobiologische Aspekte der Zeichenrezeption – die Beschaffenheit des menschlichen Auges beispielsweise und die Stimulierung von Hirnrezeptoren nach visuellem Reiz – blieben ausgeklammert. Und dies soll im Wesentlichen auch so bleiben²⁶⁴.

Mit einer Ausnahme, die interessant und aufschlussreich für den hier interessierenden Prozess der Zeichenrezeption ist: die abtastende Pupillenbewegung des menschlichen Auges im Leseprozess. Das Fixieren von Text also, das nach übereinstimmender Erkenntnis nicht linear-gleichmäßig, sondern in Blick-Sprüngen erfolgt, sogenannten Sakkaden²⁶⁵. Das Auge springt also gleichsam über die Zeilen, um nach jedem dieser Sprünge in einer kurzen „Fixation“²⁶⁶ verweilen, in der das Erfasste verarbeitet wird²⁶⁷. Am Ende einer jeden Zeile erfolgt dann ein Rückschwung an den Beginn der darunter liegenden Zeile. Jede Lese-Fixation dauert laut Wendt etwa 0,25 bis 0,35 Sekunden²⁶⁸, der Sprung vom Zeilenende zu darunter liegenden Zeilenanfang etwas länger. Auch kommt es gelegentlich (in etwa 10 % aller Fälle²⁶⁹) zu Rück-

²⁶³ Anderson 2001, S. 64

²⁶⁴ Vgl. sehr ausführlich Goldstein 2002, S. 41ff.; vgl. ferner Lindsay/Norman 1981, S. 149ff.; vgl. auch Zimbardo/Gerrig 2004, S. 121ff.; vgl. zusammenfassend Wendt 1992, S. 272 sowie übersichtlich Stocksmeier 2002, S. 111; vgl. ferner Ziefle 2002, S. 42ff.

²⁶⁵ Vgl. Rehe 1981, S. 18f.

²⁶⁶ Vgl. Bergner 1990, S. 20

²⁶⁷ Vgl. Bürgel/Neumann 2001, S. 134

²⁶⁸ Vgl. Wendt 1992, S. 272

²⁶⁹ Vgl. Ziefle 2002, S. 20

sprünge, sogenannten Regressionen, wenn der Fokus quasi vergewissernd auf eine bereits übersprungene Textpassage zurückgelenkt wird²⁷⁰.

Ungeklärt ist nach wie vor, welche Faktoren in welchem Maße Tempo, Weite und Qualität dieser „Sakkaden-Sprünge“ beeinflussen. Gross zufolge herrscht inzwischen die Meinung vor, dass Wort- und Textbedeutung das Abtast-Verhalten des Auges durchaus mit bedingen: „Häufigkeit, räumliche Spannbreite und Richtung der Saccaden (so geschrieben im Original, M.L.) hängen von der formalen Struktur des Textes und vom Textverständnis ab“²⁷¹. Widerlegt scheinen jedenfalls jene älteren Modelle, die die Augenbewegung als alleinigen Ausfluss eines vom Individuum gänzlich unmanipulierbaren, motorischen „Alleingangs“ des Gehirns betrachteten („oculomotor control model“) oder lediglich die visuelle, nicht aber die inhaltlich-kontextuale Eigenschaft eines Textes als Einflussfaktor beschrieben („cognitive process model“)²⁷². Eine einzige, eindeutige und individuumsunabhängige Sakkadenbreite zumindest existiert offenkundig nicht.

1.4 Typografie im alphabetischen System

Aus den vorangegangenen Ausführungen ist deutlich geworden: Der kognitive Prozess des „Lesens“ ist alles andere als auserklärt. Derzeit scheint das Modell der Merkmalsanalyse in Verbindung mit der Annahme einer intellektuellen „Top-Bottom“-Verarbeitung verschriftlichter Information beherrschende Grundlage theoretischer Erwägungen in der Fachwissenschaft zu sein. Ungeklärt ist nicht zuletzt – und das ist bedeutsam für den Kontext der vorliegenden Arbeit –, ob sich die beschriebenen Modelle der Zeichenrezeption unmittelbar auf das Lesen von selbststrahlenden Bildschirmen übertragen lassen. Mangels einschlägiger Quellen unterstelle ich dies im Folgenden jedoch.

Was allen diskutierten Modellen völlig abgeht, ist die Einbeziehung des Faktors „Schriftgestalt“ jenseits der basalen Konstruktionseigenheiten einer Letter (zum Beispiel der, dass ein „E“ aus einem Vertikal- und drei Horizontalstrichen besteht). Detail-Typografie ist kein Gegenstand der Zeichenrezeptions-Erklärung. Dies erstaunt beispielsweise und nicht zuletzt deshalb, als gerade der häufige Buchstabenbestandteil „Serife“ in klassischen Druckschriften den zentralen, rezeptionsrelevanten Merkmalen eines Buchstabens zur Seite tritt: Warum werden Serifen als der Zierat erkannt, der sie sind? Warum treten sie offenkundig nicht in visuellen Konflikt mit den basalen visuellen Markmalen des Buchstabens? Ziefle hebt ausdrücklich hervor: „Trotz der mehr als hundertjährigen Forschung zur Lesbarkeit von Druckschriften ist bisher kein die Einzelbefunde übergreifendes theoretisches Konzept der Lesbarkeit entstanden, das die Wirkung typographischer Faktoren (...) in die Modellierung des Lesevor-

²⁷⁰ Vgl. Luidl 1989, S. 93

²⁷¹ Gross 1994, S. 8f.

²⁷² Vgl. Gross 1994, S. 8f.

gangs integriert hat“²⁷³. Auch Bonsiepe beklagt, dass „in der Regel der Typographie in der Sprachwissenschaft kaum Beachtung geschenkt wird“²⁷⁴. Noch 2004 beklagte Stöckl ausdrücklich, dass die visuelle Textgestalt als, wie er meint, integrales Element des Text-Publizierens über den Prozess des inhaltlichen Verfassens hinaus kaum Beachtung finde in den linguistischen Fakultäten²⁷⁵. Warum dies ein erstaunliches Defizit darstellt insbesondere in Zeiten, da fast jeder Mensch druckfähige Werke zu erstellen imstande ist, wird weiter unten insbesondere im Zusammenhang mit dem typografischen Phänomen „Schriftart“ noch ausführlicher zur Sprache kommen. Vorläufig bleibt nicht mehr, als diese Tatsache als solche hier festzustellen.

²⁷³ Ziefle 2002, S. 27

²⁷⁴ Bonsiepe 1996, S. 80

²⁷⁵ Vgl. Stöckl 2004, S. 5ff.

2 Lesen vom Papier, Lesen am Bildschirm

Die herausstechende Besonderheit von Text in Webbrowsern ist, wie gezeigt, dass sie überwiegend im Status der Virtualität verharren. Sie werden meist am Monitor gelesen und nicht in Form eines Drucks materialisiert. Doch induziert der Unterschied der Plattformen auch Unterschiede des Rezeptionsprozesses?

Es existieren durchaus Indizien dafür, dass Lesen am Bildschirm qualitativ verschieden ist vom klassischen Leseprozess von Papier. Indizien, wohlgemerkt, Anhaltspunkte, die aus sich selbst heraus noch keine qualitative oder gar quantitative Abgrenzung der Medien „Bildschirm“ und „Papier“ generieren, sondern höchstens Vermutungen, Thesen. Diese Anmerkung ist wichtig, denn verwertbare empirische Befunde sind rar und auch widersprüchlich zu der Frage: Liest man – zunächst allgemein und an dieser Stelle vorläufig ohne Einbeziehung des Faktors „Typografie“ – am Bildschirm „schlechter“ als von Papier?

2.1 Besonderheiten des Bildschirm-Lesens

2.1.1 Leseabstand

Der Abstand zwischen Auge und Medium beeinflusst nach Ansicht diverser Autoren das Leseverhalten. Während man bei gedruckten Medien Köhler zufolge von „ca. 30 bis 50 cm“²⁷⁶ Leseabstand auszugehen hat, sind es am Monitor eher „45 – 55 cm“²⁷⁷. Kiehn und Tietzmann gehen gar von „40 bis 70 Zentimeter(n)“²⁷⁸ Leseabstand an Monitoren aus. Allerdings ist davon auszugehen, dass diese Größenordnungen sich in erster Linie auf stationäre Rechner mit separatem Monitor beziehen; an Notebooks dürfte die durchschnittliche Distanz kleiner ausfallen, da an mobilen Geräten Tastatur und Monitorfläche sehr nahe beieinander liegen – und wer betätigt eine Tastatur schon aus einem Abstand von 40 Zentimetern?

Doch zumindest einen ersten Schluss legen die obigen Annahmen nahe: Größere Lesedistanz erfordert vermutlich größere Schrift, um in der finalen Rezeption gleichwertig aufgenommen zu werden. Schriften am Monitor müssen in dieser Logik also größer ausfallen als auf Papier. Eine These, die zu überprüfen sein wird.

2.1.2 Flimmern und Blendungen

Bildschirme sind Selbststrahler. Sie flimmern. Und das Auge nimmt, – wenn auch in der Regel unbewusst – laut Köhler das stetige Flimmern des Bildschirms wahr²⁷⁹. Dies gilt für alle

²⁷⁶ Köhler 2002, S. 260

²⁷⁷ Bürgel/Neumann 2001, S. 109

²⁷⁸ Kiehn/Tietzmann 1998, S. 43

²⁷⁹ Erben 1998, S. 24

Monitor-Bauarten, wenn auch in deutlich eingeschränkterem Maße für TFT-, vulgo: Flachbildschirme, die sich zunehmend am Markt durchzusetzen scheinen²⁸⁰. Es spricht also einiges dafür, dass Lesen am Bildschirm die Augen des Betrachters einer höheren physischen Beanspruchung aussetzt und daher der Rezeptionsprozess tendenziell ermüdender ausfällt als das Lesen von Papier²⁸¹.

Ein weiterer Faktor, der das Lesen am Bildschirm potenziell erschweren kann, sind Blendungserscheinungen durch reflektierendes Umgebungslicht an der Bildschirmoberfläche²⁸². Auch dieses Phänomen reduziert sich an Flachbildschirmen merklich; ausgeschaltet ist es jedoch nicht.

2.1.3 Sitzhaltung und Betrachtungswinkel

Thissen weist darauf hin, „ dass Bildschirmarbeiter dazu neigen, eine recht starre Sitzhaltung einzunehmen und auf den Monitor zu ‚stieren‘, d.h. weniger (die Augen befeuchtende) Lidschläge zu tun“²⁸³. Auch Puscher äußert sich zur anatomischen Komponente des Bildschirmlesens: „Der natürliche Blickwinkel des Menschen in einer entspannten Sitzhaltung richtet sich mit ungefähr 45 Grad nach unten. Die meisten Computermonitore stehen auf Augenhöhe oder sogar darüber. Einzig am Notebook wird eine für die Wirbelsäule entspannende Haltung erreicht“²⁸⁴. Diese Argumente legen durchaus den Schluss nahe, dass Textarbeit am Bildschirm deutlich stärker ermattet als Lesen von Papier.

2.1.4 Monitorauflösung

Ein wichtiges und oft genanntes Unterscheidungskriterium zwischen Papier- und Online-Text ist die Auflösungsqualität. Text an Bildschirmen ist meist aus erheblich weniger Bildpunkten konstruiert als gedruckter Text (dieser Aspekt wird später noch ausführlich zu behandeln sein). Für Heitz, Täschler und Blum steht jedenfalls fest: „Wegen der Darstellungsweise der Schrift auf dem Monitor werden längere Texte unleserlich und ermüden schnell die Augen“²⁸⁵. Ebenfalls auf die geringe Auflösungsqualität führen Kiehn und Tietzmann zurück, dass die „Augenbewegung (...) öfter innehalten (muss), um z.B. kritische Buchstabenpaare, deren Darstellung durch die geringere (Bildschirm-)Auflösung leidet, eindeutig zu erkennen und voneinander zu trennen (z.B. ‚il‘)“²⁸⁶.

²⁸⁰ Vgl. Thissen 2003, S. 92

²⁸¹ Vgl. Ziefle 2002, S. 55ff.

²⁸² Kiehn/Tietzmann 1998, S. 28

²⁸³ Thissen 2003, S. 93

²⁸⁴ Puscher 2001, S. 37

²⁸⁵ Heitz et al. 1998, S. 28

²⁸⁶ Kiehn/Tietzmann 1998, S. 28

2.1.5 Weiteres Vorgehen

Es liegen also durchaus mannigfaltige Indizien vor, dass Lesen am Bildschirm nicht nur eine andere, sondern vor allem eine mühsamere, stärker ermüdende Tätigkeit ist als das „klassische“ Lesen vom Papier. Ist dies tatsächlich so? Die empirische Erkenntnislage ist uneinheitlich. Sie wird im Folgenden beschrieben.

2.2 Die vermeintliche Entdeckung der Langsamkeit

Die Faustregel hält sich hartnäckig. Thissen ist nur einer von Vielen, der – leider streng referenzfrei – festhält: „Studien haben gezeigt, dass die Lesegeschwindigkeit am Bildschirm 25–30% langsamer ist und dass der Leser unkonzentrierter und ungenauer liest“²⁸⁷. Auch Hellbusch warnt, die Lesegeschwindigkeit nehme „im Vergleich zu einem Buch um 25% ab“²⁸⁸, auch Nielsen behauptet eine Verlangsamung um 25 Prozent²⁸⁹. Wirth dagegen tritt in dieser Frage in interessante Gegnerschaft zu sich selbst. Denn in derselben Publikation, in der er – ganz im Sinne von Thissen und Hellbusch – verkündet, dass „Lesen am Bildschirm (...) im Vergleich ca. 25% länger (dauert), (...) anstrengender für die Augen“²⁹⁰ sei, fällt er sich selbst in den Rücken, indem er 120 Seiten später unter der Überschrift „Fakten“ doziert, die „vielerorts noch in Granit gemeißelte Weisheit, die Lesegeschwindigkeit sei am Bildschirm gegenüber Papier um 25 – 30% verlängert, ist (...) nicht mehr haltbar“²⁹¹.

Was also stimmt denn nun? Ich habe in meinen Recherchen den Eindruck gewonnen, dass ein gewichtiger Grund für die Nichterwähnung von Quellen zur Stützung dieses „25-Prozent-Mythos“ in den Quellen einen simplen Grund hat: Die vermeintlich argumentationsstützenden Studien sind – der Ausdruck sei gestattet – uralt.

2.2.1 Die Studien der 80er und frühen 90er Jahre

In meinen Recherchen nach den Ursprüngen der „Verlangsamungstheorie“ stieß ich auf keine Studien neueren Datums, die die „25-Prozent-These“ untermauern würden. Studien neueren Datums, wohlgemerkt: Oder sollte es tatsächlich wahr sein, dass die oben zitierten Autoren in ihren Feststellungen auf jene Studie von Muter, Latrémouille, Treurniet und Beam referenzieren, die in der Tat eine Verlangsamung der Geschwindigkeit von Computer-Text (projiziert auf einen Fernseher!) von etwa 29 Prozent gegenüber dem Lesen von Papier feststellten (und eine um 30 Prozent verschlechterte Fehlerauffindungs-Leistung)²⁹² – und zwar

²⁸⁷ Thissen 2003, S. 92

²⁸⁸ Hellbusch 2005, S. 19

²⁸⁹ Vgl. Nielsen 1997a

²⁹⁰ Wirth 2004, S. 87

²⁹¹ Wirth 2004, S. 208

²⁹² Vgl. Muter et al. 1982, S. 501ff.

im Jahre 1982, dem Erfolgjahr des Commodore 64? Als Texte am Computer höchst grob aufgelöst erschienen und der Begriff „Website“ noch längst nicht kreiert war? Und wenn dies so ist: Warum dominiert eine Studie nach wie vor die Literatur, die tatsächlich weiße Schrift auf blauem TV-Bildschirmgrund in Konkurrenz setzte zu schwarzer Schrift auf weißem Papier?

Ähnlich wenig anwendbar auf das Erscheinungsbild im Allgemeinen und die Typografie heutiger Webseiten im Besonderen dürfte aus ähnlichem Grunde die Studie von Gould und Grischkowsky aus dem Jahre 1984 sein: Die Forscher konfrontierten ihre Versuchsteilnehmer mit hellgrünem Text auf dunkelgrauem Grund²⁹³ und verglichen die Rezeptions-Messergebnisse mit denen, die sie für vergleichbare Texte in schwarzer Schrift auf weißem Papier fanden; dass die Forscher Unterschiede feststellten, überrascht angesichts derart wenig vergleichbarer Materialien kaum. Dass diese Unterschiede in den Lesegeschwindigkeiten deutlich unter 20 Prozent, auf jeden Fall aber nicht im Bereich jener ominösen „25 bis 30 Prozent“ lagen, sei dennoch erwähnt. Ebenfalls wenig verwertbar erscheint die Studie von Wilkinson und Robinshaw aus 1987²⁹⁴: Die Forscher präsentierten ihre Bildschirmtexte ebenfalls in „hellgrün auf dunkelgrau“; die Unterschiede zwischen Papier und Monitor lagen wiederum etwa bei 20 Prozent zugunsten des Papiers.

Ebenfalls wohl nicht gemeint sein von den „25-Prozent-Dozenten“ dürfte definitiv die Studie von Creed, Dennis und Newstead von 1988²⁹⁵, in der die Forscher – wiederum an Monitoren, die mit ihrer grünen Schrift auf dunkelgrauen Grunde heutigen kaum vergleichbar sein dürften – fanden, dass Korrekturlesen vom Bildschirm etwa 9 Prozent länger dauerte als dieselbe Verrichtung von Papier, und dass 10 Prozent mehr Fehler am Monitor übersehen wurden als in der „Hard Copy“.

Und warum wird offenbar gerne jene Übersichts-Studie von Dillon, McKnight und Richardson aus dem Jahre 1988 unterschlagen, die ergab, dass „es trotz vieler jüngster Befunde so scheint, als das Lesen von Monitoren so schnell und so genau ausfallen kann wie von Papier“²⁹⁶? Warum findet sich seltenst ein Verweis auf Osborne und Holton, die im selben Jahr in einem empirischen Experiment zum selben Schluss gelangten²⁹⁷? Und was ist mit der empirischen Untersuchung von Muter and Maurutto, die mit dem Erscheinungsdatum 1991 zwar inzwischen auch einen gewissen Reifegrad erlangt hat, aber unter anderem in derselben, nichtsdestoweniger bemerkenswerten Erkenntnis mündet, dass „das Lesen von Computerbildschirmen, die gemeinhin im Jahre 1991 erhältlich sind, äquivalent ausfallen kann zum Lesen aus Büchern, was Geschwindigkeit und Verständnisleistung betrifft“²⁹⁸? Immerhin tes-

²⁹³ Vgl. Gould/Grischkowsky 1986, S. 323ff.

²⁹⁴ Vgl. Wilkinson/Rubinshaw 1987, S. 125ff.

²⁹⁵ Vgl. Creed et al. 1988, S. 467ff.

²⁹⁶ Vgl. Dillon et al. 1988, S. 457ff.

²⁹⁷ Vgl. Osborne/Holton 1988, S. 1ff.

²⁹⁸ Muter/Maurutto 1991, S. 264 (Übersetzung M.L.)

teten Muter und Maurutto mit als erste tatsächlich schwarze Schrift auf weißem Bildschirm im Abgleich mit schwarzer Schrift auf weißem Papier! Es erstaunt, dass diese Erkenntnis – anders als offensichtlich jene der „25 Prozent“ – weit weniger von Autor zu Autor wechselseitig abgeschrieben wird. Es sei allerdings nicht unterschlagen, dass Muter und Maurutto im erwähnten Experiment auch herausfanden, dass das „Überfliegen“ von Text am Bildschirm deutlich langsamer (um 20 Prozent) und inhaltlich und in Bezug auf die Sinnentnahme zumindest etwas weniger effektiv (etwa 10 Prozent) auszufallen scheint als auf Papier.

Zwischenfazit: Ich habe keine Studie vorgefunden, die zumindest aus der Mitte der 1990er Jahre stammt – also der Frühphase des World Wide Web –, geschweige denn aus der Zeit nach der Jahrtausendwende, die belegte, dass die Verlangsamung der Leseleistung an Monitoren, verglichen mit Papierdrucken, wesentlich ausfiel. Jene ominösen „25 bis 30 Prozent“ Leseverlangsamung im Vergleich von Print und Bildschirmarbeitsleistung jedenfalls fand ich in dieser nachvollziehbaren Form nur bei Muter et al. 1982 wieder. Sollte es aber tatsächlich stimmen, dass die Legende von der „25-Prozent-Verlangsamung“ am Bildschirm nicht nur auf das Jahr 1982 zurückgeht, sondern sich sogar auf eine Studie bezieht, die an einem Fernseh-Bildschirm unter Darbietung von weißer Schrift auf blauem Grund gewonnen wurde, muss bezüglich der Seriosität vieler zeitgenössischer „Ratgeber“ zum Webdesign erheblicher Zweifel angemeldet werden.

Insgesamt muss angezweifelt werden, dass insbesondere die Studien der 80er Jahre in ihren Ergebnissen noch auf moderne Bildschirme anwendbar sind – geschweige denn auf Schrift-darstellung in zeitgenössischen Webbrowsern²⁹⁹. Es empfiehlt sich ein Blick auf empirische Erkenntnisse moderneren Datums.

2.2.2 Studien aus der „Internet-Epoche“

Das Interesse am direkten empirischen Abgleich der Leistungsfähigkeit von Print- und Bildschirmpräsentation in Bezug auf Lesegeschwindigkeit und Leseverständnis scheint mit der Durchsetzung des Heim-PCs und auch des Internets nicht etwa zugenommen haben. Tatsächlich stieß ich auf nicht einmal eine Handvoll einschlägiger Studien aus der Zeit nach 1991, dem Jahr jener angesprochenen Studie von Muter und Maurutto, in der die Forscher gar eine Gleichwertigkeit der Medien „Bildschirm“ und „Papier“ feststellten. Auch Ziefle fand in ihrer erschöpfenden Diskussion des Literaturstands keine relevanten Ergebnisse nach 1991³⁰⁰ (weswegen sie eigene, weiter unten darzustellende Untersuchungen anstellte).

Wie ist dieses frappierende Nachlassen des Forschungsinteresses zu erklären – just in einer Epoche, da sich mit *Windows 3.11* die erste grafische Benutzeroberfläche als nicht nur wirk-

²⁹⁹ Dyson 2004, S. 378

³⁰⁰ Vgl. Ziefle 2002, S. 73ff., 124ff. und 152ff.

lich massentauglich, sondern auch als kommerzieller Erfolg erwies? In einer Ära zudem, in der das Internet begann, nicht mehr nur verschworene Gemeinschaft von Computertüftlern zu sein, sondern Massenmedium? Eventuell ist schlicht davon auszugehen, dass die Grundfrage „Papier oder Bildschirm“ mit den Studien der 1980er Jahre schlicht als geklärt galt und sich das Forschungsinteresse neuen Problemen zuwandte, die sich beispielsweise auftraten durch die Potenzierung verfügbarer Schriftarten am Bildschirm, durch die zunehmende Marktdurchdringung des Betriebssystems *Windows* oder eben den beginnenden Siegeszug des Internets.

2.2.2.1 Wissenserwerb – online und von Papier

Eine Studie mit – gelinde gesagt – eigenwilliger Methodik legten O’Hara und Sellen 1997 vor³⁰¹. Die Autorinnen konfrontierten 10 Versuchspersonen mit einem „vier Seiten langen“ Text aus einer wissenschaftlichen Publikation. Fünf Teilnehmer erhielten den Text auf Papier vorgelegt, fünf andere wurden mit demselben Text in Bildschirmaufbereitung konfrontiert. Im Anschluss an den Lesevorgang wurden die Probanden aufgefordert, eine 200 bis 300 Wörter lange Zusammenfassung des Textinhalts anzufertigen. O’Hara und Sellen stellten fest, dass das Platzieren von Notizen sowie das Markieren zentraler Textpassagen den Print-Lesern leichter fiel, dass sich die Print-Leser zudem „sicherer“ im Textgesamt bewegten und die Online-Leser mehrheitlich ausdrücklich eine Papiervorlage wünschten, da sie die Aufgabe auf diesem Medium besser zu lösen meinten. Ziel der Studie war nach Angaben der Autorinnen explizit, die in ihren Augen nicht verlässlichen Funde früherer Studien zu widerlegen, Texte auf Papier und am Monitor seien gleich gut lesbar. Dieses Ziel erreichten sie; ob die eingesetzte Methodik freilich wirklich eine „reale Lesesituation“ simulierte, die eine Verallgemeinerbarkeit der Ergebnisse rechtfertigte, wie O’Hara und Sellen behaupten, muss angezweifelt werden; Leseprozesse unter der erwarteten Anschluss-Anforderung, Kurzzusammenfassungen des Rezipierten anzufertigen, sind wohl maßgeblich im schulischen Kontext üblich, aber nicht im alltäglichen informativen Lesen im Web. Insgesamt maßen die Forscherinnen ausschließlich qualitativ; belastbare numerische Daten (zum Beispiel in Bezug auf die Lesegeschwindigkeiten) erhoben sie nicht.

2.2.2.2 Papier und Bildschirm – eine vergleichende Studie

Weisenmiller kommt das Veienst zu, 1999 eine recht sorgfältig konzipierte Studie durchgeführt zu haben, in der er Lesegeschwindigkeiten und Leseverständnis sowohl am Bildschirm als auch auf Papier testete³⁰².

³⁰¹ Vgl. O’Hara/Sellen 1997

³⁰² Vgl. Weisenmiller 1999

Weisenmiller legte insgesamt 264 Probanden in 12 Gruppen à 22 Personen einen Text von 601 Wörtern vor; je ein Viertel der Teilnehmer bekam den Text in einer der vier Schriftarten *Georgia*, *Times*, *Verdana* oder *Arial* zu lesen. Die 66 Teilnehmer einer jeden „Schriftart-Gruppe“ waren wiederum gedrittelt und aufgeteilt in eine Gruppe, die den Text als Papierausdruck in 600 dpi Auflösungsqualität vorgelegt bekam, eine weitere, die den Text am Bildschirm in ungeglätteter Form las (Weisenmiller spricht hier von „1-Bit-Qualität“) und eine dritte, die den Text am Monitor in geglätteter (bei Weisenmiller „8-Bit-Qualität“) zu rezipieren hatte.

Die „Bit“-Angaben bezogen sich hierbei auf die Anzahl der Farben, die zur Darstellung der Buchstaben am Monitor zugelassen wurden. Buchstaben in „1-Bit-Qualität“ wurden dabei ausschließlich aus schwarzen und weißen Pixeln konstruiert (siehe Abb. 2.1 oben), während in den „8-Bit“-Darstellungen auch Graustufungen zur Herausarbeitung von Unterschieden in den Strichstärken innerhalb der Buchstaben zur Verfügung standen (Abb. 2.1 unten). Allerdings zeigt das nebenstehende Beispiel auch, dass Lankau Recht hat: Lettern im Anti-Alias-Darstellungsmodus wirken heller, „da die Zwischentöne zu einem anderen optischen Farbwert (...) der Gesamtform führen“³⁰³.

Da Weisenmillers Studie hier zunächst vor allem im Kontext des intermedialen Vergleichs „Bildschirm – Papier“ vorgestellt wird, werde ich zunächst nicht weiter auf die Besonderheiten der Bildschirmdarstellung von Buchstaben und auf das Phänomen der virtuellen „Schriftglättung“ eingehen; dies geschieht ausführlich in Kapitel 3.4.2.2. Es sei jedoch darauf hingewiesen, dass der absolut überwiegende Anteil zeitgenössischer Webseiten im „1-Bit-Modus“ konsumiert werden dürfte. Dies könnte sich allerdings mit der Einführung des neuen *Windows*-Betriebssystems *Vista* im kommenden Jahr grundstürzend wandeln – denn es deutet einiges darauf hin, dass mit dem neuen Betriebssystem der Einsatz des Schrift-Glättungssystem „ClearType“ zur systemischen Voreinstellung und damit absehbar auch zum weltweiten Quasi-Standard werden wird (auch dazu mehr im Kapitel „Schriftart“).

Die Studie erfolgte nach dem Modell des Nelson-Denny-Tests. Das hieß: Die Versuchsteilnehmer wurden gebeten, den ihnen vorliegenden Text zu lesen; nach einer Minute wurden sie aufgefordert, das letzte von ihnen gelesene Wort zu markieren. Danach wurden die Teilnehmer gebeten, den Text weiter zu lesen; dafür blieben jeweils weitere 19 Minuten freie Le-

1-Bit- und 8-Bit-Darstellung



Times One-Bit Text
Times Eight-Bit Text

Diesen Vergleich von 1-Bit- und 8-Bit-Darstellung legt Weisenmiller im Anhang seiner Dissertation vor.

ABBILDUNG 2.1

³⁰³ Lankau 2000, S. 283

sezeit. Im Anschluss wurden die Probanden gebeten, auf Fragebögen Verständnisfragen zum Text zu beantworten.

Die höchste Lesegeschwindigkeit stellte Weisenmiller bei Probanden fest, die den Text am Bildschirm in einer 8-Bit-*Arial* gelesen hatten (239,82 gelesene Wörter), es folgten mit merklichem Abstand die „8-Bit-*Verdana*“ (221,23 Wörter) und die „8-Bit-*Times*“ (219,14 Wörter). Die schlechtesten Werte erzielten die „1-Bit-*Verdana*“ (187,82 Wörter), nur knapp schneller wurden die „1-Bit-*Arial*“ (188,50) und die *Arial* im Papierausdruck rezipiert (190,45 Wörter).

Einen signifikanten Einfluss auf die Lesegeschwindigkeiten insgesamt konnte Weisenmiller in der Tat der Darbietungsform des Textes zuschreiben – dieser ergab sich hauptsächlich zwischen den beiden Formen der Bildschirmpräsentation (1-Bit- und 8-Bit-Aufbereitung) sowie zwischen 1-Bit-Bildschirm- und Papierdarbietung. Die Schriften im 1-Bit-Modus erwiesen sich dabei als die merklich am schlechtesten lesbaren. Keinen signifikanten Unterschied fand Weisenmiller zwischen 8-Bit-Bildschirm- und Papierpräsentation.

Die Messung des Textverständnisses (das in einem Punktesystem operationalisiert war, dessen Detail-Konstruktion hier nicht weiter erläutert werden soll) ergab ein sehr ähnliches Bild wie die der Lesegeschwindigkeit. Wiederum schnitt die „8-Bit-*Arial*“ am besten ab (28,591 von 36 erreichbaren Punkten), wiederum gefolgt von der „8-Bit-*Verdana*“ (28,000). Das drittbeste Ergebnis erzielte die *Verdana* in Papierdarbietung (27,455). „Schlusslichter“ waren die „1-Bit-*Georgia*“ (23,000), die „1-Bit-*Verdana*“ (23,273) und die „1-Bit-*Arial*“ (23,818).

Auch diese Messung ergab keine signifikanten Differenzen zwischen den einzelnen Schriftarten; wohl aber wiederum Unterschiede zwischen den Darbietungsformen. Und zwar dieselben wie bei der Geschwindigkeitsmessung: zwischen 1-Bit- und 8-Bit-Darstellung einerseits sowie 1-Bit- und Papierdarstellung andererseits.

2.2.2.3 Ziefles Studien

Martina Ziefle führt kurz vor und um die Jahrtausendwende verschiedene Experimente durch, in denen sie teilweise auch den Medienvergleich „Papier-Bildschirm“ thematisierte. In einem Experiment, in dem sie ihre 16 Probanden mit schwarzen Texten unterschiedlicher Kontraststärke zum jeweiligen Weißgrund konfrontierte, fand sie, dass die Lesegeschwindigkeit mit zunehmender Kontrastierung merklich zunahm, und zwar auf Papier wie am Bildschirm gleichermaßen – zwischen den beiden Darbietungsmedien allerdings fand sie keinen signifikanten Unterschied. Merkliche Differenzen fand Ziefle dagegen, als sie die Genauigkeit verglich, in der die Texte bearbeitet wurden. Die Teilnehmenden hatten orthografisch fehlerhafte Wörter in den Texten zu identifizieren; auf Papieridentifizierten die Probanden im Schnitt um die 90 Prozent der Fehler, am Bildschirm lediglich 78 Prozent³⁰⁴. Mit wachsender

³⁰⁴ Vgl. Ziefle 2002, S. 94ff.

Kontraststärke näherten sich allerdings auch diese Werte einander an; in einem Folgeexperiment konnte Ziefle zudem praktisch keinen eindeutigen Unterschied in der Lesegenauigkeit mehr feststellen zwischen elektronischem und textilem Medium. Ziefle formuliert es so: „Bei keiner (...) Einzelkontrastberechnung konnte ein Bildschirmvorteil gefunden werden“³⁰⁵. Man könnte es auch anders herum formulieren: Eine Papierüberlegenheit wurde nicht festgestellt.

Etwas deutlichere Unterschiede ergaben sich in einem weiteren Experiment, in dem Ziefle mit 20 Probanden die Leseleistung von Papier abglich mit der Leseleistung an einem Bildschirm, der eine Auflösung von 60 dpi aufwies, und einem weiteren Monitor mit einer Auflösung von 120 dpi. Selbst am hochauflösenden Bildschirm fiel die Lesegeschwindigkeit um gut 10 Prozent langsamer aus als von der Papiervorlage, am niedriger aufgelösten Schirm waren es sogar runde 15 Prozent. Die Papiervorlage erwies sich damit in diesem Experiment als signifikant überlegen gegenüber beiden Monitorarten.

Was die Lesegenauigkeit anging (wiederum gemessen anhand identifizierter Orthografie-Fehler), bestätigte sich das Ergebnis; wiederum schnitten die Leistungen vom Papier am besten ab, und zwar um etwa 6 Prozent (und damit signifikant) besser als die Ergebnisse, die am hochauflösenden Bildschirm erzielt wurden, und um etwa 11,5 Prozent (und damit ebenfalls signifikant) besser als am niedrig aufgelösten Monitor. Der Unterschied zwischen den Monitorarten erwies sich als statistisch nicht bedeutsam. In einer Nachbefragung erklärten 16 der 20 Teilnehmer, dass sie die Lesearbeit von Papier als die angenehmste der drei gestesteten Varianten empfanden.

2.2.3 Würdigung der Ergebnisse: Papier versus Bildschirm

Was ist aus den vorgenannten Studien zu schließen? Zumindest dies: Die Legende von der um 30 Prozent verlangsamten Lesearbeit am Bildschirm ist fast ein Vierteljahrhundert alt, wurde mit Negativschrift auf einem Fernsehbildschirm getestet und gehört endgültig zu den Akten gelegt. Die moderneren Erhebungen an eben auch technisch höher entwickelten Bildschirmen, die die Lesbarkeit von Texten auf Papier und vom Monitor verglichen, fanden in Bezug auf die Lesegeschwindigkeiten in beiden Medienformen keine oder Unterschiede von 10 bis 15 Prozent zugunsten der Papierdarbietung. Die Lesegenauigkeit scheint, wenn überhaupt, bei der Lesearbeit an Papier ebenfalls um 5 bis 10 Prozent verbessert gegenüber Monitorarstellungen.

³⁰⁵ Ziefle 2002, S. 121

2.3 Der Leseprozess – Scannen, Skimmen, Lesen

„They don’t“: ein Satz, der unter Webdesignern inzwischen nahezu Kultstatus besitzt. Es sind dies die einleitenden Worte zu Jakob Niensens Aufsatz „Wie User im Web lesen“, publiziert 1997 in der berühmten Web-Kolumne „alterbox“ des Usability-Vorfechters³⁰⁶, und der Autor fährt fort mit den Worten: „Menschen lesen kaum Wort für Wort auf Webseiten; stattdessen scannen sie die Seite, dabei einzelne Wörter und Sätze herauspickend“.

Folgt also das Lesen im Web ganz allgemein anderen Gesetzen als der Textkonsum von Papier? Bedingt der Wechsel des Mediums also nicht nur eine eventuell leichte Verlangsamung der Lesegeschwindigkeit, sondern viel mehr, eine Mutation des Rezeptionsprozesses an sich nämlich?

Für Wirth steht zumindest fest, dass die Lesemotivation am Monitor sinkt: „Der Anteil der spontan und freiwillig lesenden Internet-Nutzer ist (...) sehr gering“³⁰⁷. Lankau hält dagegen: „Unterschätzen Sie die Lesefreude (...) der Anwender nicht. Es sind im Netz bei weitem nicht nur faule ‚Freaks‘ und ‚Kids‘ unterwegs, die bei allem, was länger als zwei Zeilen ist, abschalten oder sich wegklicken. Bei optisch und inhaltlich gut strukturierten Texten dürfen Sie durchaus mit einem Publikum mit Durchhaltevermögen rechnen“³⁰⁸.

Sicherlich wäre es verfehlt, von einem gleichsam universellen „Lesestil“ auszugehen, der noch dazu ausschließlich durch das Medium induziert würde. Welcher Lesestil zur Anwendung kommt, dürfte in hohem Maße determiniert sein

- von der Qualität der angebotenen Information,
- dem inhaltlichen Interesse beziehungsweise der Motivation der Lesenden,
- der zur Verfügung stehenden Zeit und
- individuellen Lesegewohnheiten.

Dennoch scheint weitgehender Konsens, dass Leseprozesse sehr häufig nach einem mehrstufigen, verallgemeinerbaren Muster von aufeinander aufbauenden Rezeptionsphasen unterschiedlicher Qualität und Intensität entstehen:

- Scanning (grobes Text-„Abtasten“),
- Skimming (Text-Überfliegen),
- verstehendes Linear-Lesen,
- reflektierendes Linear-Lesen³⁰⁹.

Nicht jeder Rezeptionsprozess führt dabei zwangsläufig über alle Stufen; viele Lesungen enden bereits in einer der Vorstufen. Zudem ist die Terminologie keineswegs Konsens: Spool et

³⁰⁶ Vgl. Nielsen 1997b, (Übersetzung M.L.)

³⁰⁷ Wirth 2004, S. 207

³⁰⁸ Lankau 2000, S. 372

³⁰⁹ Vgl. George 1995, S.62f.

al. erblicken im „Skimming“ offenkundig keine andere Tätigkeit als im „Scanning“; bei diesen Autoren gelten beide Begriffe scheinbar als Synonyme³¹⁰, ebenso wie Alkan³¹¹.

Wie dem auch sei: Das „Scanning“ kann allgemein definiert werden als eine Phase „schwebender Aufmerksamkeit“³¹²; der Betrachter gleitet in dieser ersten Phase über die Inhalte einer Seite, die grob überflogen werde³¹³. Von einem eigentlichen „Lesen“ kann hier noch keine Rede sein; höchstens Überschriften werden, gleichwertig mit Links und Grafiken, grob erfasst, eventuell noch kleinste, bewusst gesuchte Informationseinheiten³¹⁴.

Phase zwei ist der Prozess des „Skimmens“. Nun werden Teile von Fließtexten, insbesondere Zusammenfassungen, Absatzanfänge und hervorgehobener Text, eingehender betrachtet, wenn die Aufmerksamkeit in der Scan-Phase auf diese Elemente gelenkt wurde³¹⁵. Der Nutzer verschafft sich also in der „Skimming“-Phase einen weiteren, diesmal genaueren Überblick über allgemeine Inhaltsstruktur und taxiert in diesem Zusammenhang insbesondere die Würdigkeit eines angebotenen Beitrags, intensiver rezipiert zu werden. Jedoch dominiert auch in der „Skim“-Phase noch die „schwebende Aufmerksamkeit“³¹⁶. In dieser Beschreibung entspricht das „Skimmen“ dem, was bei Willberg und Forssmann in ihrer Typisierung der „Lesetypografie“ als „informierendes Lesen“ klassifiziert wird³¹⁷ (wird bei Willberg und Forssmann allerdings nicht als „Zwischenprozess“ gewertet, sondern als medienspezifisches Einzelphänomen, das insbesondere beim Zeitunglesen zur Anwendung komme).

Das eigentliche Lesen beginnt erst in Phase drei, die auf das Skimmen folgt, wenn die überflogenen Passagen den Betrachtenden eine tiefer gehende Befassung wert scheinen. Dies ist die Phase „fokussierter Aufmerksamkeit“³¹⁸, die bei Willberg und Forssmann als „lineares Lesen“ verstanden wird³¹⁹ und laut Köhler endlich in jener Konzentration auf Texteinheiten jenseits der Überschriften und Schlagwörter mündet: „Im Unterschied zu kurzen Texten müssen längere Texte bewusst ‚gelesen‘ werden“³²⁰.

O’Hara und Sellen differenzieren diese dritte Phase auf eine interessante Art und Weise zusätzlich aus: in „verstehendes“, also sinnentnehmendes Lesen einerseits und „reflektiertes“, also gleichsam „mitdenkendes“ Lesen andererseits³²¹. Der qualitative Unterschied zwischen diesen beiden Rezeptionsintensitäten ist zweifellos enorm; einer detaillierteren Diskussion würdig erscheint er im Angesicht der Ziele der vorliegenden Arbeit jedoch weniger.

³¹⁰ Vgl. Spool et al. 1999, S. 70

³¹¹ Vgl. Alkan 2003, S. 31f.

³¹² Vgl. Wirth 2004, S. 209

³¹³ Vgl. Hellbusch 2005, S. 19

³¹⁴ Vgl. O’Hara/Sellen 1997

³¹⁵ Vgl. Wirth 2004, S. 209

³¹⁶ Kol/Scholnik 2000, S. 70

³¹⁷ Vgl. Willberg/Forssmann 2005, S. 14

³¹⁸ Vgl. Wirth 2004, S. 210

³¹⁹ Vgl. Willberg/Forssmann 2005, S. 14

³²⁰ Köhler 2002, S. 73

³²¹ Vgl. O’Hara/Sellen 1997

Ohnehin stellt sich die Frage, ob Scannen und Skimmen ausschließlich web-spezifische Phänomene bilden. Kol und Scholnik bejahen dies mit Einschränkungen: Insbesondere die Phase des „Skimmings“ halten die Autorinnen für unterschiedlich charakterisiert im Vergleich zwischen Print- und Online-Angeboten. Während nämlich auf Papier eine überschlägige Gesamteinschätzung von Textbeiträgen und -strukturen in aller Regel leicht falle – Zeitungsbeiträge sind beispielsweise in ihrer Länge meist „auf einen Blick“ abschätzbar –, träten im Web die mediumspezifischen Tätigkeiten des Scrollens und des Folgens von Verlinkungen stärker in den Vordergrund; der „Skimming“-Prozess als beurteilende Verrichtung sei also am Bildschirm tendenziell vielgestaltiger und auch meistens langwieriger³²².

Insgesamt halte ich das beschriebene dreistufige Erklärmodell „Scannen – Skimmen – Lesen“ für ein ausgesprochen plausibles. Es steht auch meines Erachtens keineswegs in Konkurrenz zu der verbreiteten These, dass Lese-Intensität und nachhaltiges Erinnern von Textinhalten weniger durch die visuelle Form der Textdarbietung beeinflusst ist als durch die schlichte, primär inhaltlich begründete Lese-Motivation, in der ein Rezipient an eine schriftliche Aufbereitung herantritt³²³, dass also „Umfang, Schwierigkeit und Inhalt des zu lesenden Materials“³²⁴ die Qualität von Leseprozessen geradezu hauptfaktoriell beeinflusst³²⁵. Es mag durchaus sein, dass die Prozesse des „Scannens“ und „Skimmens“ bei grundsätzlicher Affinität eines Rezipienten für bestimmte Textinhalte nachhaltig abgekürzt sind – dass sie gänzlich entfallen, halte ich für unwahrscheinlich.

Was folgt aus diesen Erkenntnissen für die typografische Gestaltung von Webseiten? Zunächst ist wohl davon auszugehen, dass Niensens Unterstellung des nur mehr „scannenden“ Lesers wohl eher ein Negativ-Szenario beschreibt, aber gewiss nicht den „Web-Leseprozess“ des wie auch immer psychogrammatrisch beschriebenen Web-Nutzers abschließend und als solchen. Dennoch legen die oben beschriebenen Überlegungen die Vorgabe nahe, dass Typografie, wenn möglich, in ihrer Einrichtung dazu beitragen sollte, sicherzustellen, dass Nutzer potenziell rasch und bequem über die Phasen von „Skimming“ und „Scanning“ hinwegkommen. Eine bedeutendere Rolle kann der Schriftgestaltung zweifellos insbesondere in der „Skimming“-Phase zukommen, vor allem durch typografische Herausstellung von Schlüsselwörtern oder -sätzen, die Lesende rasch in Kenntnis setzen über zu erwartende Inhalte im gesamten Folgetext. Im klassischen Sinne spricht dies insbesondere für das Vorhalten zumindest einer erweiterten Unterzeile oder, besser noch, eines knappen Vorspanns. Diese Elemente sollten mithin auch typografisch herausgestellt werden – beispielsweise durch Fettung oder eine heraufgesetzte Schriftgröße.

³²² Kol/Scholnik 2000, S. 70

³²³ Vgl. Frutiger 2001, S. 199

³²⁴ Wandmacher 1993

³²⁵ Vgl. McLean 1992, S. 48

3 Typografische Faktoren

3.1 Die Elemente von Schriftanordnungen

Gäbe es die eine, die optimale Schriftkonstellation, wir wären wohl alltäglich umgeben von typografisch uniformierten Drucksachen. Dass augenfällig das Gegenteil der Fall ist, darf schon fast allein als Beweis dafür gelten, dass die „Universalschrift“, so sie denn bereits oder potenziell existiert, noch nicht als solche identifiziert ist. Das „Backrezept“ der idealen Typografie ist also entweder noch nicht geschrieben – oder es wird sehr flächendeckend als solches nicht erkannt.

Eine Ursache für die Nichtexistenz eines typografischen „Königsweges“ ist sicherlich, dass Schriftgestaltung, neben der Inhaltsvermittlung, immer auch visuell-ästhetischer Ausdruck ist, der mithin künstlerischen und zeitgeistigen Strömungen und Wandlungen unterliegt. Der Hauptgrund allerdings dürfte darin zu finden sein, dass eine schriftliche Mitteilung ihre visuelle Gesamtqualität – auch ihre Rezeptionsfähigkeit – aus einem durchaus komplexen Wechselspiel diverser typografischer Faktoren heraus entfaltet, die in variablem Umfang interagieren. Es sind die folgenden „Stellschrauben“, die nach herrschender Meinung Lesbarkeit und Wirkung eines Textes beeinflussen:

- Schriftart,
- Schriftgröße,
- Schriftschnitt,
- Schriftfarbe,
- Schriftausrichtung,
- Zeilenbreite,
- Zeilenabstand,
- Laufweite,
- Untergrundfarbe,
- Ausgabemedium³²⁶.

Die mehr oder minder gezielte Ausgestaltung und Kombination der vorgenannten typografischen Faktoren ergeben ein Gesamt, das ich im Folgenden als „typografische“ oder „Schriftanordnung“ bezeichnen werde. Der Darstellung und Diskussion der Einzelfaktoren und ihrer Wirkungen auf Schriftanordnungen ist Kernaufgabe dieses Hauptkapitels.

3.2 Interdependenzen der typografischen Faktoren

Es herrscht weitgehende Einigkeit in der Literatur darüber, dass die Attribute eines Schriftbildes teilweise wechselseitige Einflüsse aufeinander ausüben, einander gelegentlich sogar in ih-

³²⁶ Vgl. Zuffo 1993, S. 76f.

ren Ausprägungen bedingen, zumindest jedoch in mehr oder minder starkem Maße interdependieren³²⁷. Einige dieser Abhängigkeiten sind dabei rein physikalischer Natur: Buchstaben in einer Größe von weniger als 4 Punkt beispielsweise sind für Menschen mit durchschnittlicher Sehkraft schlicht nicht mehr identifizierbar; ein Zeilenabstand, der deutlich unter der Buchstabengröße des Textes liegt, bewirkt eine Überlagerung der Zeilen und bewirkt vergleichbar verhängnisvolle Effekte.

Doch haben sich auch andere Überzeugungen durchgesetzt. Nach herrschender Lehre bedingt beispielsweise die Verbreiterung einer Zeile bei gleichzeitiger Beibehaltung der Schriftgröße eine Vergrößerung des Zeilenabstandes, soll die Lesbarkeit des Textes auf konstantem Niveau gehalten werden. Köhler erwähnt noch die Bedeutung der Relation „Wortabstand/Zeilenabstand“ sowie die Interaktion der Schriftgröße mit der Zeilenlänge sowie mit dem Zeilenabstand³²⁸.

Auch auf diese Wechseleinflüsse innerhalb der typografischen Faktoren soll in diesem Kapitel eingegangen werden. Ihre Beschreibungen sind in die Einzel-Kapitel zu den isolierten typografischen Faktoren eingeflochten, und zwar stets an Stellen, an denen die Diskussion der jeweiligen Einzelfaktoren in Isolation bereits erfolgt ist. Aussagen über interagierende Zusammenhänge zwischen Zeilenbreite und Zeilenabstand beispielsweise finden sich im Kapitel „Zeilenbreite“, das weiter unten platziert ist als das zum Thema „Zeilenabstand“.

3.3 Das Vorgehen in diesem Kapitel

Bis auf den Faktor „Ausgabemedium“, der gleichsam als „Querschnittsmerkmal“ Einfluss auf die übrigen Faktoren nimmt, sollen die oben genannten typografischen „Stellschrauben“ im Folgenden näher beleuchtet werden. Dabei wird für jeden der oben genannten Faktoren geklärt:

- welche Begrifflichkeiten, Maßeinheiten und/oder Kategorisierungen bezüglich dieses Faktors in der Typografie gebräuchlich sind,
- in welcher Art und welchem Umfang der Faktor bedeutsam ist für die Gesteuerung einer typografischen Anordnung – auch in Kombination und Wechselwirkung mit anderen Einzelfaktoren,
- ob, in welchem Umfang und wie verlässlich sich das Attribut für die Darstellung in Webbrowsern mittels HTML/CSS einrichten und manipulieren lässt,
- ob und welche technisch basierten Unterschiede in der Bildschirmdarstellung existieren im Allgemeinen Vergleich zum hergebrachten Papiersatz,
- welche Empfehlungen sich der Literatur und qualifizierten online-basierten Quellen zur

³²⁷ Vgl. Brekle 1994, S. 221ff.

³²⁸ Vgl. Köhler 2002, S. 74

Handhabung des Attributs für den Bildschirmsatz entnehmen lassen,

- welche Ergebnisse vorliegende empirische und auf Bildschirm-Medien bezogene Studien bezüglich dieses Attributs erbracht haben und inwieweit diese Ergebnisse weitere Empfehlungen nahe legen.

3.3.1.1 Schwerpunkt der Betrachtung: Brottext

Insgesamt haben sich die meisten der klassisch journalistischen Textsorten ins Medium Internet herüber gerettet. So das

- Überschriften-Ensemble, das neben der Haupt-Schlagzeile Dach- und/oder Unterzeilen als inhaltlich und stilistisch ergänzende Elemente kennt, das
- Eröffnungs-Ensemble, das aus Vorspann, Spitzmarke und Autorenangabe bestehen kann, sowie, in variabler Länge, sowie der
- Brottext, der Fließtext, der „Beitrag“.

Die vorliegende Arbeit wird sich im weiteren Verlaufe praktisch ausschließlich mit der Rezeption von „Fließtext“ befassen. Dies nicht nur, jedoch auch aus räumlichen Erwägungen; eine Befassung mit Überschriften, Autorenzeilen, Zwischentiteln und anderen Sorten hätte eindeutig den Rahmen dieser Arbeit gesprengt.

3.3.1.2 Nicht berücksichtigt: empirische Studien vor 1990

Ein wichtiger Hinweis ist an dieser Stelle zu platzieren: In der Referierung vorliegender empirischer Erkenntnisse habe ich mich in diesem Kapitel auf Studien konzentriert, die nach dem Jahre 1990 durchgeführt und publiziert wurden; diese Grenzziehung ist zum Teil, aber eben nicht völlig willkürlich gesetzt. Denn die Bildschirme der 1980er Jahre hatten, definitiv andere Darstellungsqualitäten, eine andere Größe, andere Auflösungen als zeitgenössische Monitore vorzuweisen – viele, wenn auch nicht alle Monitore dieser Epoche zeigten Schriften invers an, also hell auf dunkelfarbenem Grund; eine große Anzahl der heute im Web üblichen Schriften existierten noch nicht, *Netscape* war noch nicht gegründet, der Begriff *Internet Explorer* hätte bei Zeitgenossen wohl höchstens ein Schulterzucken induziert. Dyson weist entsprechend darauf hin, „dass diese Resultate (der 80er Jahre-Studien, M.L.) von fragwürdiger Aussagekraft sein dürften“³²⁹. 1991 überraschten Muter und Marutto zudem mit ihrer bereits erwähnten Studie, in der sie keine Unterschiede mehr fanden in den Lesegeschwindigkeiten von Papier oder Bildschirm – und sie erklärten diese den meisten Studien der 80er Jahre völlig zuwiderlaufende Erkenntnis ausdrücklich mit den Verbesserungen in der Rechner- und Monitortechnologie: „Die Fortschritte (...), zum Beispiel bei der Auflösung, durch klarere und

³²⁹ Dyson 2004, S. 378

variablere Schriften, haben die Lesbarkeit am Bildschirm verbessert und erlauben mehr Flexibilität in der Textpräsentation“³³⁰. Dillon bestätigte diese Ansicht in seiner ausgezeichneten Übersichtsstudie 1992³³¹ und mahnte Forschungsbedarf an unter den Bedingungen fortgeschrittener Technik. So nahm ich das Erscheinungsjahr von *Windows 3.0* (der ersten *Windows*-Version, die in optischer Hinsicht wirklich die 1990er Jahre einläutete und entsprechend markterfolgreich war³³²) als Stichdatum: eben das Jahr 1990.

3.3.1.3 Die Rolle der printbasierten Lesbarkeitsforschung

Die Lesbarkeitsforschung blickt auf eine inzwischen gut 150-jährige Tradition zurück³³³. Insbesondere in den 1930er, 1940er und 1950er Jahren wurden typografische Phänomene sehr intensiv auch empirisch auf ihre Wirkung überprüft. Es stellte sich die Frage, wie mit diesen Quellen im Rahmen der vorliegenden Arbeit umzugehen war.

Es geht hier, dies sei ausdrücklich betont, nicht um die bereits diskutierte Frage, ob Print- oder Bildschirmdarstellung allgemein und per se „besser lesbar“ sei. Sondern ums Detail: um die einzelnen typografischen Faktoren und ihr Zusammenwirken. Darum, ob Webbrowser-Typografie zu merklichem Anteil eigene Gesetzmäßigkeiten entfaltet etwa bezüglich Schriftartwahl, Schriftgröße oder Zeilenabstand im Vergleich zum klassischen Papierdruck. Diese Frage reicht in ihrer Bedeutsamkeit in jedem Falle über den Rang einer rein akademischen Kategorisierung-Diskussion hinaus. Denn wer unterstellt, dass Schriftsatz am Bildschirm mit den Prinzipien des tradierten Papier- und Textilsatzes kaum noch Gemeinsamkeiten aufweist, nimmt implizit in Kauf, dass bewährte Regeln und – vor allem – die zahlenmäßig doch recht stattlichen Erkenntnisse der printbasierten Lesbarkeitsforschung für die Bewertung von Schrift auf dem Bildschirm keine oder höchstens eingeschränkte Gültigkeit besitzen.

Viele Autoren scheinen in dieser Hinsicht schwankend: „Trotz vieler technisch bedingter Eigenschaften gelten im World Wide Web vielfach die Gesetze der klassischen Typografie“³³⁴, meint etwa Neutzling. Auch George rät dazu, nicht gleich alle überkommenen typografischen Regularien über Bord zu werfen, „nur weil der Text am Bildschirm präsentiert wird“³³⁵. Bollwage stößt ins selbe Horn: „Im Internet gelten die gleichen Kriterien für gute Typografie wie für die Printmedien“³³⁶, erklärt er, während Kommer und Mersin vorsichtiger argumentieren. „Die Konventionen, die für die Druckerpresse und den Bleisatz galten, sind nicht unbedingt auch für den Monitor gültig“³³⁷, bemerken die Autorinnen, um später beispielsweise

³³⁰ Vgl. Muter/Marutto 1991, S. 264 (Übersetzung M.L.)

³³¹ Vgl. Dillon 1992, S. 1322

³³² Vgl. Makowski 2006

³³³ Vgl. Bergner 1990, S. 20

³³⁴ Neutzling 2002, S. 17

³³⁵ George 1995, S. 62 (Übersetzung M.L.)

³³⁶ Bollwage 2001, S. 102

³³⁷ Kommer/Mersin 2002, S. 106

die Diskussion empfehlenswerter Schriftgrößen am Bildschirm praktisch ausschließlich auf einer viel zitierten, aber ausdrücklich printbasierten Studie von Tinker und Paterson aus dem Jahre 1931³³⁸ aufzubauen³³⁹. In puncto Schriftgröße, so ist zu folgern, halten Kommer und Mersin die Erkenntnisse der printbasierten Forschung also für übertragbar auf den Bildschirm; ganz anders als beispielsweise Köhler. Er folgert aus einem von ihm unterstellten größeren Leseabstand am Bildschirm, Schriften seien für die Monitorausgabe um ein Viertel bis ein gutes Drittel im Vergleich zu gedruckten Texten zu vergrößern³⁴⁰. Inkongruente, vor allem aber unverwertbare Ansätze! Kolers et al. jedenfalls mahnen sehr nachdrücklich, klassische printbasierte Erkenntnisse könnten „nicht einfach und im Ganzen auf das neue Medium übertragen werden“³⁴¹.

Nach langem Abwägen habe ich mich entschieden, die Diskussion der folgenden Seiten – was die Darstellung empirischer Ergebnisse zur Lesbarkeit typografischer Konstellationen angeht – ausschließlich unter bildschirmspezifischen Gesichtspunkten vorzunehmen. Diese Entscheidung fällt ich ausdrücklich nicht, weil ich die Erkenntnisse der printbasierten Lesbarkeitsforschung en bloc für untauglich befunden hätte als Grundlagen der vorliegenden Arbeit. Die Gründe waren profanerer Natur.

Denn der gewichtigste Grund für meinen Verzicht auf die Diskussion der Print-Empirie war und ist schlicht darin zu finden, dass die Detail-Diskussion der Resultate printbasierter Lesbarkeits-Experimente den Rahmen dieser Arbeit eindeutig gesprengt hätte (Rehe widmete sich in seiner Dissertation 1981 praktisch ausschließlich der Rekapitulation und Bewertung der zu diesem Zeitpunkt vorliegenden über 200 printbasierten Studien, ohne, wie ich es tat, zusätzlich relevante, jedoch nicht-empirische Literaturempfehlungen zusammenzutragen oder gar eigenständige empirische Anstrengungen zu unternehmen³⁴²). Zudem liegen – neben Rehes hervorragender Arbeit – weitere Metastudien zur printbasierten Lesbarkeitsforschung vor³⁴³, die im Rahmen dieser Arbeit weder wiederholt noch gar übertroffen werden konnten und sollten. Letztlich gab den Ausschlag für den Verzicht, dass, wie zu zeigen sein wird, Studien zur Lesbarkeit an Monitoren zwar quantitativ noch längst nicht im Range der printbasierten Lesbarkeitsforschung vorzufinden sind, aber doch inzwischen meist so ausreichend viele Fingerzeige geben zur monitorbasierten Typografie, dass mir eine Beschränkung auf eben diese Studien zulässig schien.

³³⁸ Vgl. Tinker/Paterson 1931, S. 72ff.

³³⁹ Vgl. Kommer/Mersin 2002, S. 194

³⁴⁰ Vgl. Köhler 2002, S. 261

³⁴¹ Kolers et al. 1981. S. 526f. (Übersetzung M.L.)

³⁴² Vgl. Rehe 1981

³⁴³ Vgl. vor allem Tinker 1963; vgl. ferner Tinker 1965; vgl. ferner Wendt 1992; vgl. ferner Dyson 2002; vgl. ferner Hartley 1987

3.3.1.4 Noch eine Bemerkung zur nicht-empirischen Quellenlage

Wissenschaftliche Auseinandersetzungen mit Typografie am Bildschirm im Allgemeinen, in Webbrowsern im Besonderen sind bis heute ausgesprochen rar. Unter den Autoren der einschlägigen nicht-empirischen Literatur stellen die ausgewiesenen Praktiker die überwältigende Mehrheit – Menschen also, die aus der beruflichen Befassung mit und aus meist agentur-täglicher Tätigkeit im Webdesign Ratschläge, Fingerzeige entwickelt und publiziert haben.

Wohlgemerkt: Diese Praxis-Basierung gereicht vielen Werken keineswegs zum Nachteil. Fakt ist jedoch, dass ein wissenschaftlicher Abwägungsprozess zumindest *expressis verbis* den Ratschlägen und Setzungen kaum einer der im Folgenden zitierten Publikationen zugrunde liegt. Die meisten Autoren empfehlen mehr oder minder ausschließlich auf Grundlage reflektierten Praktizierens. Nachweislich existente Gegenmeinungen werden dabei so gut wie nie diskutiert, der Anschein empirischer Erhärtung wird tendenziell umso mehr vermieden, je selbstgewisser die Autoren vermeintliche Standards im Basta-Stil verordnen. Und wenn gelegentlich doch einmal empirische „Studien“ zitiert werden zur Untermauerung der eigenen These (oder gerne auch „mehrere Studien“, die „inzwischen gezeigt haben“), so bleiben die Quellen mehrheitlich ungenannt. Viele vermeintlich felsenfest empirisch untermauerte Thesen wiederum finden sich allerdings so verdächtig oft und gleichzeitig unbelegt, noch dazu teilweise nahezu wortgleich, in diversen Büchern, dass der ernste Verdacht nahe liegt: Hier wird etwas vermeintlich immer wahrer, nur weil es der eine von der anderen abschreibt. Die bereits erwähnte und unhaltbare Theorie, dass Lesen am Bildschirm um 25 bis 30 Prozent langsamer erfolgt als von Papier, ist nur eines der schillerndsten Beispiele für die Karriere eines Apodiktums, das durch Weitergabe von Sekundärquelle zu Sekundärquelle zur herrschenden, damit aber keineswegs richtigen Meinung aufgestiegen ist.

Einem ganz besonderem Phänomen möchte ich an dieser Stelle noch ein paar spezielle Zeilen widmen: Jenem Buch mit dem vielversprechenden Titel „Webdesign“ nämlich von Dieter Rottermund, der es fertig bringt, keine einzige müde Zeile den Themen Schriftart, Zeilenabstand, Schriftgröße oder auch nur irgend einem anderen typografischen Faktor zu widmen³⁴⁴. Es waren Werke wie diese, die mich gelegentlich zweifeln ließen, ob meine Literaturrecherche überhaupt wissenschaftlich haltbare Erkenntnisse würde zutage fördern können.

3.4 Schriftart

3.4.1 Die Schriftart als typografisches Phänomen

Den zweifellos größten Einfluss auf den visuellen Gesamtcharakter eines Textes nimmt die Schriftart, in der er gesetzt ist. Muss es angesichts dieser Tatsache nicht etwas verkürzt anmuten, bündig und abschließend nicht mehr zu fordern als: „Schrift muss gut lesbar sein“?

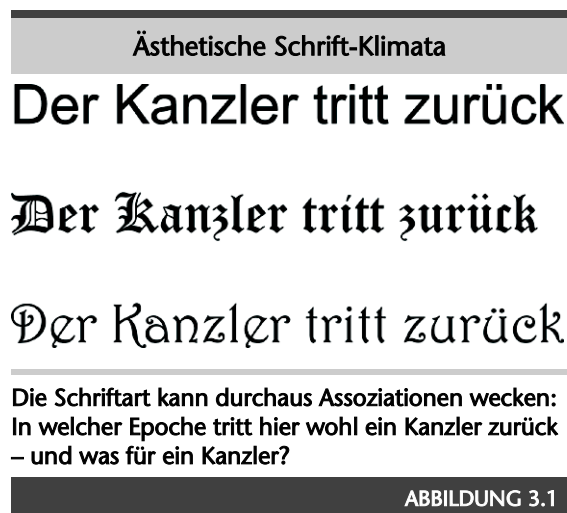
³⁴⁴ Vgl. Rottermund 2002

Denn so banal (und grundsätzlich zweifellos richtig) diese Aussage ist, so teilweise nur benennt sie die Funktion der Schriftart in der Mikrotypografie. Schriftzeichen sind mehr als bloß Träger von Information. Sie kreieren Atmosphäre, eine Anmutung³⁴⁵, „die den Leser in eine besondere Stimmung versetzt“³⁴⁶: „Der Besucher überträgt bereits über die Schriftart unbewusst vieles aus seinem Erfahrungsschatz auf (...) die dargestellten Inhalte“³⁴⁷. Die „Schriftform drückt durch die grafische Note der Buchstaben bereits viel aus“³⁴⁸, kann durch attraktives Gesamterscheinungsbild motivieren, aktivieren³⁴⁹. „Schrift ist keine Verzierung“³⁵⁰, mit dieser Ermahnung liegen Kommer und Mersin durchaus richtig – aber Schrift zu verwenden, ohne gleichzeitig eine schier ästhetische Entscheidung zu fällen, ist schlichtweg unmöglich. Funktionselement und ästhetisches Mittel: das ist der Doppelcharakter jedes Schriftzeichens „als verbale und visuelle Kommunikationsweise“³⁵¹.

3.4.1.1 Schrift als ästhetisches Gestaltungselement

Das optisch-ästhetische Erscheinungsbild einer Schriftanordnung kann sich durchaus grundstürzend wandeln, wenn eine Schriftart durch eine alternative ersetzt wird³⁵² – das visuelle „Klima“ eines Textblocks insgesamt kann erhebliche Wandlungen erfahren, wenn beispielsweise eine *Arial* durch eine Frakturschrift ersetzt wird oder durch eine verspielte Schmucktype. Abbildung 3.1 verdeutlicht dies – in einer Fortbildungs-Veranstaltung, in der ich dieses Beispiel an die

Wand projizierte, formulierte ein Teilnehmer treffend: „Oben ist Schröder zurückgetreten, in der Mitte Hitler und unten der Vorsitzende vom Kleingärtnerverein“. Für Stöckl ist Typografisches entsprechend „Stilmittel und soziales Signal“ zugleich, „ein Beschreibungsaspekt von Textstil, und zwar ein gewichtiger, weil er mit verschiedenen anderen Textaspekten in systematischer Verbindung steht, so z.B. mit der thematischen Struktur, dem Textinhalt, der Kommunikationssituation und der Partnerbeziehung etc.“³⁵³. Bonsiepe fordert entsprechend, „die ‚Literacy‘ – die Fähigkeit, Texte zu lesen und zu schreiben – durch ‚Visual Literacy‘ – die



³⁴⁵ Vgl. Gulbins/Kahrman 1992, S. 136f.

³⁴⁶ Lamprecht 2002, S. 38

³⁴⁷ Schweizer 2003, S. 150

³⁴⁸ Schäffer 2001, S. 180

³⁴⁹ Vgl. Willberg/Forssmann 2005, S. 14

³⁵⁰ Vgl. Kommer/Mersin 2002, S. 103

³⁵¹ Lynch/Horton 1999, S. 79 (Übersetzung M.L.)

³⁵² Vgl. Peck 2000, S. 154

³⁵³ Stöckl 2004, S. 7

Kompetenz im Umgang mit visuellen Distinktionen, in denen Texte erscheinen –³⁵⁴ zu erweitern. Betrachtet man den – höflich formuliert – sorglosen Umgang vieler Webseiten mit typografischen Features, ist Bonsiepe zuzustimmen; George stellt ebenfalls fest: „Aus irgendwelchen Gründen verlieren Menschen, die nicht im Traum daran denken würden, einen Brief mit einem schlechten Layout zu erstellen, beim Einrichten von Bildschirmtext ihr Gespür für Gestaltung.“³⁵⁵ Köhler spricht in diesem Zusammenhang – wohl augenzwinkernd – auch von „Hobby-Typografie“³⁵⁶.

Schriften haben eine visuelle Tonalität, eine sublimale Eigenschaft, die Neuschwander als gewichtig und gar dem eigentlichen Leseprozess vorgeschaltet erachtet: „Der emotionale Einfluss der Zeichenform wird vor ihrem literarischen Inhalt kommuniziert“³⁵⁷. Diese Tonalität beziehen Zeichen offenbar meist aus ihrer historischen Gestehung (Frakturschriften wie die mittlere in Abbildung 3.1 werden aufgrund ihrer weiten Verbreitung im Kaiserreich und in der Weimarer Republik anscheinend immer noch mit einer eher konservativen Gesinnung in Verbindung gebracht³⁵⁸), aber auch aus ihrem bisweilen „aufdringlichen“ Charakter (die untere Schrift ist allzu verspielt, als dass ein ernstes politisches Erdbeben damit in Verbindung zu bringen wäre). Köhler zieht einen interessanten Vergleich zwischen Schriftbild und gesprochenem Wort: „Im Zusammenhang mit Typografie lässt sich der Ton und die Stimmlage am ehesten mit den Schriftarten vergleichen. Jede Schriftart löst aufgrund ihres Erscheinungsbildes beim Leser bestimmte Assoziationen und Eindrücke aus“³⁵⁹.

So ist also eine der zentralen Forderungen an Typografie, dass Textinhalt und Schriftanmutung korrespondieren sollten³⁶⁰, dass „Form und Inhalt des Wortbildes zur Deckung zu bringen“³⁶¹ sind. „Oft muss das Gesamtkonzept auch die Corporate Identity (...) des Auftraggebers berücksichtigen“³⁶². Dass diese Kernforderungen gern übergangene sind, haben Kommer und Mersin zutreffend beobachtet: „Vielen Mediengestaltern ist gar nicht bewusst, wie wesentlich die Entscheidung für eine bestimmte Schriftart für die Wirkung eines Druckerzeugnisses ist“³⁶³, schreiben die Autorinnen.

3.4.1.2 Der Aspekt der Lesbarkeit

Nun ist die schiere visuelle Ästhetik einer Schriftart aber eben nur ein Aspekt eines Fonts. Der andere ist der der Rezipierbarkeit: Dass nämlich der oft augenfällige ästhetische Einfluss einer

³⁵⁴ Bonsiepe 1996, S. 80

³⁵⁵ George 1995, S. 62 (Übersetzung M.L.)

³⁵⁶ Köhler 2002, S. 7

³⁵⁷ Neuschwander 2001, S. 31 (Übersetzung M.L.)

³⁵⁸ Vgl. Stöckl 2004, S. 25f.

³⁵⁹ Köhler 2002, S. 74

³⁶⁰ Vgl. Kommer/Mersin 2002, S. 136f.

³⁶¹ Bürgel/Neumann 2001, S. 135

³⁶² Kiehn/Titzmann 1998, S. 39

³⁶³ Kommer/Mersin 2002, S. 134

Schriftart unmittelbar korreliert mit der Lesbarkeit einer Schriftanordnung insgesamt, ist keineswegs ausgemacht. Wie verhält es sich mit dem Zusammenspiel beider Teilphänomene?

Für Weidemann ist die Gewichtung klar: „Schriften kann man nach zwei Grundkriterien bewerten. Zum einen nach ihrem ästhetischen Anspruch, zum anderen unter dem Aspekt der Funktionalität, der Lesbarkeit. Ist der erstere zu vordergründig ausgeprägt, drängt er sich zwischen Botschaft und Leser und behindert die Aufnahme. Im letzteren Falle (steht) die Lesbarkeit (...) im Vordergrund. Die Erfüllung beider verkörpert den Idealfall“³⁶⁴. Auch Kapr betont, dass „die wichtigste Eigenschaft einer Schrift ihre gute Lesbarkeit ist“³⁶⁵. Im selben Sinne betonen Kiehn und Tietzmann: „Besonders bei umfangreichen Texten hat die Lesbarkeit der verwendeten Schrift Vorrang vor einer originellen Gestaltung“³⁶⁶. Auch Böhringer et al. beziehen klar Position: „Ursprünglich hat die Typografie (...) das Ziel, einen Text möglichst lesbar zu gestalten. (...) Mit dem Erfolg des Computers ist aber eine gewisse Beliebigkeit der Schriften-Verwendung einhergegangen. Es geht mehr um die (vermeintliche) Schönheit als um Lesbarkeit“³⁶⁷. Neutzling betont gleichfalls: „Gute Typografie transportiert die Inhalte angemessen und unterstützt das Lesen, ohne sich in den Vordergrund zu drängen“³⁶⁸.

3.4.1.3 Und was ist „gute Lesbarkeit“?

Spätestens an dieser Stelle ist freilich nicht mehr zu leugnen: Was „gute Lesbarkeit“ ist (im typografischen Sinne wohlgermerkt!), ist in der vorliegenden Arbeit noch nicht eingehender untersucht, gar abschließend definiert worden. Unabhängig von der Tatsache, dass die vorliegende Arbeit mit diesem Versäumnis einen festen Schulterschluss mit fast der gesamten Fachliteratur zum Thema „Webdesign“ eingeht, bedarf dieser Makel in der wissenschaftlichen Systematik selbstverständlich einer Erklärung.

Der Hauptgrund für den bisherigen Definitionsverzicht ist darin zu finden, dass die vorliegende Arbeit nicht zuletzt die Vorstellung der Ergebnisse eines von mir durchgeführten Online-Experiments zum Gegenstand hat. In der Konzeption des Designs eben dieser Studie aber hat gerade die Operationalisierung des Betrachtungsgegenstandes „Lesbarkeit“ – und damit nicht zuletzt die Charakterisierung dieser typografischen „Eigenschaft“ in einem ausführlicheren Ansatz – eine ausgesprochen zentrale Rolle eingenommen. Eine synchrone Analyse des Betrachtungsgegenstands und dessen Potenzialen einer empirischer Operationalisierung erschien mir nicht nur sinnvoll, sondern zwingend. Eine separate Klärung des Begriffs „Lesbarkeit“ am Beginn dieser Arbeit jedoch hätte diesem Ziel entgegengestanden – oder die Konstruktion unnötig komplexer und zahlreicher Wechselbezüge über diverse Seiten hinweg

³⁶⁴ Weidemann 1996, S. 15

³⁶⁵ Kapr 1983, S. 49

³⁶⁶ Kiehn/Tietzmann 1998, S. 40

³⁶⁷ Vgl. Böhringer et al. 2003, S. 47

³⁶⁸ Neutzling 2002, S. 17

und wohl auch diverse überflüssige inhaltliche Wiederholungen notwendig gemacht. Zudem referenzierte ich in der Operationalisierungsphase meines Experiments ausdrücklich auch auf Methodiken, die bei bereits durchgeführten „Lesbarkeits“-Studien zur Anwendung gekommen waren; die Darstellung dieser Studien allerdings erfolgt ebenfalls erst auf den nun folgenden Seiten.

Aus diesen Gründen heraus soll auch an dieser Stelle vorläufig nur eine „provisorische“ Definition gesetzt werden: Gut lesbar ist danach ein Text,

- der mit hoher Geschwindigkeit rezipierbar ist und
- das Textverständnis zumindest nicht behindert.

Eine detailliertere Diskussion und Abwägung erfolgt, wie angedeutet, weiter unten. Für die Besprechungen der folgenden Seiten ist die obige Kurz-Beschreibung allemale ausreichend als Basis.

3.4.1.4 Das Verhältnis Lesbarkeit – Schriftästhetik

Welche Prioritäten also sollten Gestaltende zugrunde legen bei der Auswahl einer Schriftart für eine Publikation – sei es nun eine papierne oder eine elektronische? Die Parallelen zur bereits besprochenen „Usability“-Debatte und der Debatte um die Rolle der Ästhetik innerhalb dieses Konzepts sind offenkundig. Daher rekapituliere ich an dieser Stelle nur kurz, was ich im Rahmen der überblicksweisen Darstellung der Usability-Diskussion bereits als für mich leitendes Prinzip erkannt und erklärt habe.

Insgesamt plädiere ich dafür, Schriftgestaltung tatsächlich vorrangig unter dem Aspekt der Lesbarkeit zu berücksichtigen. Jedoch sollte die Schriftästhetik – ganz im Sinne des oben beschriebenen „Joy of Use“-Ansatzes – bei der Schriftwahl auch nicht die stark oder völlig untergeordnete Rolle einnehmen, die ihr meines Erachtens oft zugewiesen wird. In diesem Sinne betrachte ich die Lesbarkeit einer Schriftart der Anmutung bedeutungshierarchisch vorgeordnet – dies jedoch ausdrücklich ohne einen „Durchschlagseffekt“ in dem Sinne, dass die „Lesbarkeit“ die Schriftästhetik als Kriterium völlig neutralisierte. Für die folgenden Aussagen gilt also: Lesbarkeit ist Kerneigenschaft einer „guten“ Schriftanordnung; jedoch ist die visuelle Schriftästhetik ein zwar nachgeordnetes, dabei aber keinesfalls ignorierbares weiteres Charakteristikum einer tauglichen Typografie.

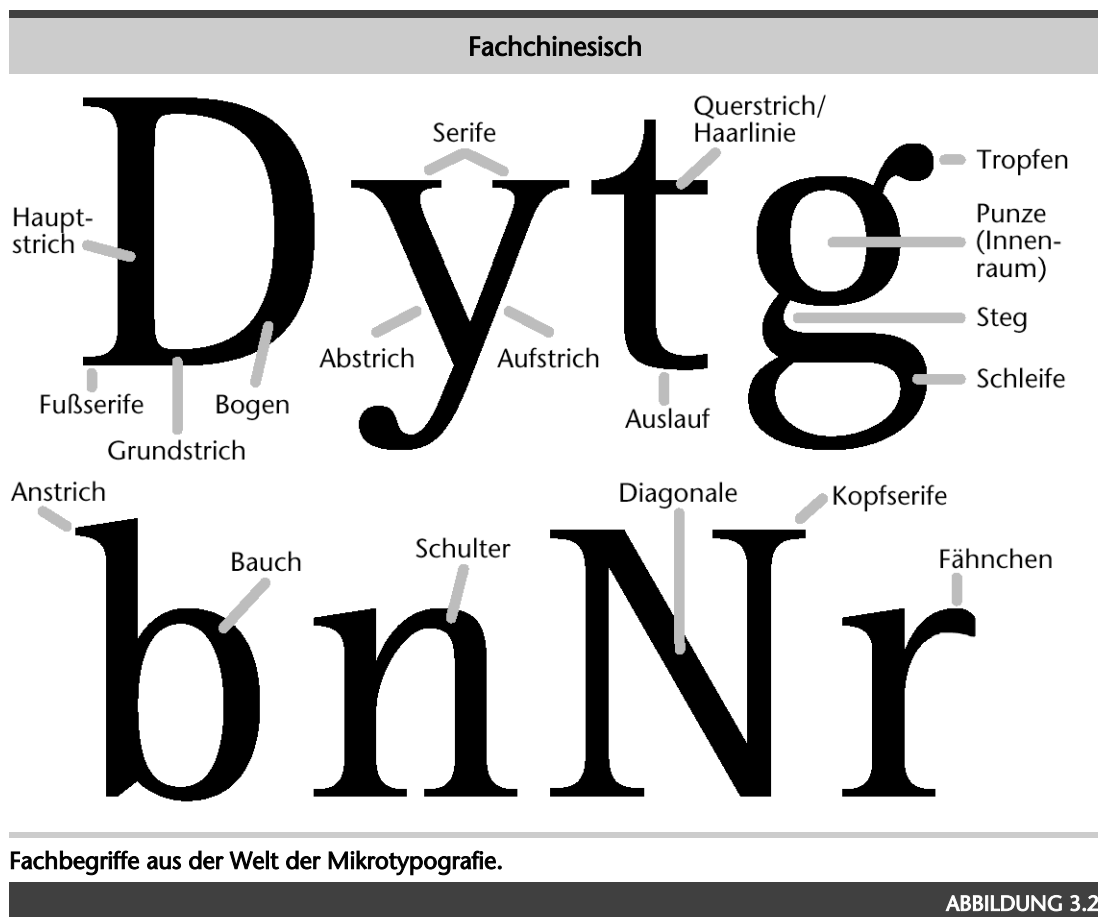
3.4.1.5 Geschichte der Schrift

Einige Rückbezüge auf die Entstehung der alphabetischen Schriften unseres Kulturkreises wurden in der vorliegenden Arbeit bereits geleistet. Auf eine erschöpfendere und systematischere Nachzeichnung der Kulturgeschichte der Schrift soll jedoch verzichtet werden; für

Thema und Ziel dieser Dissertation sind Ursprung und Weiterentwicklung von Buchstaben und Alphabeten über die Epochen hinweg zu wenig relevant, als dass eine allzu vertiefende Darstellung der Genese zu rechtfertigen wäre³⁶⁹. Lediglich einige ausgewählte historische Aspekte werden im Folgenden zur Sprache kommen, sofern sie dem Verständnis des hier interessierenden Status Quo dienen: der Browser-Typografie im Jahre 2006.

3.4.1.6 Mikrotypografische Begriffe

Die Schriftsetzer als Angehörige einer stolzen Gilde hielten es wie alle geschlossenen Gesellschaften: Sie entwickelten Fachtermini. Die Verwendung diverser dieser Ausdrücke wird in der weiteren Diskussion dieser Arbeit nicht vermeidbar sein – Abbildung 3.2 veranschaulicht daher übersichtsweise die zentralen Begrifflichkeiten.



³⁶⁹ Zur Kulturgeschichte der Schrift vgl. vor allem Földes-Papp 1966; vgl. ferner Friedrich 1966; vgl. ferner McLean 1992, S. 11ff.; vgl. ferner Raible 1991; vgl. ferner Barthel 1972; vgl. ferner Bergner 1990, S. 12ff.; vgl. ferner Frutiger 2001, S. 111ff.; vgl. ferner Luidl 1989, S. 15ff.; vgl. auch Aicher 1989, S. 20ff.; vgl. auch Tschichold 1961; vgl. auch Menzel 2000, S. 23ff.

3.4.1.7 Schriftarten – eine deutsche Kategorisierung

Die bis heute wohl meistarangezogene Kategorisierung von Schriftarten hierzulande ist die Deutsche Industrie-Norm (DIN) 16 518, die im Kern auf das Jahr 1964 zurückgeht und in Schriftgruppen unterteilt ist. Die Kategorien dieser Einteilung sollen im Folgenden vorgestellt werden; konkurrierende Klassifizierungsansätze wie etwa der des französischen Designers Maximilien Vox³⁷⁰ oder des ehemaligen DDR-Standards „TGL 37 776“³⁷¹ sollen an dieser Stelle nicht diskutiert werden, da sie im Wesentlichen zu vergleichbaren Eingruppierungen gelangen, wenn auch in anderen Kategorisierungsreihenfolgen. Im Entwurf existiert auch bereits eine neue DIN 16 518–1998, die aber noch nicht offiziellen Status erlangt hat³⁷². Obgleich die Neufassung unweit zeitgemäßer einherschreitet als ihre angestaubte Vorgängerin, werde ich im Folgenden den Fokus auf die noch gültige 60er-Jahre-Version legen – unter anderem auch deswegen, weil auf ihrer Grundlage nachvollziehbar ist, welche bedeutsamen Wandlungen die Typografie allein in den vergangenen 40 Jahren durchlebt hat. Ob eine national begrenzte Neuordnung der Schriften überhaupt und prinzipiell noch Sinn ergibt, soll an dieser Stelle nicht weiter erörtert werden – und damit auch nicht Bollwages Vorhersage; er prophezeit, dass sich im Zuge des Desktop-Publishings langfristig ohnehin anglo-amerikanische Bezeichnungen und Kategorisierungen durchsetzen werden³⁷³. Bis es soweit ist, gilt die offizielle Grundlage: die DIN 16 518.

Über die Alltagstauglichkeit dieser Norm, dies sei vorweggeschickt, lässt sich in der Tat trefflich streiten. Die Norm ist nicht nur durchzogen von der Aura durchgeistigten Akademikertums, die vor allem Herkunftsepochen und zeichnerische Kleinstphänomene in den Buchstaben zu Unterscheidungs-Merkmalen erhebt³⁷⁴; Zeitgeist-Bezüge dagegen, mögliche Verwendungsschwerpunkte (Zeitung, Werbung, Buchdruck), Lesbarkeit, Alltagspräsenz oder wechselseitige Kombinationsfähigkeiten der Schriften spielen in der DIN 16 518 keine oder höchstens eine untergeordnete Rolle.

Zudem offenbart die Norm erstaunliche Varianzen in der Tiefe der Differenzierung. So weist sie beispielsweise gleich zwei Gruppen von „Renaissance-Antiquas“ aus, die beide im modernen Print- und Digitalsatz nur noch eine sehr marginale Rolle spielen und sich untereinander hauptsächlich durch die Stellung des Querstrichs im „e“ unterscheiden. Insgesamt weist die Norm sogar fünf Untergruppen der „gebrochenen Schriften“ aus, die jenseits der Anhängerschaften bestimmter alternativer Popmusikstile kaum Popularität genießen und im Allgemeinen als „altdeutsch“, wenn nicht gar als „Nazi-Schriften“ eine sehr zweifelhafte Reputation im Gepäck führen. Dagegen ist den serifenlosen Schriften gerade einmal eine einzige Gruppe reserviert (Gruppe VI, „Serifenlose Linear-Antiqua“), was der Vielfalt der Schrift-Schöpfun-

³⁷⁰ Vgl. McLean 1992, S. 58ff.

³⁷¹ Vgl. Bergner 1990, S. 62ff.

³⁷² Vgl. Böhringer et al. 2003, S. 47

³⁷³ Vgl. Bollwage 2001, S. 67

³⁷⁴ Vgl. Gorbach 2001, S. 45

gen, die notgedrungen in dieser Kategorie zusammen zu pferchen sind, nicht annähernd gerecht wird. Und auch nicht der Bedeutung der Serifenlosen im Web (vor allem *Arial* und *Verdana*), das diese – wie noch genauer zu zeigen sein wird – ganz überwiegend prägen in typografischer Hinsicht.

Woher rührt diese evidente Geringschätzung der serifenlosen Schriften in der Norm? Aicher verweist darauf, dass die Herausbildung der serifenlosen Typen vor allem zu Beginn des 20. Jahrhunderts (die ersten „serifenlosen“ Experimente fanden Anfang des 19. Jahrhunderts statt, aber ohne größere öffentliche Wirkung³⁷⁵) mehr war als eine bloße Weiterentwicklung der Typografie im Rahmen der Konventionen des klassischen Schriftzeichnens – Künstler wie Paul Renner, Gustav Klimt und Jan Tschichold verstanden ihre typografischen Experimente als revoltierende Kunst, als visuelle Ausdrücke des Geists der Moderne: „Der Gegner war der Historizismus, der Klassizismus, die Typographie eines Garamond, Bodoni, Didot oder Caslon, die Typographie der klassischen Antiqua“³⁷⁶. Neue Schriften wie etwa die *Futura* wurzelten im Rationalismus, waren generiert aus streng geometrischen Rastern aus Kreisen, Quadraten und Linien. Das Konzept, die Idee überlagerte in vielen der frühen serifenlosen Schriften das Streben nach bestmöglicher Lesbarkeit; ein Frevel im Verständnis der traditionsbewussten Branche und wohl die Grundlage jenes „typografischen Kriegszustandes“ zwischen Bewahrern und Neuschöpfern, den Aicher als prägend für einen Großteil des 20. Jahrhunderts identifiziert hat³⁷⁷. Die Fronten blieben verhärtet zwischen den Traditionalisten, die die Neulinge unter den Schriftentwürfen spöttisch als „schlecht lesbar und seelenlos“³⁷⁸, gar als „grotesk“ bezeichneten³⁷⁹, – eine abfällige Titulierung, die bis heute zur Bezeichnung von serifenlosen Schriften, „Grotesken“ also, gebräuchlich ist – und der Welt der Schriftzeichnung einen ermüdenden Dauerstellungskrieg bescherte: „Wer für Bauhaus oder Le Corbusier war, (...) der stand auf der Seite der Grotesk, (...) wer auf der Seite der ewigen Werte und des Abendlandes stand, aber auch in der Tradition des absoluten Staates, der stand auf der Seite der Antiqua“³⁸⁰. Da half es offenbar wenig, dass spätestens in den 50er und 60er Jahren mit der *Frutiger* und der *Univers* Schriften auf den Markt traten, die auch der konservativste Typograf nicht mehr „grotesk“ nennen konnte – denn diese Schriften waren zwar immer noch (praktisch) serifenfrei, aber stark auf Lesbarkeit hin gezeichnet und mit durchaus unmonotonen, „lebendigeren“ Aspekten und eindeutigen Anleihen bei den „konservativen“ Schriftklassikern versehen, funktional in jedem Falle und definitiv „unpolitisch“. Im angelsächsischen Raum werden diese moderneren Schriftgestalten unter den serifenlosen Schriftarten auch mit dem Adjektiv „humanistisch“ versehen, um deren visuelle Anlehnung an die Neuzeit-Schriften zu signalisieren (die klassischen Grotesken werden dagegen als „geometrisch“ be-

³⁷⁵ Vgl. Friedl 1998, S. 4ff.

³⁷⁶ Aicher 1989, S. 167

³⁷⁷ Vgl. Aicher 1989, S. 166ff.

³⁷⁸ Friedl 1998, S. 9

³⁷⁹ Vgl. Lankau 2000, S. 289; vgl. ferner Bollwage 2001, S. 66

³⁸⁰ Aicher 1989, S. 168

zeichnet)³⁸¹. Doch all diese Weiterentwicklungen der Serifenlosen konnte die Fronten offenbar nicht aufweichen. Und so fristet die Grotteske ihr Dasein als Mauerblümchen in der DIN 16 518.

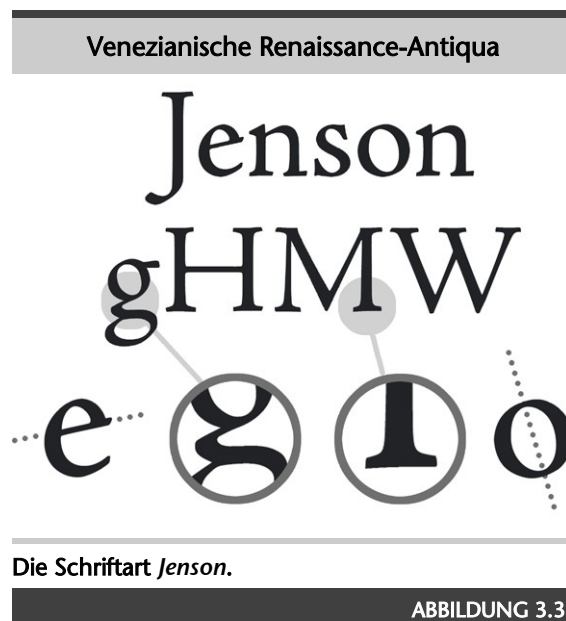
Wohlgemerkt: Dies ist kein Plädoyer gegen die differenzierte Gliederung der serifentragenden Antiquas in insgesamt 5 Gruppen in der Norm – wohl aber eines für eine detailliertere Betrachtung der serifenlosen Schriftarten. Vor allem soll hier ausdrücklich nicht einer simplifizierenden Neueinteilung von Schriften in „serifentragende“ und „serifenlose“ Schriften das Wort geredet werden. Doch beherrschen diese beiden „Meta-Gruppen“ nicht nur definitiv die heutige Medien-Landschaft, sie bieten auch wiederkehrenden Anlass zur Debatte, welche Gruppe besser lesbar sei. Schweizer referiert eine immer wieder bemühte Typografen-Grundregel, die in Wahrheit wohl kaum mehr ist als eine durchaus begründete Vermutung: „Das Auge findet an den ausgeprägten Serifen besseren Halt und kann die Buchstaben besser unterscheiden“³⁸².

Insgesamt ist festzuhalten, dass im zeitgenössischen Print- und Digitalsetz vor allem Barock-Antiquas und serifenlose Schriften im Mengensatz dominieren. Die Werbung greift gelegentlich noch auf serifenbetonte Linear-Antiquas zurück, in den Überschriften meist konservativer Zeitungen finden sich gelegentlich noch klassizistische Antiquas.

Gruppe I: Venezianische Renaissance-Antiqua (Ursprung: 15. und 16. Jahrhundert)

Die Geschichte der Druckschriften beginnt – wie sollte es anders sein – mit Erfindung des Buchdrucks durch Johannes Gutenberg. Die frühesten Druckschriften waren von einer Charakteristik, die man heute mit dem Begriff „Venezianische Renaissance-Antiqua“ bezeichnet. Schriften dieser Gruppe (wie beispielsweise die *Jenson*, die in Abb. 3.3 rechts abgebildet ist) zeichnen sich durch sichtbare, aber im Vergleich zu den übrigen Antiqua-Gruppen nicht allzu ausgeprägte Unterschiede zwischen den Stichstärken aus. Auch die Serifen

sind recht massiv ausgeformt. Die senkrechten Achsen der runden Buchstaben (beispielsweise des „o“) sind merklich nach links geneigt. Besonderes Merkmal der Venezianischen Re-



³⁸¹ Vgl. McLean 1992, S. 58ff.

³⁸² Schweizer 2003, S. 151

naissance-Antiqua ist der schräg stehende „e“-Strich. Insgesamt wirken Schriften dieser Gruppe recht robust.

Schriften dieser Gruppe spielen in der heutigen Typografie nur noch eine sehr untergeordnete Rolle. Am häufigsten trifft man Venezianische Renaissance-Antiquas noch im Buchdruck an. Im Nachrichten-Journalismus findet man sie praktisch nicht mehr.

Gruppe II: Französische Renaissance-Antiqua (Ursprung: 16. Jahrhundert)

Die Verwandtschaft der Französischen zur Venezianischen Renaissance-Antiqua ist schon durch die Benennung evident. Freilich wirken die Schriften dieser Gruppe (das Beispiel rechts ist die *Palatino*) in den Augen des heutigen Betrachters deutlich „moderner“ und weniger schwerfällig. Dies liegt vor allem daran, dass die Strichstärken-Differenzen hier merklich ausgeprägter sind als bei der Venezianischen Renaissance-Antiqua³⁸³; der „e“-Strich ist gerade ange-
setzt, während die Links-Neigung der vertikalen Achse bei runden Buchstaben

immer noch sehr stark ausformuliert ist. Die Serifen sind außerdem feiner, spitzer geschnitten. Ein weiterer Unterschied zur Venezianischen Renaissance-Antiqua: die Oberlängen von Kleinbuchstaben wie „d“ oder „h“ überragen die Oberlängen der Versalien.

Französische Renaissance-Antiquas werden durchaus noch eingesetzt auch in zeitgenössischen, vor allem Print-Entwürfen – zumindest deutlich häufiger als die Venezianischen Renaissance-Antiquas. Bekanntester „Einsatzort“ der *Palatino* ist vielleicht die angesichts des Tabak-Werbeverbots wohl bald aussterbende Plakatkampagne von „Lucky Strike“.

Gruppe III: Barock- oder Übergangs-Antiqua (Ursprung: 18. Jahrhundert)

Der Kontrast der Strichdicken ist bei Schriften dieser Gruppe noch stärker ausgeprägt als bei den beiden vorherigen Gruppen. Die Serifen sind nochmals feiner gezeichnet, die senkrechten Achsen der runden Buchstaben sind exakt oder in der Tendenz senkrecht, höchstens noch schwach links-lastig angelegt³⁸⁴. Der Begriff der „Übergangs-Antiqua“ kennzeichnet die



³⁸³ Vgl. Redelius 1998, S. 38

³⁸⁴ Vgl. Redelius 1998, S. 39

charakterliche Nähe der Barock-Antiquas sowohl zu den Schriften der vorangegangenen Renaissance- als auch der nachfolgenden klassizistischen Epoche.

Als „Protagonistin“ für diese Schriftgruppe habe ich hier die *Times* gewählt. Dass diese Type nicht in der Ära des Barock geschnitten wurde, sondern 1929, ist für sich bereits aussagekräftig: Der Typus der Barock-Antiqua hat sich als derart (druck-)tauglich und zeitlos erwiesen, dass er bis heute Grundcharakter der absolut überwiegenden Serifenschriftarten ist. Fast alle Zeitungs-Mengentexte sind heute in einer Barock-Antiqua gesetzt.

Woraus erklärt sich der Erfolg vieler Schriftarten dieser Kategorisierungsgruppe? Wohl vor allem daraus, dass Barock-Schriftarten wie die *Times* in vielerlei Hinsicht Kompromisse eingehen, die nicht „faul“ sind: die Strichstärken differenziert genug, um visuellen Reiz zu entfalten, aber nicht zu ausufernd variiert, als dass Unruhe im Schriftbild entstehen könnte. Die Serifen fein und elegant, aber nicht zu ziseliert, als dass sie im Mengendruck Schaden nehmen oder gar ganz „verschluckt“ werden könnten. Selbst eine der beliebtesten Schriftarten der Postmoderne, die *Rotis Serif* von Otl Aicher, darf in ihrem Charakter als Barock-Antiqua gelten.

Gruppe IV: Klassizistische Antiqua (Ursprung: 18. und 19. Jahrhundert)

Der Kontrastreichtum der Strichstärken wie die Feinzeichnung der Serifen ist in Schriften dieser Gruppe quasi zum Zenit vorangetrieben; die in der Zeit der Entstehung dieses Schrifttyps in Blüte stehende Technik des Kupferstichs ließ auch unter Typografen den Wunsch laut werden, extrem feinstrichige Buchstaben zur Verfügung zu erhalten. Die Schriften wirken dadurch sehr elegant, distinguiert, bestechen durch eine „ausgewogene Schlichtheit“³⁸⁵, bergen aber auch Risiken: So können klassizistische Antiquas in geringerer Druckqualität und/oder sehr kleinen Größen stark „leiden“, wenn die feineren Striche „verschluckt“ werden.



³⁸⁵ Warkentin 1992, S. 348

Aus diesem Grunde werden Klassizistische Antiquas (in Abbildung 3.6 repräsentiert durch den „Klassiker“ *Bodoni*) vor allem in größeren Schriftgraden verwendet, also als Überschriften-Typen (zum Beispiel in der „Zeit“ oder der „Financial Times Deutschland“), selten dagegen in Mengensatz.

Es wird später noch zu zeigen sein, dass die klassizistischen Antiquas insbesondere im Bildschirmsatz die heftigsten Charakter-Einbußen verzeichnen.



Gruppe V: Serifenbetonte Linear-Antiqua (Ursprung: 19. Jahrhundert)

In Schriften dieser Gruppe wirken alle Strichstärken, inklusive derer der Serifen, optisch gleich stark. Diese Schriftgruppe wird auch mit dem Namen „Egyptienne“ gekennzeichnet, da die ersten Fonts dieser Gruppe Anfang des 19. Jahrhunderts unter dem Eindruck einer großen „Ägypten-Euphorie“ entstanden vor allem in England, das 1799 Napoleon seine ausgegrabenen Schätze aus dem Land des Nils abgejagt und in britische Museen verbracht hatte³⁸⁶.

Als Beispielschrift ist hier die *Rockwell* angeführt.



Schriften dieser Gruppe erleben immer wieder modische Phasen, in denen sie gerade in der Werbung häufiger zum Einsatz kommen; Phasen sind dies aber, die meistens nicht sehr lange dauern: ihren jüngsten Frühling erlebten die „Egyptiennes“ in der Phase der New Economy – unter anderem die Telekommunikationsfirma „Arcor“ setzte Ende der 1990er-Jahre auf die der *Rockwell* sehr verwandte Schriftart *Memphis*.

³⁸⁶ Vgl. Bollwage 2001, S. 66

Gruppe VI: Serifenlose Linear-Antiqua (Ursprung: 20. Jahrhundert)

Es kam bereits zur Sprache: Diese eine Gruppe deckt, grob geschätzt, wohl zumindest 30 bis 40 Prozent aller heute in Massenmedien verwendeten Schriften ab, im Web eher noch mehr. Schon die schiere Menge deutet an, dass diese Gruppe VI sehr holzschnittartig konstruiert ist und Typen versammelt, die gerade einmal ein Merkmal verbindet: die völlige oder überwiegende Absenz von Serifen.

Bereits das zweite, in der DIN 16 518 definierte Charakteristikum der Gruppe VI jedoch suggeriert eine Uniformität der Schriften, die einem Abgleich wie dem nebenstehenden (Abbildung 3.8) absolut nicht gerecht wird: danach zeichnen sich „Serifenlose“ durch weitestgehende optische Gleichförmigkeit ihrer Strichstärken aus. In der Tat: Die *Futura*, Ende der 1920er-Jahre ganz im Geiste des Rationalismus gezeichnet, wirkt sehr geometrisch (das „e“ leitet sich unverkennbar aus der Grundform „Kreis“ her), mechanistisch, und tatsächlich sind die Strichstärken hier durchweg praktisch monoton. Auch die *Helvetica* (als Kopie wohl besser bekannt als *Arial*³⁸⁷) wirkt noch sehr formalistisch, schnörkellos-nüchtern, die Striche sind auch hier immer noch ausgesprochen wenig ausdifferenziert. Ich möchte die beiden vorgenannten Schriftarten daher als „geometrische Serifenlose“ bezeichnen – und damit eindeutig abgrenzen von denen, die ich „humanistische Serifenlose“ nennen will. Die drei unteren Beispiele in Abbildung 3.8 zeigen Lettern von Schriftarten, die ich dieser Gruppe zurechne: *Frutiger*, *Stone* und insbesondere *Rotis* sind eindeutig befreit von



Während die *Futura* als „Konzept-Schrift“ streng geometrisch eingerichtet ist, sind Schriften wie *Rotis* und *Frutiger* handgezeichnet, wirken bewusst „ungeometrischer“, und strahlen durchaus eine größere Lebendigkeit aus.

ABBILDUNG 3.8

³⁸⁷ Vgl. Herrmann 2005, S. 218

geometrisch-formalistischen Zwängen, strahlen durch differenziertere Strichvarianten Lebendigkeit und Lebhaftigkeit aus.

Die übrigen Gruppen

Die verbleibenden Schriftgruppen der DIN 16 518 werden hier nur noch in einem gesammelten, kurzen Abriss dargestellt, da sie weder im Buch- noch im Zeitungs- oder Zeitschriftendruck eine bedeutsamere Rolle spielen – und auch im Web höchstens gelegentlich zum Einsatz kommen. Gruppe VII bezeichnet dabei die sogenannten „Antiqua-Varianten“; Gulbins und Kahrmann bezeichnen diese Sparte zu Recht als „eine Art Abfallgruppe, da hier all jene Antiqua-Gruppen eingeordnet werden, die nicht in die Gruppen I bis VI hineinpassen“³⁸⁸. Auch Warkentin erkennt hier eine „Sammelsurium-Klasse (...), in der alle all jene Schriften zu finden sind, die in den neun anderen Klassen keinen Unterschlupf finden konnten“³⁸⁹.

Gruppe VIII umfasst die „Schreibschriften“, also immer noch Drucktypen, aber eben solche, die stark an die Handschrift, die „Kalligrafie“, angelehnt sind. Fein wiederum davon zu differenzieren sind die Schriften der Gruppe IX „Handschriftliche Antiqua“, die laut Definition an die Antiqua angelehnt, aber durch Elemente des Handschriftlichen modifiziert sind.

Gruppe X umfasst die „Gebrochenen Schriften“, die Fraktur also, die erstaunlich fein in nochmals fünf Untergruppen separiert werden – eine Feindifferenzierung, die, legt man den zeitgenössischen Einsatz in Medien zum Einsatz, heute wohl eher den Serifenlosen zustehen würde als jenen Typen, auf die höchstens noch die „Frankfurter Allgemeine“ als Überschriftenart zurückgreift für politische Kommentare. Gruppe XI fasst die quantitativ nicht unerhebliche Gruppe der „Fremden Schriften“ zusammen.



ABBILDUNG 3.9

3.4.2 Manipulation von Schriftarten auf HTML-basierten Webseiten

So alltäglich und gegenwärtig das Faktum im Geschäft zeitgenössischer Webdesigner ist, so bizarr mutet es nach wie vor an: Es ist bis heute praktisch unmöglich, mithilfe von HTML und CSS die Schriftart festzulegen, in der potenzielle Betrachter die Texte einer gestalteten

³⁸⁸ Vgl. Gulbins/Kahrmann 1992, S. 22

³⁸⁹ Warkentin 1992, S. 357

Webseite zu Gesicht bekommen sollen. Möglich ist nur eine Näherung mit Unwägbarkeiten. Und selbst wenn diese Näherung von Erfolg gekrönt ist, bleiben jene Probleme, die spezifisch induziert sind durch das Ausgabemedium: den Bildschirm.

3.4.2.1 Problem Nummer 1: Schriftvorschlag statt Schriftfestlegung

So eröffnen HTML und CSS zwar die grundsätzliche Möglichkeit, eine oder sogar mehrere, prioritär geordnete Schriftarten zu bestimmen, in der Texte und Textpassagen der gestalteten Seite auf dem Bildschirm dargestellt werden sollen. Doch reicht diese vermeintliche Festsetzung der Schriftart tatsächlich nicht über den Status einer „Empfehlung“, der Formulierung eines „Wunsches“ hinaus, in welcher Schriftart die Texte auf dem Bildschirm des Users erscheinen sollen³⁹⁰.

Denn die gewünschten Schriftarten können im HTML-/CSS-Quelltext lediglich durch Niederschrift ihres Namens definiert und damit quasi empfehend an den Browser weitergereicht werden, mittels dessen der Betrachter die Webseite begutachtet³⁹¹. Der Schriftsatz als solcher, und das ist entscheidend, ist jedoch nicht Bestandteil des HTML-/CSS-Quellcodes³⁹², was Web-Anwendungen entscheidend von CD-ROM- und DVD-Produktionen unterscheidet: wiewohl beide für den Bildschirm produziert werden, sind Webdesigner noch deutlich eingeschränkter in der Schriftwahl als die Hersteller von Offline-Anwendungen, die meistens die im jeweiligen Angebot benötigten Schriftarten schlicht mit ausliefern auf der CD³⁹³. Im Web ist das – im Prinzip, siehe unten – nicht möglich.

Folglich kann Text auf Webseiten nur in jenen Schriftarten dargestellt werden, die sich auf dem Rechner des die Seite betrachtenden Users befinden³⁹⁴. Gorbach taxiert diese Wahrscheinlichkeit in derart bemerkenswerter Diktion, dass das Zitat hier unmöglich ausgespart bleiben darf: „Außer den Standards ist das eher zufällig der Fall und fällt damit praktisch aus“³⁹⁵.

Was bedeutet das in der Praxis? Programmiert (das heißt also konkret: empfiehlt) der Gestalter einer Webseite die Darstellung eines Textes beispielsweise in der Schriftart *Univers 55* und ist diese Schriftart auf dem Computer des Betrachters nicht vorhanden, bleibt die Anweisung ohne Wirkung. Der Text wird in diesem Falle in der sogenannten „Standard-Schriftart“ des Browsers dargestellt³⁹⁶ (und das ist meistens *Times New Roman*). Die Folge: „Das Erscheinungsbild der Site wird (...) unkalkulierbar“³⁹⁷. Das gilt erst recht, verzichtet der Seitengestalter gleich ganz auf die Benennung einer eindeutigen Schriftart, wie es in CSS grundsätzlich

³⁹⁰ Vgl. Bollwage 2001, S. 102

³⁹¹ Pring 2000, S. 14

³⁹² Vgl. Herrmann 2005, S. 217

³⁹³ Vgl. Köhler 2002, S. 262

³⁹⁴ Vgl. Lankau 2000, S. 105

³⁹⁵ Gorbach 2001, S. 255

³⁹⁶ Vgl. Lamprecht 2002, S. 38

³⁹⁷ Köhler 2002, S. 273

ebenfalls möglich ist: Stattdessen wird in diesem Modell nur noch der Wunsch formuliert, der betreffende Text solle in einer bestimmten Schriftgattung dargestellt werden, von denen CSS fünf vorsieht: „serif“, „sans-serif“, „cursive“, „fantasy“ und „monospace“³⁹⁸. Es sind dies die sogenannten „generischen Schriftfamilien“, „CSS-Schlüsselwörter, die einen Satz von ähnlichen Schriften definieren“³⁹⁹. Bezeichnet wird also „nur eine Art Schrift und nicht eine konkrete Realisierung“⁴⁰⁰ – welche Schriftart letztlich erscheint, hängt vom Browser ab⁴⁰¹. Da dieses Vorgehen freilich weniger mit Design als mit einem Lotteriespiel gemein hat, verzichten die meisten Webdesigner dankend auf die Definition von CSS-Generika.

Stattdessen suchen die meisten Webgestalter das Risiko, dass die gewünschte Schrift unavailable ist auf dem Nutzer-Rechner, dadurch zu minimieren, dass sie, als „Fallback-Lösung“⁴⁰², in ihren Quellcodes Schriftarten empfehlen, von denen anzunehmen ist, dass ein ganz überwiegender Anteil von Betrachtern diese auf ihrem Rechner vorhält⁴⁰³. Diese Strategie freilich schränkt die Wahlfreiheit erheblich ein. Zwar ist davon auszugehen, dass etwa 90 Prozent der Computerbesitzer weltweit das Betriebssystem *Windows* der Firma *Microsoft* auf ihrem Rechner installiert haben – und damit über die im Standard-Lieferumfang des Betriebssystems enthaltenen Schriftsätze verfügen⁴⁰⁴. Bereits die Besitzer von Rechnern der Firma *Apple Macintosh*, auf denen in aller Regel das Betriebssystem *MacOS* installiert ist, verfügen jedoch nur über einen Bruchteil der in *Windows* bereit gestellten Schriftarten (und umgekehrt). Reduziert man den Pool verfügbarer Fonts nun auf die Schnittmenge der in beiden Betriebssystemen vorgehaltenen Zeichensätze, verbleiben nur noch wenige Schriftarten zur verlässlichen Verwendung. Diese werden im Folgenden als „web-sichere Schriftarten“ bezeichnet. Doch selbst innerhalb dieses bereits kleinen Pools halbwegs „verlässlicher“ Schriftarten muss qualitativ differenziert werden nach Bildschirm-Tauglichkeit. Hier tritt neben die bloße Verfügbarkeit einer Schriftart ein weiterer Aspekt: der der Auflösungsstabilität einer Schriftart⁴⁰⁵.

3.4.2.2 Problem Nummer 2: die Bildschirmauflösung

Die Qualität einer visuellen Aufbereitung – auf Bedruckstoff wie Papier oder anderen Textilien, aber ebenso am Bildschirm – wird maßgeblich bestimmt von der Höhe ihrer optischen Auflösung. Konkret ist dabei die basale Frage: In wie viele kleinste Einheiten ist ein visuelles Konstrukt – hier eben: das vereinzelt Schriftzeichen – materialisiert in seiner optischen Realisierung?

³⁹⁸ Vgl. Kübler/Struppek 1996, S. 75

³⁹⁹ Spona 2006, S. 165

⁴⁰⁰ Maurice/Rex 2006, S. 58

⁴⁰¹ Baumgardt 1998, S. 111

⁴⁰² Hellbusch 2005, S. 92

⁴⁰³ Vgl. Kesper et al. 2004, S. 47f.

⁴⁰⁴ Vgl. Dengler/Volland 2000, S. 52

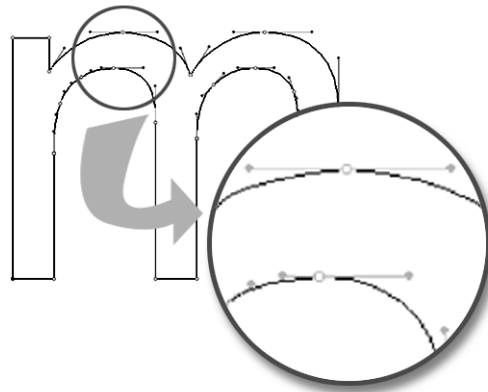
⁴⁰⁵ Vgl. Schweizer 2003, S. 155f.

Die zeitgenössisch gängige Maßeinheit der Auflösung im Druck- wie im bildschirmgestaltenden Gewerbe ist „dots per inch“ (das bedeutet: [Farb-]Punkte pro Zoll, kurz dpi). Gelegentlich ist auch von „pixel per inch“ die Rede (der angelsächsische Inch ist dabei überliefert als „Länge des Daumens König Heinrichs VIII.“⁴⁰⁶), also „Pixel pro Zoll“, abgekürzt ppi; „Pixel“ ist dabei eine bildschirmbezogene Wortschöpfung und bezeichnet die einen Bildschirmeindruck als kleinste visuelle Einheiten generierenden Bildquadrate, die „picture elements“ also – abgekürzt eben im Akronym „Pixel“.

Das Problem: Zeitgenössische Schriften werden nicht in „Pixel“-Form, also als „Bitmaps“, vertrieben – zumindest nicht für jede Größe und Auflösung separiert, wie es für differenzierte typografische Gestaltung unerlässlich ist⁴⁰⁷ –, sondern als mathematische Beschreibungen (oder „Outline“⁴⁰⁸, siehe Abbildung 3.10). Dieses Prinzip, Buchstaben algebraisch über relative Positionen und Tangenteneigenschaften zu beschreiben, bietet im Wesentlichen zwei Vorteile: Zum einen wird der Schriftsatz sehr ökonomisch definiert, das Schriftpaket ist kaum speicherintensiv. Zum anderen ist jede Letter praktisch beliebig skalierbar, muss also auch nur einmal definiert werden im Lieferumfang; was für ein Fortschritt gegenüber dem Bleisatz, als für jede Schriftgröße ein eigener kompletter Satz an Buchstaben vorhanden sein musste!

Am Bildschirm ergibt sich daraus freilich ein Problem: Der vektorisierte Buchstabe muss in die „Währungseinheit“ des Bildschirm umgerechnet werden – eben den Pixel. Diese „Herrunterrechnung“ der Original-Buchstabenkontur auf Bildquadrate wird als „Aufrasterung“ bezeichnet: Im Grundprinzip heißt dies schlicht, dass „ein Raster über den Buchstaben gelegt

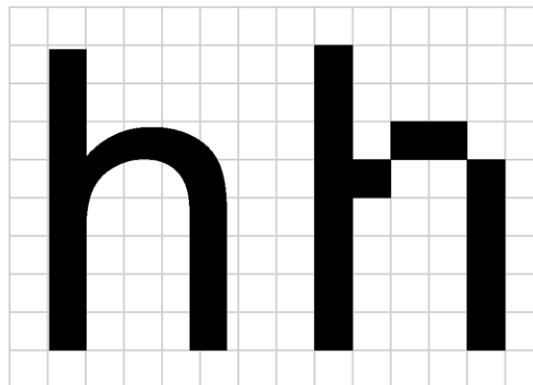
Vektorenbeschreibung einer Letter



Zeitgenössische Druckschriften liegen heute meistens als „Vektorenbeschreibungen“ vor. Das heißt: Eckpunkte und Kurven werden mathematisch beschrieben. Rundungen werden dabei mithilfe von Tangenten definiert.

ABBILDUNG 3.10

Primitive Aufrasterung einer Letter



Umsetzung eines vektorisierten in einen „Pixel“-Buchstaben.

ABBILDUNG 3.11

⁴⁰⁶ Weinman/Lentz 1998, S. 175

⁴⁰⁷ Vgl. Böhringer et al. 2003, S. 115

⁴⁰⁸ Vgl. Bayer 2003, S. 24

und von jedem Punkt gesagt (wird), ob er schwarz oder weiß ist“⁴⁰⁹ (siehe Abbildung 3.11). Doch nicht genug, dass Buchstaben in diesem Transformationsprozess erheblich „leiden“ können in ihrer visuellen Gestalt – in dieser primitiven Form der Aufrasterung kann es zudem sehr leicht zu sogenannten „Raster-Unfällen“ in der Monitordarstellung kommen⁴¹⁰ (siehe das „m“ in Abbildung 3.12).

Aus diesem Grunde versehen zumindest die Anbieter hochwertigerer Schriften heute ihre Outline-Fonts mehrheitlich mit „Anweisungen“, die die Bildschirm-Rasterung (mehr oder minder) intelligent steuern⁴¹¹. Im Hause *Apple* wird diese Ergänzung als „instruction“ bezeichnet, bei *Adobe* spricht man von „hinting“⁴¹². Letzterer Begriff hat sich inzwischen als Sammelbezeichnung für verschiedenste Ausgleichstechnologien unterschiedlichster Schriftenshersteller eingebürgert.



Die obigen Bildbeispiele lassen bereits deutlich werden: Sehr grobe Raster ziehen die Original-Konturen von Buchstaben potenziell sehr stark in Mitleidenschaft. Und grobe Raster sind auf den meisten handelsüblichen Monitoren Realität. Zum Vergleich: Qualitativ hochwertige Drucksachen werden heute in einer Auflösung von mindestens 300 dpi, eher mehr (in der Spitze bis 600 dpi) digital aufbereitet und ausgedruckt. Zeitgenössische Bildschirme (wozu auch die Displays von PDAs und Handys zu rechnen sind) weisen jedoch eine Standard-Auflösung von gerade einmal 65 bis 110 ppi auf (auf diese erstaunliche Varianz werde ich später noch eingehen). Selbst in der höchsten Ausprägung steht diese Normalauflösung also in erheblicher Diskrepanz zu Druckstandards, die auch und vor allem für die Darstellung von Schrift auf Bildschirmen durchaus folgenreich ist. Denn am Bildschirm stehen für Buchstaben, gerade in kleineren Größen, schlichtweg „nicht genügend Bildpunkte zur Verfügung(,) um eine Form in allen Einzelheiten korrekt darzustellen“⁴¹³.

Dies zeitigt Folgen vor allem für die Monitordarstellung von Schriftarten mit ausdifferenzierteren Gestalteigenschaften: für Schriftarten also mit elaborierten Unterschieden in den Strichstärken, für Lettern mit feinen Serifen, für Buchstaben mit eleganten Rundungen, kleinen Punzen oder detailverliebten Einbuchtungen – all diese Besonderheiten geraten in der Bildschirmaufbereitung potenziell unter die Räder. Die Schriftdarstellung am Bildschirm nivelliert auf Grund der geringeren Auflösung die variierenden Strichstärken einer Type zu ei-

⁴⁰⁹ Funke 1992, S. 79

⁴¹⁰ Vgl. Redelius 1998, S. 14

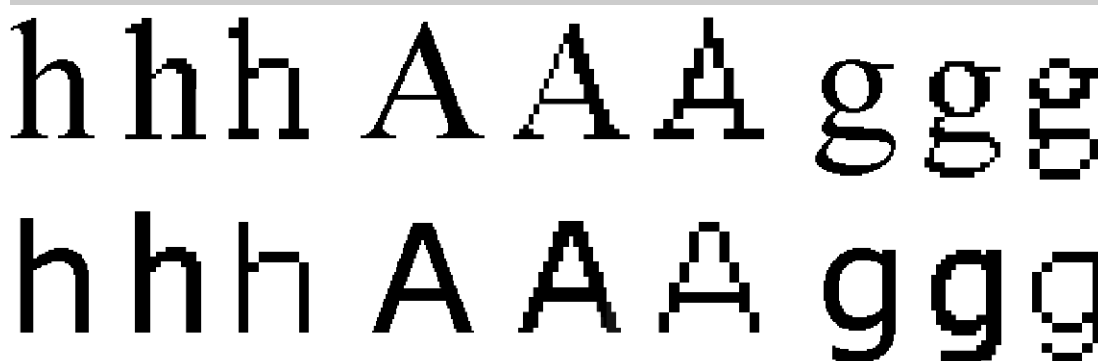
⁴¹¹ Vgl. Böhringer et al. 2003, S. 119

⁴¹² Vgl. Karow 1992, S. 113f.

⁴¹³ Neutzling 2002, S. 18

ner Einheits-Dicke, verwandelt Diagonalen oder Rundungen in „Treppen“⁴¹⁴ (oft ist hier vom „Sägezahneffekt“ die Rede⁴¹⁵), die Schrift wirkt „pixelig“⁴¹⁶, gleichsam „angefressen“⁴¹⁷ – und verwandelt rund auslaufende Serifen in Ein-Pixel-Klötze, die geschwüränlich an den Hauptlinien der Buchstaben kleben und der Eleganz der Schrift eher Abbruch tun – „halbe Pixel gibt es nun einmal nicht“⁴¹⁸. Kurz: Das Schriftbild, der eigentliche Charakter einer Schriftart kann am Bildschirm erheblichen Schaden nehmen, wenn nicht gar ihn zerstören. Für Nielsen und Tahir steht daher fest: „Alle Computer-Bildschirme sind aus typografischer Sicht schlecht, und dies wird sich auf Jahre hinaus nicht ändern, weil hochauflösende Bildschirme so teuer sind“⁴¹⁹. In diesem Sinne ist sogar fraglich, ob Adrian Frutiger, lebte er noch, Bildschirmbuchstaben überhaupt das Sammelprädikat „Typografie“ zugestehen würde; immerhin enthielt er 1978 sogar Schreibmaschinenschrift diese Adellung vor und klassifizierte sie stattdessen – fast schon ein wenig despektierlich – als „Zweckschriften“ (seine dritte und letzte Kategorie in diesem Zusammenhang sind im Übrigen die „Handschriften“)⁴²⁰.

Times und *Verdana* in verschiedenen Auflösungen



Die Schriftarten *Times* (oben) und *Verdana* (unten) im Vergleich. Um einen akkuraten Vergleich der Schriftbilder herzustellen, wurden die beiden Schriftarten auf dieselbe x-Höhe gebracht. Zur Darstellung dieses Mittellängen-Bereichs wurden jeweils einmal 48 Pixel in der Höhe verwendet (jeweils der linke Buchstabe), einmal 16 Pixel (mittlerer Buchstabe) und einmal 8 Pixel (rechter Buchstabe).

ABBILDUNG 3.13

Ein Beispiel mag die Charakter-Verzerrung illustrieren, die Schriftarten durch die mangelnde Auflösungsqualität von Bildschirmen erfahren können (vergleiche Abb. 3.13 oben). Dargestellt sind jeweils die Buchstaben „h“, „A“ und „g“, in der oberen Zeile in der Schriftart *Times*, darunter in der Schriftart *Verdana*. Jeder Buchstabe ist dabei von links nach rechts in verschiedenen Auflösungs-Varianten abgebildet: für die jeweils linke Letter einer identischen Buchstaben-Triade sind für die Repräsentation der Mittellänge, also der x-Höhe, 48 Bildquadrate vorgesehen, in der jeweils mittleren Letter sind es 16 Pixel, in der rechten 8 (letztere x-

⁴¹⁴ Neutzling 2002, S. 18

⁴¹⁵ Vgl. Kommer/Mersin 2002, S. 120

⁴¹⁶ Neutzling 2002, S. 18

⁴¹⁷ Kesper et al. 2004, S. 46

⁴¹⁸ Karow 1992, S. 114

⁴¹⁹ Nielsen/Tahir 2002, S. 51 (Übersetzung M.L.)

⁴²⁰ Vgl. Frutiger 2001, S. 199

Höhe entspricht im Falle der *Times* recht genau der Gesamtschriftgröße „13 Punkt“ – also einer nicht einmal sonderlich kleinen Ausprägung).

Beginnen wir die Analyse mit den Serifen der *Times*. Beim Buchstaben „h“ beträgt, billigt man der Letter 48 Pixel in der Höhe zur Darstellung der x-Höhe zu, das Verhältnis von Serifen-Auslaufhöhe (2 Pixel) einerseits zur Mittellänge (48 Pixel) andererseits 2:48, bereinigt also 1:24. Reduziert man die Auflösung des Buchstabens jedoch auf 16 Pixel für die Mittellänge, so sorgt der Vereinfachungs-Algorithmus dafür, dass das oben beschriebene Verhältnis sich auf 1:16 reduziert – also auf zwei Drittel des Ausgangswertes⁴²¹. Treibt man das Beispiel noch ein Moment weiter und billigt man nur noch 8 Pixel zur Darstellung der Mittellänge zu, so beansprucht die ursprünglich feine Serifenmitte mit der Höhe von wiederum 1 Pixel sogar ein Achtel der x-Höhe. Durch die Verkleinerung der Auflösung wurde die feine *Times*-Serife hier also zum typografischen „Klumpfuß“ – wenn sie nicht, in sehr kleinen Schriftgrößen, gänzlich „verschluckt“ wird⁴²².

Sehr augenfällig tritt die Mutation vor allem der *Times* ebenfalls zutage, vergleicht man die Strichstärken miteinander unter den Bedingungen der drei Auflösungen. Das Verhältnis Querstrich/Linksstrich/Rechtsstrich beträgt in bei dem Buchstaben „A“ in der Auflösung „48 Pixel x-Höhe“ 3:4:9 Pixel, in der nächstkleineren Auflösung (16 Pixel) 1:1:3, in der kleinsten 1:1:1. Das bedeutet: In der mittleren Auflösung behält die Type zumindest noch einen Teil ihres Grundcharakters bei, indem die Strichstärkenunterschiede noch zutage treten können; in der kleinsten Auflösung gleichen sich Quer- und Rechtsstrich, in der Original-Zeichnung noch im Breitenverhältnis 1:3 vorgesehen, auf die Einheitsbreite von einem Pixel an. Stellt man also die Anforderung, dass Schriften wie die *Times* mit ausgeprägteren Strichstärkendifferenzen am Bildschirm ihren Charakter behalten sollen, so scheint bei 16 Pixel Mittellänge eine gewisse Grenze erreicht. Da dies in etwa 36 Punkt vorgegebener Schriftgröße im *Internet Explorer* entspricht, liegt zumindest folgende Feststellung nahe: Eine *Times*, die auf einer Webseite ihrem Print-Original entspricht, muss in einer Größe dargeboten werden, die sie höchstens als Überschriften-Schrift qualifiziert.

Und die *Verdana*? Wie Abbildung 3.13 verdeutlicht, bleibt diese Schriftart auch bei geringerer Auflösung zumindest in ihren wesentlichen Eigenschaften deutlich weniger verletzt als die *Times*. Das hat Gründe: Serifen kennt die *Verdana* praktisch nicht (lediglich das große „i“ weist zwei Abschlussstriche auf, die „l“ einen), und die Strichstärken sind von vorneherein weitgehend monoton, „uniform“ angelegt, so dass Verzerrungen wie bei der *Times* nicht drohen. Natürlich häufen sich auch bei der *Verdana* die „Sägezahn-Effekte“ und die pixeligen Rundungen bei fortschreitender Vergrößerung der Auflösung; der Grundcharakter aber bleibt weit stabiler gewährleistet.

⁴²¹ Vgl. erläuternd auch Kommer/Mersin 2002, S. 122

⁴²² Vgl. Lankau 2000, S. 368

All diese Erkenntnisse führen immerhin vorläufig zu einem unbestreitbaren Schluss: Schriftarten mit ausgeprägteren Strichstärken-Unterschieden, mit zahlreichen, zumal feiner gezeichneten Serifen und kleinen Punzen verändern auf dem Bildschirm definitiv stärker ihren visuellen Charakter als Typen mit eher gleichförmigen Strichstärken und ohne Serifen⁴²³.

Es erweist sich aber auch dies: Schriftart und Schriftgröße interdependieren am Monitor sehr stark, definitiv stärker als im Papierdruck. Zwar leidet die Darstellungsqualität von Schriftzeichen auch im Papierdruck durchaus in Maßen, setzt man die Lettern in sehr kleiner Ausfertigung; am Monitor jedoch nivelliert die geringe Auflösung, wie gezeigt, subtile Zeicheneigenschaften geradezu in die Nichtexistenz.

Die *Arial* in drei Schriftgrößen

Dies ist die *Arial* in einer Größe von zehn Pixeln.

Dies ist die *Arial* in einer Größe von elf Pixeln.

Dies ist die *Arial* in einer Größe von zwölf Pixeln.

Die *Arial* ändert ihren Charakter schon bei minimaler Manipulation der Schriftgröße.

ABBILDUNG 3.14

Doch kann man am Bildschirm überhaupt noch von „der“ *Verdana* sprechen oder „der“ *Times*? Wie Abbildung 3.14 beweist, bewirkt die grobe Rastervorgabe des Bildschirms, dass Schriften oft schon bei Vergrößerung oder Verkleinerung um einen einzigen Punkt sehr unterschiedlich ausfallen können: „Während die *Arial* in 10 Pixel Schriftgröße recht breit (...) läuft, zeigt die gleiche Schrift bei der Darstellung in 11 Pixeln ein bereits viel engeres Satzbild. In 12 Pixeln dominiert die stark vertikale Anmutung der Schrift. Gleichzeitig gibt es nur noch wenig Differenzierung im Satzbild, das heißt, es wird schwierig einzelne Buchstaben zu erkennen“⁴²⁴. Man beachte nur die Mutation des Buchstabens „a“ in den drei Größendarstellungen und die verschiedenen Abstände der Buchstabenkombination „vo“ in den verglichenen Ausprägungen – kann man unter diesen Umständen überhaupt noch reinen Gewissens von „einer“ Schriftart sprechen, die lediglich in ihrer Dimension variiert wird? Notgedrungen werde ich dies weiterhin tun; warum ich es unter Bauchschmerzen tue, dürfte deutlich geworden sein.

Seit einigen Jahren versuchen verschiedene Hersteller, dem Problem der Grobrasterung am Bildschirm mit einer Technologie beizukommen, die allgemein als „Schriftglättung“ bezeichnet wird. Auf Rechnern der Firma *Apple Macintosh* findet dieses Prinzip bereits seit der Jahrtausendwende Anwendung; Konkurrent *Microsoft* konterte 2002 – nach ersten, zaghaften Versuchen bereits in *Windows 98*⁴²⁵ – mit „ClearType“.

⁴²³ Vgl. Kiehn/Titzmann 1998, S. 39

⁴²⁴ Neutzling 2002, S. 30

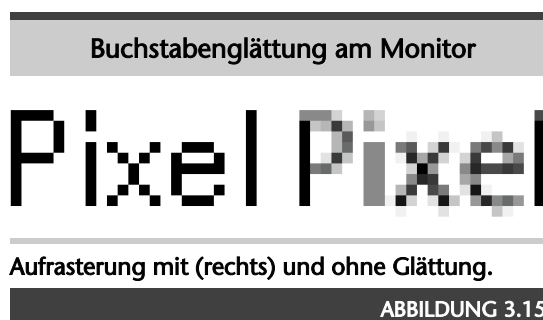
⁴²⁵ Vgl. Boyarski et al. 1998, S. 87ff.

Unabhängig vom Anbieter ist der Grundgedanke jedoch derselbe: Eine spezielle Hinting-Technik sorgt dafür, dass Bildschirmschriften nicht im herkömmlichen „1-Bit-Modus“ dargestellt werden, also als ausschließliche Kombination je einer Schrift- und einer Untergrundfarbe, sondern geglättet. „Pixel-Treppen“, aber auch überzogen generierte Buchstabenabstände werden dabei optisch durch Zwischenfarbtöne ausgeglichen; insbesondere Rundungen in Lettern erscheinen auf diese Weise potenziell harmonischer auf dem Monitor⁴²⁶. Allerdings ist auch festzuhalten, dass geglättete Buchstaben auf dem Monitor eben auch tendenziell unschärfer, verschwommen, ja „matschiger“ wirken⁴²⁷; insbesondere für Schriftgrößen ab etwa 11 Punkt abwärts ist die Technologie daher praktisch unbrauchbar⁴²⁸; Neutzling findet sogar, dass Glättung „ bei den meisten Schriften ab einer Größe von 14 Punkt zu empfehlen“ sei⁴²⁹. Turttschi meint gar, geglättete Schrift sei erst ab 16 Punkt Größe vertretbar⁴³⁰.

Wie viele Nutzer zum Zeitpunkt der Erstellung dieser Arbeit Schriften auf ihrem Bildschirm „geglättet“ betrachteten, muss ungeklärt bleiben. Diesbezügliche Erhebungen sind mir nicht bekannt; hinzu kommt: „ClearType“ ist zwar eine kostenlose Software, bedarf aber bis zur *Windows*-Version *XP* der nachträglichen Installation – eines Arbeitsschrittes also, den vermutlich eher wenige Nutzer auf sich nehmen. Auch aus diesem Grunde ist davon auszugehen, dass sich das Kleinprogramm bislang sehr wenig durchgesetzt haben dürfte. Abzuwarten bleibt, ob „ClearType“ als Standardaktivierung im für Januar 2007 angekündigten neuen *Windows*-Betriebssystem *Vista* vorgesehen ist. Wenn dies der Fall wäre, dürfte sich die Schriftglättung allerdings binnen weniger Jahre durchsetzen – vorausgesetzt, *Microsoft* versteht sein Quasi-Monopol auf dem Markt der Betriebssysteme auch mit *Vista* zu konservieren; und es spricht wenig dagegen, dass es so kommen wird.

3.4.3 Empfehlungen und Vorgaben aus der Literatur

Was ist aus den obigen Darlegungen zu folgern für den Schrifteinsatz im Web? Vorläufig hat sich die – durchaus nachvollziehbare – Mehrheits-Meinung durchgesetzt, dass auf dem Bildschirm Schriftarten einzusetzen seien, deren Original-Charakter in der Bildschirm-Darstellung möglichst wenig Schaden nehmen kann⁴³¹: Für die meisten Autoren impliziert dies vor allem



⁴²⁶ Vgl. Böhringer et al. 2003, S. 187

⁴²⁷ Vgl. Bayer 2003, S. 26

⁴²⁸ Schweizer 2003, S. 155f.

⁴²⁹ Neutzling 2002, S. 58

⁴³⁰ Vgl. Turttschi 2000, S. 317

⁴³¹ Vgl. Balzert 2004, S. 179; vgl. ferner Köhler 2002, S. 271

einen Verzicht auf Schriftarten mit Serifen⁴³², doch nicht alle Quellen gehen so weit wie Böhringer et al.⁴³³, mit den Serifen gleich eine ganze Schriftgruppe aus dem Web zu verbannen, oder wie bei Kesper et al., die praktisch ausschließlich die Schriftarten *Arial*, *Helvetica* und *Verdana* für den Webeinsatz für tauglich halten⁴³⁴ (wobei festzuhalten ist, dass die *Arial* nach herrschender Auffassung nicht mehr ist als eine Kopie der *Helvetica*).

Differenziertere Abhandlungen heben darauf ab, es sei neben der Serifenträgerschaft insbesondere darauf zu achten, ob die Lettern mit insgesamt stabilen⁴³⁵, dabei möglichst wenig variierenden Strichstärken⁴³⁶ und großen Innenräumen⁴³⁷ ausgestattet seien. Bürgel rät zusätzlich (und etwas schwammig) zur „Verwendung von Schriftarten mit ausgeprägten Wortkonturen“; aus seinen diesem Zitat voranstehenden Ausführungen lässt sich zumindest entnehmen, dass er damit Schriftarten mit markanten Oberhälften meint⁴³⁸.

Zwar gehen die Präferenzen der meisten Autoren durchaus in Richtung der serifenlosen Schriftarten; einen grundsätzlichen Bann über serifentragende sprechen allerdings die wenigsten aus. Thissen beispielsweise ist nicht grundsätzlich gegen „Antiquas“ auf dem Monitor; er glaubt jedoch, erst „ab einer gewissen Größe (je nach Schriftart ab 16 pt) lassen sich auch Serifenschriften problemlos darstellen“⁴³⁹. Auch Seibert und Hoffmann raten dazu, Serifenschriften höchstens für größer bemessene Überschriften zu verwenden⁴⁴⁰. Nielsen setzt die Größen-Schwelle bereits deutlich niedriger an: „Größer geschriebener Text (als 10 pt, Anm. M.L.) kann in Serifen-Fonts verfasst werden, wenn es besser in das Gesamtbild der Website passt“⁴⁴¹. Selbst Neutzling lässt sich bei aller Tendenz pro „Groteske“ zu der Aussage hinreißen, die *Times* gehöre „noch zu den am besten lesbaren Serifenschriftarten am Bildschirm“⁴⁴².

Gänzlich frei von „Lagerdenken“ bezüglich der Schriftgruppen gibt sich Gorbach, der erklärt, dass „*Arial*, *Verdana* und auch die *Georgia* generell bildschirmtauglich“ seien⁴⁴³. Kommer und Mersin raten sogar ausdrücklich zur Verwendung der *Georgia*, „die auf dem Bildschirm genauso scharf und sauber aussieht wie die meisten Schriften auf Papier“⁴⁴⁴. Die *Georgia* ist eine serifentragende Schriftart, die sich allerdings von der *Times* grundlegend darin unterscheidet, dass sie (siehe unten) dezidiert für den Monitoreinsatz entworfen wurde. Hier zeigt sich: Eine differenzierte Betrachtung lohnt in Bezug auf serifentragende Schriften nicht

⁴³² Vgl. Lamprecht 2002, S. 39; vgl. ferner Lankau 2000, S. 368; vgl. ferner Bürgel/Neumann 2001, S. 110; vgl. ferner Köhler 2002, S. 260; vgl. ferner Hellbusch 2005, S. 92

⁴³³ Vgl. Böhringer et al. 2003, S. 133

⁴³⁴ Vgl. Kesper et al. 2004, S. 47f.

⁴³⁵ Vgl. Gorbach 2001, S. 255

⁴³⁶ Vgl. Kommer/Mersin 2002, S. 122

⁴³⁷ Vgl. Neutzling 2002, S. 24

⁴³⁸ Vgl. Zeifle 2002, S. 17; vgl. ferner Bürgel/Neumann 2001, S. 134

⁴³⁹ Thissen 2003, S. 95

⁴⁴⁰ Vgl. Seibert/Hoffmann 2006, S. 277

⁴⁴¹ Nielsen 2000a-, S. 126

⁴⁴² Neutzling 2002, S. 24

⁴⁴³ Vgl. Gorbach 2001, S. 255

⁴⁴⁴ Kommer/Mersin 2002, S. 122

nur bezüglich der Schriftgröße – und damit der verfügbaren Pixel – sondern auch hinsichtlich der Schriftgestalt.

Für Schweizer bietet sich dagegen eigentlich gar keine der auf zeitgenössischen Webseiten gängigen Schriftarten wirklich an; die *Arial* schon gar nicht, denn diese Schriftart sei eigentlich „die schlechteste Wahl in puncto Lesbarkeit und hat auch keinen wirklichen Charakter“⁴⁴⁵. Auch die *Times* gehört für Schweizer „zu den am wenigsten geeigneten, wenn auch am häufigsten verwendeten Schriftarten“⁴⁴⁶. Für erträglich hält er zumindest die *Verdana*.

Einzig Stocksmeier stellt sich konträr zu den oben zitierten Kollegen, und zwar gründlichst; er propagiert als einziger einen klaren Vorrang für serifentragende Schriften. Ihm zufolge sollten „serifenlose Schriften ausschließlich für kurze Texte und Überschriften verwendet werden (...). Längere Texte sollten in Schriftarten mit Serifen geschrieben werden, da die kleinen Querstriche dem Auge beim Lesen helfen, den Buchstaben zu folgen“⁴⁴⁷. Stocksmeier vertritt damit als einziger die klassische „Zeitungs-Anordnung“ auch als Modell fürs Web; eine derart klare Positionierung pro Serifen-Schriftarten habe ich bei keinem anderen Autoren gefunden.

Abschließend sei noch auf Nielsen und Tahir verwiesen, die auf die grundsätzliche Vorläufigkeit aller Empfehlungen bezüglich der Schriftwahl im Web hinweisen: Viele Empfehlungen könnten sich erheblich relativieren, sobald „hochauflösende Bildschirme Standard werden“⁴⁴⁸.

3.4.3.1 Zwischenbemerkung: Verunstaltete Schrift – oder neuer Font?

Eine Zwischenbemerkung sei an dieser Stelle gestattet. Denn durch die einschlägige Literatur zieht sich eine Einsicht wie ein roter Faden, die zu hinterfragen ist. Die in der Tat unbestreitbare wie nachweisbare Verfälschung von Schriftbildern am Bildschirm wird nämlich fast durchgehend als Nachweis auch für die Verschlechterung der Lesequalität eben dieser Schriftarten verstanden. Je weniger ein Schriftbild sich in der Bildschirm-Rasterung visuell vom Original entferne, so die offenkundig herrschende Ansicht, desto besser. Doch ist diese unterstellte Kausalität nicht ein wenig kurz gegriffen?

Immerhin ist bislang selten die Frage gestellt worden, ob beispielsweise die am Bildschirm „verunstaltete“ *Times New Roman* nicht in den Rang einer ganz neuen, quasi maschinell erzeugten Schriftart erhoben werden sollte, deren Lesbarkeits-Qualitäten dann auch notwendigerweise neu, individuell, auf jeden Fall gelöst von ihrem Print-Vorbild erforscht werden

⁴⁴⁵ Schweizer 2003, S. 155

⁴⁴⁶ Schweizer 2003, S. 155

⁴⁴⁷ Stocksmeier 2002, S. 95

⁴⁴⁸ Nielsen/Tahir 2002, S. 51. Übersetzung M.L.

müssten? Erben erkennt dies als einer von Wenigen an: „Wenn aus einem Produkt durch eine Neuerung ein neues hervorgeht, kann man nicht immer die Gesetze, die für das Vorausgegangene bestanden und funktioniert haben, übernehmen“⁴⁴⁹. Kurzum: Ist die „verkrüppelte“ *Times* am Bildschirm überhaupt noch die *Times*, die wir vom Papier kennen? Oder nicht vielmehr eine ganz neue Type, die höchstens „zufällig“ den selben Namen trägt wie ihr verschwistertes Print-Original? Und: Können die hergebrachten Erkenntnisse der Lesbarkeitsforschung bezüglich der Print-Tauglichkeit einzelner Schriftarten angesichts dieser Verzerrungen im Schriftbild überhaupt noch relevant sein für die Bewertung dieser Typen auf dem Bildschirm? Ist es insbesondere ein zulässiger Kausalschluss, dass „neue ästhetische Gebilde einer Schrift (...) das charakteristische Wesen einer zugrunde liegenden Druckschrift verändern können und so (...) (die) Wahrnehmung durch das Merkmalsanalyse-System erschweren können“⁴⁵⁰? Unter anderem dieser Frage widmete sich auch meine empirische Erhebung, die weiter unten beschrieben ist.

3.4.3.2 Konsequenzen für die Schriftwahl im Web

Weiter oben wurden bereits verschiedene Schriftart-Empfehlungen für den Web-Gebrauch zitiert. Es wurden unter anderem erwähnt die Schriftarten *Arial*, *Helvetica*, *Verdana*, *Times* und *Georgia*. So bedauerlich es ist: Im wesentlichen war's das auch schon. Denn Webdesigner können, wie gezeigt, Seiten nur insofern typografisch kontrollieren, als sie Schriftarten „empfehlen“; wenn ein Nutzer jedoch eben diese Schriftart gar nicht installiert hat auf seinem Rechner, bleibt die „Definition“ wirkungslos. Es tut also Beschränkung not auf Schriften, die auf so vielen Rechnern wie möglich, unabhängig vom Betriebssystem zudem, vorgehalten werden. Und dieser Pool ist klein. Es ist daher davon auszugehen, dass nahezu alle Webseiten, die Texte in lateinischen Buchstaben anbieten, eine der folgenden fünf „Web-Standard-schriftarten“ empfehlen⁴⁵¹:

Die Web-Standardschriftarten

Times (New Roman)
in einem Beispieltext

Arial (Helvetica)
in einem Beispieltext

Verdana
in einem Beispieltext

Georgia
in einem Beispieltext

Courier
im Beispieltext

Übersicht der „web-echten“ Schriftarten.

ABBILDUNG 3.16

⁴⁴⁹ Erben 1998, S. 10

⁴⁵⁰ Bayer 2003, S. 39

⁴⁵¹ Vgl. Lazar 2005, S. 149; vgl. ferner Balzert 2004, S. 180

- *Times* (oder das Derivat *Times New Roman*)
- *Helvetica* (oder das Derivat *Arial*)
- *Verdana*
- *Georgia*
- *Courier*.

Nun kann man über die Vollständigkeit dieser Liste zugegebenermaßen trefflich streiten. Stellt man beispielsweise in Rechnung, dass geschätzte 90 Prozent aller Rechner weltweit mit dem Betriebssystem *Windows*, in welcher Version auch immer, gestartet werden, so könnte man durchaus argumentieren, dass zumindest der mit diesem Betriebssystem ausgelieferte Standard-Schriftsatz eigentlich vollumfänglich im Web verwendbar sei. Die 10 Prozent „sonstige Nutzer“ wären in dieser Argumentation also „quantité négligable“; dass Schriften wie *Comic Sans* oder *Tahoma* in aller Regel nicht installiert sind auf Computern, die insbesondere mit den Betriebssystemen *MacOS* oder *Linux* betrieben werden, würde in diesem Modell also als verschmerzbarer Randerscheinung abgetan. „kicker online“ empfahl bis zum Relaunch der Webseite am 1. Juni 2006 tatsächlich die Schriftart *Trebuchet* für die Darstellung seiner Seiten; einen Font, der lange Zeit nur auf *Windows*-Rechnern existierte.

Schweizer beispielsweise hält gar nichts von der Selbstbeschränkung auf die Standard-Schriftarten; dies freilich weniger aus resignativer Anerkennung des *Microsoft*-Monopols heraus, nein, er plädiert für völlige Wahlfreiheit: „Sicherlich hören Sie des öfteren, dass man eine Schrift nicht verwenden kann, weil sie auf den Zielsystemen im Web nicht installiert ist. (Doch das) ist kein Grund, eine Schrift nicht als Idealfall zu definieren und den Kompromiss im Falle einer nicht vorhandenen Installation einzugehen“⁴⁵².

Auf den ersten Blick wirkt Schweizers Gedanke in der Tat bestechend. Jedoch ist in Betracht zu ziehen, dass unterschiedliche Schriftarten in ihrer Größe oft sehr unterschiedlich ausfallen, auch wenn eben diese Größe numerisch identisch definiert ist (siehe Kapitel 3.5.1.2). Zudem fordern verschiedene Schriftarten je nach Eigengestalt bisweilen sehr unterschiedliche Zeilenabstände. Beide Faktoren – Schriftgröße und Zeilenabstand – können jedoch in HTML/CSS nicht in Abhängigkeit zur Schriftart definiert werden, sondern nur einmalig und universell. Das bedeutet: Der vermutlich seltene Fall, da im Sinne Schweizers ein Nutzer tatsächlich eine „exotische“ Schriftart auf seinem Rechner vorhält und mithin ein tatsächlich „erwünschtes“ Gesamtschriftbild auf seinem Bildschirm vorfindet, muss als Szenario abgewogen werden gegen den mutmaßlich häufigeren Fall, da Ersatzschriften zum Einsatz kommen. Wenn diese Schriften freilich in Platzbedarf und vor allem Gesamtschriftbild sich nicht allzu weit entfernen vom „Ideal“-Font, ist nichts dagegen einzuwenden, diesen auch ausdrücklich vorzusehen.

⁴⁵² Schweizer 2003, S. 155

Tabelle 3.1: Schriftarten populärer Websites

	Mengentext	Ergänzung
spiegel.de	Verdana	-
focus.de	Verdana	-
zeit.de	Verdana	-
kicker.de	Arial	-
sport1.de	Arial	-
netzeitung.de	Times	Verdana
bild.de	Arial	-
stern.de	Verdana	-
faz.de	Verdana	-
welt.de	Arial	-
taz.de	Arial	-
sueddeutsche.de	Verdana	-
frankfurter-rundschau.de	Verdana	-
tagesschau.de	Verdana	-
heute.de	Verdana	Arial
bundesliga.de	Arial	-
handelsblatt.de	Verdana	Times
ftd.de	Arial	-
n-tv.de	Verdana	-
n24.de	Verdana	-
newyorktimes.com	Georgia	Arial
the-Times.com	Arial	Verdana

Quelle: eigene Erhebung

Praktiziert wird Schweizers Prinzip hingegen kaum. Eher scheint der Autor Recht zu haben mit seiner düsteren Unterstellung: „Die Angst vor technischen Schwierigkeiten führt zu einer typografischen Einfalt, die man zuletzt aus der Zeit der Erfindung des Buchdrucks kannte“⁴⁵³. Die nebenstehende Tabelle (Tabelle 3.1), wiewohl im World Wide Web natürlich nur eine Momentaufnahme, listet auf, welche journalistisch orientierten Web-Angebote am Stichtag 08.09.2006 in ihren Mengentexten auf welche Schriftart referenzierten. Als relevanten „Mengentext“ habe ich dabei behelfsweise auf der jeweiligen Startseite der genannten Anbieter den jeweils obersten redaktionellen Beitrag herangezogen, der mir in Wortwahl und Gram-

matik als journalistischer „Fließtext“ identifizierbar schien und nicht lediglich als „Anreißer“ im Sinne einer erweiterten Überschrift. Als „Ergänzungsschriften“ (in der Tabelle verkürzt bezeichnet als „Ergänzung“) wertete ich Schriftarten, die auf der untersuchten Startseite komplementär zum „Brottext“ zum Einsatz kamen: als Überschriften-Type beispielsweise wie bei „handelsblatt.de“, als Anreißer-Schrift wie etwa bei „netzeitung.de“ oder als Schriftart in der Navigationsleiste wie bei der „New York Times“. Die Auswahl der analysierten Websites erfolgte willkürlich; berücksichtigt wurden die Internet-Ableger der fünf überregionalen Tageszeitungen in Deutschland, der auflagenstärksten politischen Magazine und Wochenzeitungen sowie die laut IVW meistaufgerufenen Sport-Anbieter. Wo erschließbar, analysierte ich die CSS-Vorgaben der jeweiligen Seite und leitete daraus die favorisierten Schriftarten ab; wo diese Information nicht unmittelbar zugänglich war, analysierte ich nach Augenschein, den ich aus der Betrachtung der jeweiligen Seiten an einem PC und einem Macintosh in den Browsern *Internet Explorer* und *Firefox* gewann.

⁴⁵³ Schweizer 2003, S. 200

Dabei offenbarte sich allein die „netzeitung“ als Verwenderin der *Times* in ihren Fließtexten (und auch den größeren Headlines), und damit als eine der offenkundig letzten Bastionen der Serifen im Web. Dabei kann die „netzeitung“ in dieser typografischen Entscheidung nicht einmal dort prominente Vorbilder für sich reklamieren, wo man sie am ehesten vermuten würde: die Website der renommierten Tageszeitung „New York Times“ (newyorktimes.com) bereitet ihre Mengentexte in der Schriftart *Georgia* auf. Selbst die Londoner *Times* gibt sich auf ihrer Website in typografischer Hinsicht erstaunlich emotionsfrei: Das Zeitungshaus, dessen Gestaltungsauftrag aus dem Jahre 1929 Typografen aller Welt immerhin die gleichnamige Schriftart verdanken, setzt in seinen Web-Artikeln auf die *Arial*, in der Navigation auf *Verdana*.

Insgesamt ist Tabelle 3.1 zu entnehmen: Die von Schweizer noch 2003 befürchtete „Arialisierung“ des Web⁴⁵⁴ ist ausgeblieben; dass stattdessen gleichsam eine „Verdanisierung“ von Webseiten zu konstatieren ist, beschreibt allerdings ein vergleichbar bedauerliches Szenario. Tatsächlich weist ausgerechnet jenes vermeintlich „innovative“ Medium World Wide Web in zwischen eine optische Uniformität auf (nicht nur in typografischer Hinsicht übrigens), angesichts derer selbst die gestalterisch lange als eher konservativ gesonnen geltende bundesdeutsche Tagespresse⁴⁵⁵ geradezu als experimentierfreudig dasteht.

Die *Verdana* hat damit eine bemerkenswerte und anhaltende Erfolgsgeschichte aufzuweisen. Dabei ist der Schriftsatz jung wie kaum ein anderer: Die *Verdana*, als Kreation in Auftrag gegeben vom US-Softwarehersteller *Microsoft*, wurde 1996 vorgestellt⁴⁵⁶ (und als Standard-schriftart mit dem Betriebssystem *Windows 98* ausgeliefert, was die Schrift auch im Web „massentauglich“ geraten ließ). Laut Aussage ihres Schöpfers Matthew Carter entstand die Schriftart aus dem Antrieb heraus und mit dem Ziel, eine dezidiert bildschirmtaugliche Schrift zu schöpfen⁴⁵⁷ – auch und insbesondere eine, die besonders in kleineren Schriftgrößen durch großzügigere Buchstabenabstände eine gute Lesbarkeit entfaltet⁴⁵⁸. Weder *Microsoft* noch Carter verleugnen im Übrigen, dass einerseits die populäre Schriftart *Frutiger* von Adrian Frutiger (entwickelt 1976⁴⁵⁹) sowie andererseits die Schriftart *Johnston Sans* des englischen Kalligrafen Edward Johnston, die 1916 für die Informations-Beschriftung der Londoner U-Bahnhöfe und das Logo der „Underground“ entworfen wurde, in sehr vielen ihrer Charakteristika Pate gestanden haben bei der Entwicklung der *Verdana*⁴⁶⁰. Weniger eine eigenständige Schriftart als vielmehr ein entwurfstechnischer Vorläufer der *Verdana* ist übrigens die *Tahoma*, die Carter 1995 entwarf – und eben später vor allem durch die Installation größerer Buchstabenabstände zur *Verdana* vervollkommnete.

⁴⁵⁴ Vgl. Schweizer 2003, S. 156

⁴⁵⁵ Vgl. Garcia 1997, S. 15

⁴⁵⁶ Vgl. Bayer 2003, S. 51f.

⁴⁵⁷ Vgl. Schäffer 2001, S. 181

⁴⁵⁸ Microsoft Corp. 1997

⁴⁵⁹ Vgl. Friedl 1998, S. 44

⁴⁶⁰ Microsoft Corp. 1997

Doch die bloße Präsenz der *Verdana* im Standard-Schriftenpool des quasimonopolischen Betriebssystems *Windows* taugt kaum zur Erklärung des offenkundigen Erfolgs dieser Schriftart. Denn eher bescheiden durchgesetzt im Web hat sich bislang die Schriftart *Georgia*, die ebenfalls aus der Feder von Matthew Carter stammt⁴⁶¹ und 1996 zeitgleich mit der *Verdana* vorgestellt wurde – und genauso Aufnahme in den *Microsoft*-Standardschriftsatz fand (zu dem sie bis heute gehört). Entworfen als webtaugliche Serifen-Schriftart und lesefreundlichere Alternative zur *Times New Roman*⁴⁶², hat die *Georgia*, wie gezeigt, erkennbar zumindest mehr Befürworter als die *Times* – für van Aaken ist sie wegen ihrer „wohligen Wärme“ gar „die einzig brauchbare Serifenschrift im Web“⁴⁶³.

Festzuhalten bleibt vorläufig: Die Schriftarten *Verdana* und *Georgia* wurden intentional für den Monitoregebrauch geschnitten. Das unterscheidet diese Fonts von den beiden anderen, nach meiner Erhebung ebenfalls durchaus zumindest noch stellenweise genutzten Schriftarten im Web: *Times* und *Helvetica*. Diese Schriftarten datieren in ihrer Entstehung auf Termine, die weit vor der Veröffentlichung der ersten Webseiten lagen; es waren und sind im Grundcharakter also Print-Schriften (auch wenn Lazar die *Helvetica*-Kopie als „Bildschirm-schrift“ identifiziert⁴⁶⁴).

Die *Times* wurde vom englischen Typografen Stanley Morison für die und im Auftrag der Londoner Tageszeitung „The Times“ entwickelt⁴⁶⁵; das Blatt erschien erstmals am 3. Oktober 1932 in eben dieser Schriftart. Die *Helvetica* wurde geschnitten vom Schweizer Typografen Max Miedinger, einem Vertreter der legendären „Schweizer Typografie“⁴⁶⁶ 1957 für den Zürcher Flughafen⁴⁶⁷.

Augenscheinlich gänzlich aus der Mode – nimmt man die Verwendung im Web als Maßstab für den allgemeinen Zuspruch – ist die *Courier*. Sie ist die einzige „Monotypeschrift“ der fünf oben genannten „Web-Standardschriftarten“, das heißt, alle Buchstaben weisen exakt die gleiche Breite auf (im Gegensatz zu den anderen vieren „Proportionalschriften“, in denen beispielsweise ein „i“ deutlich schmaler ausfällt als ein „W“). Zumindest Tabelle 3.1 auf Seite 104 zeigt: Kaum eine „wichtige“ Webseite bedient sich heute noch der *Courier* als Schriftart.

3.4.3.3 Empfehlungen zur Schriftmischung

Eine wirklich originelle Schriftmischung ist im Web kaum zu realisieren. Wie auch, angesichts von höchstens einer Handvoll „web-sicherer“ Schriftarten? Höchstens in Kombination mit

⁴⁶¹ Microsoft Corp. 2006

⁴⁶² Vgl. Kersken 2004, S. 823

⁴⁶³ van Aaken 2005b

⁴⁶⁴ Vgl. Lazar 2005, S. 149

⁴⁶⁵ Vgl. Herrmann 2005, S. 219

⁴⁶⁶ Vgl. Friedl 1998, S. 12

⁴⁶⁷ Vgl. Klein 2005, S. 13

Text-Bildern (siehe unten) lässt sich eine gewisse typografische Varianz realisieren, die auch höheren gestalterischen Ansprüchen genügt und die von Bergner propagierte Belegungswirkung aufs Gesamtlayout entfaltet⁴⁶⁸.

Insgesamt gehen die Mahnungen der Literatur eher in die Richtung „Maßhalten“: „Der Einsatz vieler Fonts ergibt ein Aussehen wie auf einem Steckbrief“⁴⁶⁹, warnt Nielsen; gemeinsam mit Tahir weist er zudem darauf hin, dass allzu viele verschiedene Fonts ein Charaktermerkmal von Anzeigen seien und daher die Textglaubwürdigkeit insgesamt leiden könne bei Verwendung allzu vieler Fonts⁴⁷⁰. Schäffer rät, „niemals mehr als zwei, drei Schriftarten gleichzeitig auf einer Seite“⁴⁷¹ vorzusehen, Hellbusch plädiert für „maximal 2“⁴⁷², Lazar rät gar dazu, eine einzige Schrift pro Seite vorzusehen⁴⁷³. Wenz versucht sich im launigen Duktus, wenn er sagt: „Zwanzig verschiedene Fonts (...) sind zwar eine coole Sache, sorgen aber für Tränen in den Augen des Betrachters und ein leichtes Grummeln in der Magengegend“⁴⁷⁴ – wie diese 20 Schriftarten aber angesichts der oben beschriebenen Einschränkungen der HTML-/CSS-basierten Schriftwahl überhaupt auf die Seite gelangen sollen, erklärt er nicht.

Insgesamt wirkt es etwas befremdlich, dass die zitierten Autoren so massiv vor der Verwendung zu vieler Schriften auf ein und derselben Webseite werben. Tabelle 3.1 hat schließlich eindrucksvoll gezeigt, dass auf zeitgenössischen journalistischen Webseiten in Sachen Schriftmixturen eher triste Einfalt als bedrohliche Vielfalt regiert: *Verdana* allenthalben, und dann mehrheitlich gleich für alle Textsorten verwendet – Überschriften, Vorspanne, Links und so weiter. Das sorgt unbestritten für optische Konsistenz⁴⁷⁵, lässt aber wiederum das vermeintlich altmodische Medium „Zeitung“ in seiner unbestrittenen typografischen Vielfalt am Kiosk geradezu als Spielwiese innovativen Designs glänzen.

3.4.4 Alternativen zu den Standardschriften

Ein Web voller *Verdana*, bereichert gelegentlich um die *Times*, die *Arial*, vielleicht die *Georgia* – lässt sich in der Welt von HTML und CSS tatsächlich nicht entfliehen aus dem demoralisierenden „Getto“ jener wenigen Schriftarten, die sich anbieten, weil sie nun mal „fast jeder auf dem Rechner hat“? Ist das Corporate Design eines Auftraggebers, zumindest in typografischer Hinsicht, tatsächlich überhaupt nicht abbildbar auf einer Webseite, es sei denn, dieses Corporate Design beruht zufälligerweise auf einer jener allein durch ihren Massengebrauch eher als uninspiriert zu bezeichnenden Schriftarten *Arial*, *Times* oder *Verdana*? Dieses Pro-

⁴⁶⁸ Vgl. Bergner 1990, S. 69f.

⁴⁶⁹ Nielsen 2000a-, S. 83

⁴⁷⁰ Nielsen/Tahir 2002, S. 23. Übersetzung M.L.

⁴⁷¹ Schäffer 2001, S. 189

⁴⁷² Hellbusch 2005, S. 93

⁴⁷³ Vgl. Lazar 2005, S. 149

⁴⁷⁴ Wenz 2000, S. 34

⁴⁷⁵ Vgl. Peck 2000, S. 153ff.

blem treibt Webdesigner und Entwickler seit frühesten WWW-Epochen um, und ein paar Alternativen existieren durchaus – freilich unbefriedigende beziehungsweise eindeutig „unausgorene“. Sie seien hier dennoch vorgestellt.

3.4.4.1 True Doc und OpenType

Große Hoffnungen setzten viele Webdesigner um die Jahrtausendwende herum in ein Konzept, das wirklich das Zeug zum Wegbereiter einer grafischen Revolution im Web zu haben schien: die „eingebettete“ Schrift beziehungsweise, neudeutsch, „font embedding“. Mit den Seiteninformationen, so der Grundgedanke, sollte der Anbieter einer Webseite auch Schriftinformationen an den Browser ausliefern können, also die Buchstaben und Zeichen der gewünschten Schriftart einfach mit übertragen, anstatt auf die Fonts zuzugreifen, die der Zielrechner vorhielt⁴⁷⁶. Die Selbstbeschränkung auf die schmale Palette der *Microsoft*-Standardschriften hätte damit in der Tat ein schönes Ende gefunden.

Leider geschah, was absehbar war: *Microsoft* als Marktführer und *Netscape* als damals noch ernstzunehmendes, innovatives Pendant entwickelten zwei unterschiedliche Technologien. Folge: Wer sich auf die Neuerung des „Schrift-Einbettens“ einlassen wollte, war gezwungen, jede Webseite zwei Mal zu entwickeln – einmal auf der Basis der Technologie „TrueDoc“ von *Microsoft* (das Schriftformat heißt hier „Embedded OpenType“, die Schriftdatei trägt entsprechend die Endung „.eot“), einmal ausgelegt auf „TrueDoc“ von *Netscape*, das die Firma in Kooperation mit dem Softwarehersteller *Bitstream* entwickelt hatte (dieses Schriftformat heißt „Portable Font Resource“ und greift auf Schriftdateien mit der Endung „.pfr“ zu⁴⁷⁷).

Doch nicht genug damit, dass Webseiten mit eingebetteten Fonts nunmehr quasi ein potenziell kostenintensives Zwillingdasein zu führen hatten – um eine Schrift mit der Webseiten-Information ausliefern zu können, genügte es noch nicht, diese zu besitzen; selbst wenn sie legal erworben und nicht räuberisch kopiert auf dem Rechner des Gestalters der Verwendung harnte. Denn die beschriebenen Formate „eot“ oder „pfr“ bildeten eben völlig neue Datei-Formate, die inkompatibel waren und sind zumindest mit jenen Schriftformaten, auf die beispielsweise gängige Betriebssysteme wie *Microsoft Windows* zugreifen. Bevor also eine Schriftart tatsächlich in eine Webseite eingebunden werden kann, muss sie zunächst in mindestens eines dieser beiden neuen Formate umgewandelt, konvertiert werden. Und dafür braucht es wiederum spezielle Software, nicht ganz billig im Erwerb selbstverständlich, die diese Umwandlung vorzunehmen in der Lage ist⁴⁷⁸.

Nun existieren inzwischen immerhin Programme, die Gestalter von Webseiten in den Stand setzen, die beiden konkurrierenden „Einbettungs“-Formate zumindest in einem gebündelten

⁴⁷⁶ Vgl. Kiehn/Titzmann 1998, S. 32

⁴⁷⁷ Vgl. Weinman 2000, S. 266

⁴⁷⁸ Pring 2000, S. 31

Arbeitsgang zu erstellen, der Kauf von zweierlei Software für die zweierlei Formate ist *Microsoft Windows* nicht mehr notwendig⁴⁷⁹. Außerdem können in diesen Programmen (wie beispielsweise „WEFT“ oder „FontLabStudio“) nicht nur vorhandene Schriftarten umgewandelt, sondern sogar eigene, neue gestaltet werden. Doch auch diese Programme kosten eben Geld, womit ein entscheidender „Sargnagel“ an das Projekt „embedded font“ angefügt war. Kostengünstige Shareware, also Kleinstprogramme für den Amateurgebrauch mit eingeschränkten Funktionalitäten, die die Font-Einbettung aber eventuell erschwinglich und damit attraktiver hätten geraten lassen, ließen auf sich warten. Aus dem W3C-offiziellen Standard-satz zumindest sind die Dateiformate „eot“ und „pfr“ inzwischen verschwunden⁴⁸⁰. „Embedded Type“ war (und ist) vermutlich schlichtweg zu teuer.

Ein weiterer Grund, und nicht der unerheblichste, warum sich keine der beiden im Prinzip hoffnungsvollen Technologien bis heute nachhaltig hat durchsetzen können, dürfte der Widerstand aus dem Lager der Schrift-Designer und der Schriftenhersteller sein, die schon früh Opposition gegen die eine wie die andere Lösung machten. Zu den technischen Widrigkeiten gesellten sich nämlich Probleme des Urheberschutzes⁴⁸¹. So ging im Lager der Schriftgestalter und -verkäufer bereits früh nach Vorstellung der Technologien die Furcht um, „dass Hacker problemlos die Konturen von eingebetteten TrueType- und OpenType-Schriften stehen könnten und dass die gespeicherten Zeichensätze von TrueDoc mithilfe von automatischen Schriftklaus-Robotern gesammelt werden könnten“⁴⁸². Bedenkt man, dass die Kreation einer hochwertigen Schrift Monate, ja Jahre konzentrierter Feinstarbeit in Anspruch nehmen kann – weshalb eine qualitativ wertige Schriftfamilie auch gut und gerne mal 1000 Euro und mehr kosten kann – waren diese Vorbehalte durchaus verständlich.

Viele Schriftenhersteller untersagen bis heute kategorisch die Einbettung aller oder eines Teils ihrer Fonts in Webseiten⁴⁸³. Dabei fanden die Entwickler-Firmen *Microsoft*, *Netscape* und *Bitstream* durchaus angemessene Lösungen für das Copyright-Problem. Denn bei genauer Betrachtung werden die Schriften bei den beschriebenen Technologien gar nicht „eingebettet“, also zum integralen Bestandteil der Webseite an sich gemacht, sondern nur „referenziert“. Das heißt: Die aufgerufene Webseite stellt ihrerseits wiederum einen Kontakt zu der verwendeten Schriftdatei her, die jedoch, für den Benutzer unzugänglich, auf dem Ursprungsrechner abgelegt ist, und verwertet diese Informationen weiter. Ein echtes „Herunterladen“ der Schriftdatei auf den Computer des Benutzers findet in diesem Modell also gar nicht statt⁴⁸⁴; damit aber sind auch die Missbrauchspotenziale erheblich eingeschränkt. Dass freilich versierte Hacker nicht auch diese „Sicherheits-Wände“ zu knacken verstehen würden, war und ist zugegebenermaßen unwahrscheinlich.

⁴⁷⁹ Vgl. Köhler 2002, S. 274ff.

⁴⁸⁰ Vgl. Maurice/Rex 2006, S. 59f.

⁴⁸¹ Vgl. Kübler/Struppek 1996, S. 28; vgl. ferner Böhringer et al. 2003, S. 871

⁴⁸² Weinman 2000, S. 267

⁴⁸³ Vgl. Schweizer 2003, S. 201

⁴⁸⁴ Vgl. Schweizer 2003, S. 201

Betrachtet man die Landschaft professionell betriebener Seiten im Jahre 2006, so muss konstatiert werden, dass das Vorhaben „embedded fonts“ zumindest vorläufig als auf der ganzen Linie gescheitert bezeichnet werden muss. Mir selbst ist keine bedeutendere Site bekannt, die auf eingebettete Schriftarten auf Basis der beschriebenen Technologien zurückgreift. Da hilft es wenig, dass zumindest ein Vorbehalt aus den frühen Entwicklungstagen von TrueDoc und Embedded OpenType längst als gegenstandslos angesehen werden darf: Dass die „eot“- und „prf“-Schriftdateien mit ihren Größen von 20 bis 40 Kilobyte zu unerträglichen Ausdehnungen der Ladezeiten führen würden⁴⁸⁵, hat sich im Zeitalter von DSL als vernachlässigbarer Schönheitsfehler im Konstrukt erwiesen. Für einen neuerlichen Anlauf einer der beteiligten Firmen, das Vorhaben doch noch zum Erfolg zu führen (oder mit einer neuen Technologie einen weiteren Vorstoß zu wagen), gibt es derzeit zumindest keine Anhaltspunkte. Und selbst wenn es sie gäbe: Ob die „Web-Gemeinde“ sie annehmen würde, steht dahin. Lankau steht sicherlich nicht allein mit seiner Ansicht, dass „(kein) vernünftiger Mensch (...) das Web auf eine Technik umstellen (wird), die einem Unternehmen die Monopolstellung und regelmäßige Lizenzeinnahmen sichert“⁴⁸⁶.

3.4.4.2 Typografie-Bilder

Was Weinman durchaus klangvoll als „grafikbasierte Typografie“⁴⁸⁷ bezeichnet, ist eine zweite, doch im Vergleich zur Font-Einbettung eher noch schwächere Alternative zum Diktat der Systemschriften. Denn natürlich kann es die in HTML und CSS „eingeschränkte(n) Funktionen der Absatz- und Textgestaltung“⁴⁸⁸ ausgleichen, Textpassagen als Bilder abzuspeichern und dann als „Text-Surrogat“ auf einer Webseite zu platzieren: Man kann Überschriften durchaus in einer noch so exotischen Schriftart im Bildprogramm *Photoshop* eingeben, dieses „Text-Bild“ webgerecht abspeichern und auf der Seite positionieren⁴⁸⁹. Zumindest teilweise ist auf diese Art und Weise zweifellos auch das Corporate Design eines Unternehmens auch auf der Webseite zu verwirklichen⁴⁹⁰. Aber sind diese „Typo-Bitmaps“ wirklich eine Hilfe? Höchstens gelegentlich. Im Alltag der Erstellung nachrichtlich orientierter Web-Angebote sicherlich nicht.

Denn diese vermeintlich „absolut unfehlbare Methode der Schriftdarstellung“⁴⁹¹ zeitigt erhebliche Nebenwirkungen. Zum einen muss jeder neue oder auch nur in Nuancen modifizierte Text zunächst in einem Bildbearbeitungsprogramm gesetzt und in ein web-taugliches Bild-Format übersetzt werden (meistens in ein GIF), bevor er als Inhalt auf der Seite erschei-

⁴⁸⁵ Wenz/Hauser 2004, S. 24

⁴⁸⁶ Lankau 2000, S. 367

⁴⁸⁷ Weinman 2000, S. 262

⁴⁸⁸ Vgl. Köhler 2002, S. 271

⁴⁸⁹ Vgl. Peck 2000, S. 147; vgl. ferner Lynch/Horton 1999, S. 96

⁴⁹⁰ Vgl. Bollwage 2001, S. 103

⁴⁹¹ Pring 2000, S. 32

nen kann⁴⁹²: Zumindest für journalistisch anspruchsvollere Angebote wie „Spiegel online“, die täglich mehrere hundert Texte neu oder modifiziert ins Web stellen, wäre das eine kaum zu bewältigende Aufgabe⁴⁹³ – und, bedingt durch den stets unerlässlichen Zwischenschritt der Bilderstellung, auch ein Aktualitätsnachteil gegenüber der Konkurrenz.

Zum zweiten werden Bilder im Web von Suchmaschinen nicht inhaltlich-typografisch dekodiert. Sie existieren nur als „Bitmaps“, als „Pixel-Haufen“, nicht als schriftliche Informationsträger; die „Maschinenlesbarkeit“ des Dokuments ist ausgehebelt⁴⁹⁴. Wer also Texte in Form von Bildern auf seine Website stellt, muss entweder auf eine qualitativ-inhaltliche Indizierung durch Suchmaschinen verzichten⁴⁹⁵, da diese Suchmaschinen Schriftelemente in Bildern nicht dekodieren können⁴⁹⁶, oder diesen Nachteil programmiertechnisch ausgleichen – sei es als „alt“-Tag im Bildformat⁴⁹⁷ oder als Referenzverweis im Kopf des HTML-Dokuments, denn diese Seiteninformationen können Suchmaschinen durchaus erfassen. Doch auch diese Alternativstrategien erfordern weitere, Kosten verursachende und den Arbeitsprozess verlängernde Man-/Woman-Power. Drittens kann die Verwendung von Text-Bildern auch zu visuellen Inkongruenzen im Schriftbild einer Webseite führen⁴⁹⁸; dann nämlich, wenn ungeglätteter HTML-/CSS-erzeugter Text neben Text-Bitmaps steht, in denen die Schrift geglättet aufbereitet ist.

Dass die Übertragungszeit für texthaltige Bitmaps größer ist als die für die schlichte Übermittlung von Textinhalten via HTML⁴⁹⁹, sei hier nur der Vollständigkeit halber und nicht mehr als wirklich stichhaltiges Argument gegen Text-Bilder angeführt; auch hier gilt, dass im Zeitalter von DSL dieser Vorbehalt nur noch bedingt abschreckende Wirkung entfalten kann. Schweizers Ratschlag, Text-Bilder aus Gründen der Übertragungszeiten möglichst nur für Überschriften einsetzen⁵⁰⁰, darf daher als überholt gelten. Auch die Anzahl jener Web-Besucher, die die Darstellung von Bildern in ihrem Browser deaktiviert haben – und denen damit der Genuss einer ausgefallenen Type entgehen könnte –, dürfte kleiner sein als von manchen Autoren heraufbeschworen⁵⁰¹.

Für Textpassagen, die häufigerer Aktualisierung unterliegen, sind Text-Bitmaps mithin keine echte Alternative. Hier bleibt wiederum meistens nur der Rückgriff auf die Standardschriftarten vom Kaliber *Times*, *Arial* und *Verdana*. Allerdings ist es, gerade auf den Webseiten etablierter Unternehmen mit elaboriertem Corporate Design, durchaus Usus, Textpassagen

⁴⁹² Vgl. Böhringer et al. 2003, S. 129; vgl. ferner Kiehn/Titzmann 1998, S. 31.; vgl. ferner Köhler 2002, S. 280

⁴⁹³ Vgl. Herrmann 2005, S. 251

⁴⁹⁴ Vgl. Lamprecht 2002, S. 38

⁴⁹⁵ Vgl. Gorbach 2001, S. 255

⁴⁹⁶ Vgl. Schäffer 2001, S. 190

⁴⁹⁷ Vgl. Seibert/Hoffmann 2006, S. 216

⁴⁹⁸ Balzert 2004, S. 176

⁴⁹⁹ Vgl. Kent 1999, S. 232; vgl. ferner Dengler/Volland 2000, S. 56; vgl. ferner Lankau 2000, S. 366; vgl. ferner Neutzling 2002, S. 32. Vgl. ferner Kommer/Mersin 2002, S. 124

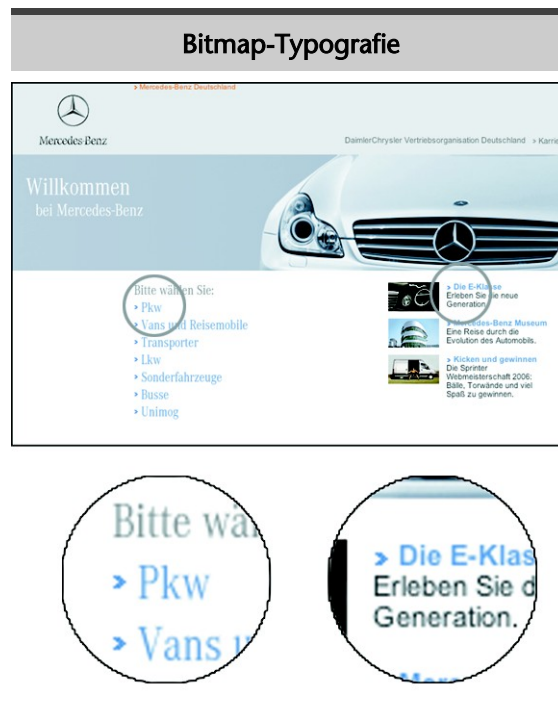
⁵⁰⁰ Vgl. Schweizer 2003, S. 158

⁵⁰¹ Vgl. Seibert/Hoffmann 2006, S. 277

mit längeren Aktualisierungszyklen als typografische Bilder anzubieten und der Webseite damit zumindest zum Teil ein unternehmensspezifisches Erscheinungsbild zu verleihen.

So sind beispielsweise die Menüpunkte in der Navigationsleiste auf der Startseite von DaimlerChrysler als Bilder abgelegt, die den jeweiligen Menüpunkt in der firmeneigenen Hausschriftart *Corporate* gepixelt darstellen (eine Schriftart also, die, zumindest offiziell, auf keinem einzigen Rechner vorgehalten werden dürfte, der nicht im weiteren Sinne mit dem Konzern in Verbindung steht). Inhalte dagegen, die häufigeren Wandlungen unterliegen, sind für die Darstellung in Systemschriften eingerichtet.

Hellbusch schließlich weist auf einen weiteren Nachteil von Bitmaps hin: Sie sind – abgesehen vom kaum verbreiteten Browser „Opera“ – nicht skalierbar. Menschen mit Sehbehinderung werden also potenziell von der Lektüre ausgeschlossen, was dem Ziel eines „barrierefreien Webdesigns“ entgegensteht.



Die Hauptnavigationen auf der DaimlerChrysler-Homepage sind als Bilder in der Hausschrift „Corporate“ platziert („Pkw“, „Vans“). Textpassagen mit kürzeren Aktualisierungszyklen hingegen („Die E-Klasse: Erleben Sie ...“) sind als browsereigene Darstellung konzipiert – was in diesem Falle zur Darstellung des Textes in der Schriftart *Arial* führt.

ABBILDUNG 3.17

3.4.4.3 Schriftsatz mit *Adobe Flash*

Ein weiterer potenzieller Ausweg aus der typografischen Enge der „web-echten“ Schriftarten ist der Einsatz des Plug-Ins *Flash*, das Web-Angebote abbilden kann, die mit dem Programm *Adobe Flash* erstellt worden sind. In Web-Filmen dieser Gattung sind Schriften beliebig einbindbar⁵⁰²; Gestalter können sich in der *Flash*-Umgebung also unbesorgt lösen von der Überlegung, welche Schriftarten auf den Zielrechnern mehrheitlich vorhanden seien. *Flash* garantiert zudem identisches Aussehen in allen Browsern⁵⁰³. Bei dieser Kurzbeschreibung belasse ich es an dieser Stelle – unter anderem deshalb, weil es sich im Sinne der Zielstellung der vorliegenden Arbeit bei Schriftsatz in *Flash* definitiv nicht um originäre „Browser-Typografie“ handelt. Da *Flash* allerdings eine bedeutende Rolle bei der Realisierung meines Experiments

⁵⁰² Vgl. Schweizer 2003, S. 162

⁵⁰³ Maurice/Rex 2006, S. 264

spiele (siehe Kapitel 4.7.2), sollen die Für und Wider des Einsatzes dieser Programmsorte an dieser späteren Stelle noch eingehender diskutiert werden.

3.4.5 Am Horizont: Neue *Windows*-Schriften

Ende September 2006 hieß es: für Privatkunden Anfang 2007. Ob *Windows Vista*, die seit Jahren angekündigte Nachfolgeversion des Betriebssystems *Windows XP* aus dem Hause *Microsoft*, in diesem besagten Monat wirklich erscheinen wird, muss Spekulation bleiben.

Fakt aber scheint zu sein: Mit dem Betriebssystem werden auch neue Schriftarten auf die Rechner der *Windows*-Kunden gelangen⁵⁰⁴, als sogenannte „ClearType Collection“⁵⁰⁵ – und sollte es *Microsoft* gelingen, sein Quasi-Monopol auch mit *Vista* aufrecht zu erhalten, dürfte der Pool der „web-echten“ Schriften in drei bis vier Jahren erheblich angewachsen sein.

Abbildung 3.18 zeigt die neuen Schriften; dabei ist nach allem, was zu lesen ist, die *Segoe* auserkoren als die neue „Haupt-Systemschrift“ von *Vista* (in die-

ser Schriftart werden also im neuen Betriebssystem Buttonbezeichnungen, Ordner und Allgemeintexte gesetzt sein), die übrigen Schriften werden sich in ihrer Beliebtheit in Briefsatz, E-Mail-Verkehr und Geburtstagsseinladungen beweisen müssen. Meines Erachtens dürfte auch Hauptkonkurrent *Apple* die neuen Fonts recht bald vorhalten; es könnte also durchaus sein, dass die Ära der *Verdana*-Dominanz in einigen Monaten ein recht abruptes Ende finden wird.

Eine detailliertere Diskussion der neuen *Microsoft*-Schriften soll an dieser Stelle jedoch nicht geführt werden. Es bleibt zunächst, wie gesagt, abzuwarten, ob und welche der typografischen Neulinge sich im Alltagstest und insbesondere im Web als konsensfähig und stetiger genutzt herausstellen werden.



⁵⁰⁴ Vgl. van Aaken 2006; vgl. ferner Shaikh et al. 2006

⁵⁰⁵ Vgl. Microsoft Corp. 2004

3.4.6 Empirische Studien zur Schriftart am Bildschirm

Ich habe bereits dargelegt, dass die Wirkung von Schriftarten am Bildschirm deutlich stärker und unmittelbarer als im Printbereich interdependiert mit der Größe (also der „Pixel-Menge“), in der die Buchstaben am Monitor dargeboten werden. Im Folgenden werden daher zunächst nur jene empirischen Studien vorgestellt, die sich eindeutig und isoliert mit dem Phänomen „Schriftart“ auseinandergesetzt haben. Untersuchungen, die zusätzlich den typografischen Faktor „Größe“ in die Forschung einbezogen, habe ich dagegen im Folgenden Hauptkapitel abgehandelt, das sich mit dem Aspekt der Schriftgröße befasst.

3.4.6.1 Lesegeschwindigkeit und Leseverständnis: 4 Schriftarten im Vergleich

Weisenmiller verglich in seiner bereits erwähnten Studie (vgl. Seite 66) die Lesegeschwindigkeiten sowie das Leseverständnis von Versuchspersonen. Typografisch aufbereitet war der Experimentaltext dabei in vier Varianten, von denen jeder seiner 264 Probanden eine vorgelegt bekam: Weisenmiller verglich im Detail zwei Schriftarten aus der Prä-Netz-Ära (nämlich *Times* und *Arial*) mit den beiden wohl bekanntesten, dezidiert für die Bildschirmdarstellung entwickelten Schriftarten (*Georgia* und *Verdana*). Alle Texte waren in 12 Punkt Größe gesetzt; ein Drittel der Teilnehmer las dabei den Text von Papier, die restlichen Probanden wurden mit dem Text an Computerbildschirmen konfrontiert. Dabei sah eine Hälfte der „Bildschirm-Leser“ den Text im „1-Bit-Modus“, also in „ungeglätteter“ Monitor-Aufrasterung, die andere Hälfte im „8-Bit-Modus“, also geglättet.

Im Zusammenhang mit der Diskussion dieser Arbeit sind die „1-Bit“-Ergebnisse Weisenmiller sicherlich die interessantesten – da, wie gezeigt, anzunehmen ist, dass diese ungeglättete Art der Schriftdarstellung am ehesten derjenigen entspricht, mit der sich die angenommene Mehrheit von Online-Nutzern im Web konfrontiert sieht. Überraschenderweise lagen in diesem Modus ausgerechnet die beiden Serifen-Schriften *Georgia* (213,59 gelesene Wörter in 60 Sekunden) und *Times* (198,09 Wörter) vor den Grotesken *Arial* (188,50) und *Verdana* (187,82)⁵⁰⁶. Die *Georgia* erzielte mithin ein um 13 Prozent besseres Ergebnis als die beiden serifenlosen Schriftarten. Als statistisch signifikant schneller lesbar als eine der jeweils anderen drei getesteten Schriftarten stellte sich dennoch keine Testvariante heraus.

Im 8-Bit-Modus dagegen kehrte sich das Verhältnis der Schriftfamilien um: hier erzielten die beiden serifenlosen Schriften die besten Resultate. Die *Arial* (239,82 gelesene Wörter) lag deutlich vor der *Verdana* (221,23), diese wiederum knapp vor der *Times* (219,14) und etwas klarer vor der *Georgia* (207,55). Statistisch signifikant schneller gelesen als eine der jeweils drei „Konkurrentinnen“ wurde allerdings auch hier keine der vier Schriftarten.

⁵⁰⁶ Vgl. Weisenmiller 1999, S. 79

In der Papierversion lag mit der *Times* ebenfalls eine Antiqua vorne (206,18), diesmal jedoch hauchdünn vor der *Verdana* (206,05) und unmerklich vor der *Georgia* (201,86). Die *Arial* lag auf Papier ganz hinten (190,45). Auch hier war keine wechselseitige statistische Signifikanz zwischen den vier Schriftarten feststellbar.

Ohne eindeutigen „Sieger“ blieb auch der Vergleich des Leseverständnisses, das Weisenmiller differenziert nach vorgelegter Schriftart mittels inhaltlicher Verständnisfragen ermittelte. Insgesamt erwiesen sich hier die ermittelten Durchschnitts-Differenzen als eher noch geringer als in der Messung der Lesegeschwindigkeiten.

Im 1-Bit-Modus lag die *Times* vorne (durchschnittlich 24,73 „Verständnis“-Punkte von 36 möglichen), gefolgt von *Arial* (23,82), *Verdana* (23,27) und *Georgia* (23,00). Im 8-Bit-Modus schnitt die *Arial* am besten ab (28,59), kurz vor der *Verdana* (28,00) und der *Times* (27,09). Erstaunlich abgeschlagen landete die *Georgia* hier auf Rang 4 (24,82).

Auf Papier wurden die besten Verständnisseleistungen bei Probanden gemessen, die den Text in der *Verdana* vorgelegt bekommen hatten (27,46) – was für sich genommen durchaus nicht ohne Reiz ist: Immerhin ist die *Verdana* dezidiert für Bildschirm-Anwendungen gezeichnet worden, entfaltete aber auf dem „fremden“ Medium Papier die beste Nachhaltigkeit. Insgesamt stellte Weisenmiller auch unter dem Kriterium der „Verständnismessung“ keinerlei statistisch signifikant „bessere“ Lesbarkeit einer der vier Schriften im wechselseitigen Vergleich zu den übrigen drei fest.

Weisenmiller stellt fest: „Die Beobachtungen, die den Daten der vorliegenden Studie zu entnehmen sind, führen zum Schluss, dass die bildschirmoptimierten Schriftarten (*Georgia* und *Verdana*) nicht notwendigerweise die Lesbarkeit verbessern“⁵⁰⁷.

3.4.6.2 Lesedauer von 12 Schriftarten

Bernard et al. untersuchten 2001 an 22 Versuchsteilnehmern im Durchschnittsalter von 25 Jahren (15 Frauen, 7 Männer), welche Schriften die Lesegeschwindigkeit beschleunigen oder dämpfen⁵⁰⁸. Sämtliche Schriften wurden in 12 Punkt Größe dargeboten an einem 17-Zoll-Monitor, jeder Teilnehmer las je einen Text in einer der zwölf interessierenden typografischen Konstellationen. Ob die Schriften in geglätteter oder 1-Bit-Darstellung präsentiert wurden, geht aus den Beschreibungen von Bernard et al. bedauerlicherweise nicht hervor; ebenfalls offen bleibt, ob ein Webbrowser zur Darstellung der Schriften verwendet wurde oder ein alternatives Darstellungs-Programm.

Insgesamt konnten Bernard et al. keinen statistisch signifikanten Effekt der Schriftart auf das Lesetempo feststellen. Im direkten Vergleich der insgesamt 12 präsentierten Schriftarten

⁵⁰⁷ Weisenmiller 1999, S. 119 (Übersetzung M.L.)

⁵⁰⁸ Vgl. Bernard et al. 2001c

wurde die *Tahoma* (wie erwähnt die etwas enger laufende Vorgängerin der *Verdana*) am schnellsten gelesen (270 Sekunden durchschnittliche gemessene Lesezeit) und war damit 11,3 Prozent und auch statistisch signifikant schneller rezipierbar als die *Corsiva*, die mit 306 Sekunden Lesezeit den insgesamt schlechtesten Durchschnittswert aufwies.

Nun ist die *Corsiva* definitiv nicht zu den verbreiteten Schriftarten im Web zu zählen. Daher sei der Blick ein wenig verengt auf jene von Bernard et al. getesteten Schriftarten, die im weiteren Sinne zu den „web-tauglichen“ Typen zu zählen sind; aufgrund ihrer Gestalt einerseits, ihrer Verfügbarkeit in gängigen *Windows*-Betriebssystemen wegen andererseits. Und hier schnitt wiederum unverhofft eine Serifenschrift gut ab: Am zweitschnellsten überhaupt wurde die *Times* gelesen (273 Sekunden), schneller als die *Verdana* (285 Sekunden) und die *Arial* (287). Überraschend abgeschlagen findet sich die *Georgia* auf Gesamtrang 9 wieder mit bemerkenswerten 289 Sekunden durchschnittlicher Lesezeit. Noch schlechter schnitt die *Courier* ab (302). Insgesamt jedoch ist zu konstatieren, dass sich alle untersuchten Schriftarten im Experiment in ihren Ergebnissen in einem sehr schmalen Korridor von 36 Sekunden bewegten, weshalb Bernard et al. auch feststellen: „(Diese) Differenzen haben keine allzu große Bedeutung für die meisten Web-Texte, sie waren dazu nicht ausgeprägt genug“⁵⁰⁹.

3.4.6.3 Individuelle Einschätzung der Lesbarkeit durch die Probanden

Im selben Experiment baten Bernard et al. ihre 22 Versuchsteilnehmer, die von ihnen vermutete Lesbarkeit der 12 vorgelegten Schriftarten auf einer Skala einzuschätzen (1 = gar nicht lesbar, 6 = sehr gut lesbar). Hier waren die Ergebnisse nicht nur deutlicher als im Test der Lesegeschwindigkeit, sondern auch überraschender: Als am besten lesbar stuften die Probanden ausgerechnet die Schreibmaschinen-Type *Courier* ein (Durchschnittswert 5,1), knapp gefolgt von *Verdana* und der Kinderschrift *Comic* (beide 4,9). Auf den Folgeplätzen landeten *Times* und *Arial* (jeweils 4,8), dicht dahinter die *Georgia* (4,7). Innerhalb dieser „Spitzengruppe“ konnten die Forscher allerdings wiederum keine statistisch signifikanten Unterschiede feststellen. Als am schlechtesten lesbar im Urteil der Teilnehmenden fielen die Schriftarten *Corsiva*, *Agency* und *Bradley* mit Durchschnittswerten um die „3“ herum durch.

3.4.6.4 Schriftart-Präferenzen von Nutzern

Schließlich erhoben Bernard et al. im angesprochenen Experiment die „allgemeine Präferenz“ der Teilnehmer in Bezug auf die Schriftarten; abgefragt wurde also die generelle Werteschätzung einer Schriftgestalt. Die meisten Teilnehmer gaben hier der *Arial* (Durchschnittswert 8,0 von maximal 10) den Vorzug vor allen anderen Schriften, mit geringem Abstand platzierten sich dahinter die *Georgia* (7,7) und die *Verdana* (7,6). Recht gut (alle im Bereich

⁵⁰⁹ Bernard et al. 2001c (Übersetzung M.L.)

um den Wert „7“) schnitten die nicht „web-typischen“ Schriftarten *Comic*, *Tahoma* und *Schoolbook* ab. Die *Courier* fand sich ebenfalls in dieser Marge (Wert 6,8). All die vorgenannten Schriftarten erwiesen sich als statistisch signifikant beliebter als die übrigen fünf, unter denen sich auch, deutlich abgeschlagen auf Rang 7, die *Times* befand – mit einem geradezu desaströsen „Beliebtheits-Wert“ von gerade einmal 4,8.

3.4.6.5 Einschätzungen von Schrift-Charakteristika durch Nutzer

Ferner befragten Bernard et al. ihre 22 Probanden nach „Charaktereigenschaften“ der 12 vorgelegten Schriftarten am Bildschirm. Ermittelt wurden Einschätzungen der Teilnehmenden, in welchem Maße die verglichenen Schriftarten „Persönlichkeit“ entfalteten, wie „eleganter“ sie wirkten, in welchem Maße sie „Business-Likelihood“ ausstrahlten und schließlich, wie weitgehend sie „Jugendlichkeit und Spaß“ vermittelten. Erbeten wurden dabei je Schriftart und untersuchtem Charakterzug Bewertungen von 1 („gar nicht“) bis 6 („vollständig“).

Ich habe die Ergebnisse dieser vier Einzel-Befragungen in Tabelle 3.2 zusammengefasst. Diese Art der Zusammenstellung erschien mir erhellender als die separierte Auflistung der Durchschnitts-Resultate nach „Eigenschaften“. Allerdings habe ich mir erlaubt, in der Tabelle lediglich die Ergebnisse für die etablierten „Web-Schriften“ aufzuführen sowie für die Schriftarten *Comic* und *Tahoma* – diese beiden sind immerhin Standardschriften im *Microsoft*-Betriebssystem *Windows* und damit potenziell für den Einsatz auf Webseiten verwendbar, wenn man sich mit einer nicht 100-prozentigen Abdeckung zufrieden gibt. Nicht aufgelistet sind damit die von Bernard et al. untersuchten Schriftarten *Bradley*, *Corsiva*, *Goudy*, *Schoolbook* und *Agency*, für die ich eine sehr geringe Verbreitung auf zeitgenössischen PCs unterstelle.

Schriftart	„personality“	„elegant“	„business“	„youthful“
<i>Arial</i>	3,1	2,5	3,7	2,6
<i>Verdana</i>	3,4	2,9	3,6	3,2
<i>Times</i>	2,6	2,6	4,8	1,9
<i>Georgia</i>	3,4	3,0	3,9	3,1
<i>Courier</i>	2,8	2,1	4,6	2,0
<i>Comic</i>	3,9	2,6	2,7	3,9
<i>Tahoma</i>	3,1	2,6	3,5	2,9

Quelle: <http://psychology.wichita.edu/surl/usabilitynews/3S/font.htm>

Auffallend ist an der Tabelle 3.2 insgesamt, dass die *Verdana* relativ stabile Werte in allen vier abgefragten Bereichen erzielt, dabei allerdings in der Summe sogar noch von der *Georgia* übertroffen wird. Herausstechend wiederum sind die Werte für die *Times* und die „Schreibmaschinenschrift“ *Courier*

im Bereich der „Business-Anmutung“. Einer bemerkenswerten Beliebtheit erfreut sich auch die „Kinderschrift“ *Comic* in allen Feldern außer „business“. Ein eindeutiger Sieger unter den verglichenen Schriftarten fand sich allerdings nicht; dennoch ist den Forschern um Bernard durchaus zuzustimmen: „Applying this information can help establish the proper mood of a particular site“⁵¹⁰.

⁵¹⁰ Bernard et al. 2001c

3.4.6.6 Die Schriftarten *Times*, *Arial* und *Courier* im Vergleich

Hill und Scharff verglichen 1997 die Lesbarkeiten der Schriftarten *Times*, *Arial* und *Courier* am Computermonitor in verschiedenen Farb-/Farbgrund-Kombinationen sowie in Kursiv- und Normalsatz⁵¹¹. Dem Gegenstand dieses Kapitels folgend, werden hier zunächst nur die Ergebnisse der Vergleiche der Schriftarten im Normalsatz rekapituliert.

43 Probanden wurden aufgefordert, verschiedene Texte zu „scannen“, die dargebotenen Texte also überblicksweise zu „überfliegen“, und dabei bestimmte Wörter zu finden, die ein grafisches Symbol verbalisierten (zum Beispiel „Stern“). Nach Auffinden dieses Symbol-Begriffs hatten die Teilnehmenden das entsprechende bildliche Symbol anzuklicken, das neben dem jeweiligen Text platziert war. Insgesamt schnitt die *Times* in den Reaktionszeiten dabei etwas besser ab als die beiden anderen Schriftarten. Die *Courier* erzielte überraschenderweise noch bessere Werte als die *Arial*; es zeigte sich allerdings kein einziger statistisch bedeutsamer Unterschied der verglichenen Schriftarten untereinander.

3.4.6.7 Die Schriftarten *Verdana* und *Georgia* in verschiedenen Vergleichskonstellationen

Boyarski et al. untersuchten 1998, kurz vor Einführung des Betriebssystems *Windows 98*, die beiden wichtigsten neuen Schriften, die *Microsoft* mit dem neuen Betriebssystem in seinen Standard-Schriftsatz einzubinden versprach: *Verdana* und *Georgia*⁵¹². Sie verwandten dieselbe Methodik wie Weisenmiller (siehe oben), um die *Georgia* mit der *Times* (der bis 1998 wohl vorherrschenden Schriftart im Web), die *Georgia* mit der *Verdana* sowie die *Verdana* im normalen und im kursiven Schnitt auf Lesegeschwindigkeit und die Unterstützung des Inhaltsverständnisses hin zu vergleichen. Dabei ergab sich praktisch kein Unterschied zwischen *Georgia* und *Times* in der durchschnittlichen Lesegeschwindigkeit der 16 Versuchsteilnehmer (376 zu 377 Sekunden Lesezeit); in der Messung des Leseverständnisses schnitt die *Times* sogar etwas besser ab als die Neuschöpfung (10,2 zu 9,8 richtige Antworten im Schnitt). Jedoch: In der Befragung der Probanden nach ihrer persönlichen Einschätzung der beiden Schriften lag eindeutig die *Georgia* vorne. Sie wurde als besser lesbar empfunden, als schärfer in der Kontur und allgemein leichter identifizierbar als die *Times*.

Deutlich besser als in Gegenüberstellung zur *Times* schnitt die *Georgia* im Vergleich mit der *Verdana* ab; sie wurde um etwa 4 Prozent schneller gelesen (363 zu 377 Sekunden), auch bei den korrekten Antworten auf die Inhaltsfragen schnitt die Serifenschrift besser ab (9,8 zu 9,0). Die Probanden schätzten die *Verdana* als etwas „leichter lesbar“ ein; beide Schriften wurden jedoch durchschnittlich als gleich „scharf“ und „erkennbar“ eingestuft.

⁵¹¹ Vgl. Hill/Scharff 1997

⁵¹² Vgl. Boyarski et al. 1998. S. 87ff.

3.4.6.8 Bevorzugte Schriftarten im Web

Shaikh et al. untersuchten 2006 in einer beachtlich umfangreichen Online-Studie⁵¹³ den „Appeal“, die „Persönlichkeit“ von 20 verschiedenen Schriftarten am Bildschirm mit insgesamt über 1000 Probanden. In einem ersten Teil wurden mithilfe von 15 Adjektiv-Pärchen die Charakteristika der Schriftarten untersucht (beispielsweise hatten die Teilnehmenden jede Schriftart danach zu beurteilen, ob sie eher „jugendlich“ wirke oder „erwachsen“, „maskulin“ oder „feminin“, „traurig“ oder „fröhlich“; zwischen beiden Polen konnten die Probanden jeweils in vier Abstufungen werten). Eine detailliertere Darstellung der Ergebnisse soll an dieser Stelle unterbleiben – 15 Adjektiv-Pärchen für 20 Schriftarten ergeben immerhin alleine 300 Konstellationen! Im Folgenden beleuchte ich daher nicht die Ergebnisse aller 20 Schriften im Einzelnen, sondern unterteilt in stilistische Gruppen; und auch dies nur sehr knapp.

Interessant ist vor allem, dass die serifenlosen Schriften (unter anderem *Arial* und *Verdana*) fast durchgehend nur sehr „unentschiedene“ Wertungen erhielten; kaum eines der abgefragten „Persönlichkeits-Merkmale“ wurde in dieser Stilgruppe als sonderlich ausgeprägt empfunden. In diesem Sinne fielen die Grotesken in der Studie also als weitgehend „seelenlos“ auf. Anders die Serifenschriften (unter anderem *Times* und *Georgia*): Diese wurden am höchsten bewertet in Eigenschaften wie „stabil“, „praktisch“, „erwachsen“ und „formalistisch“.

Schriften mit „Schreibschrift“- oder „Spaß“-Charakter (darunter die *Comic*) hatten die höchsten Rating-Werte für die Charakteristika „jugendlich“, „fröhlich“, „rebellisch“, „feminin“, „zwanglos“ und „knuddelig“. Die Schmuckschriften (unter anderem die sehr schmallettrige „Impact“) erfuhren hohe Bewertungen wie „maskulin“ und „selbstbewusst“, jedoch auch eher unschmeichelhafte Belegungen wie „unhöflich“, „traurig“ und „grob“ (inwiefern diese drei Eigenschaften wiederum als „maskulin“ einzustufen seien, wurde nicht weiter hinterfragt). Die Monospace-Schriftarten (Schreibmaschinenschriften) schließlich wirkten auf die Teilnehmenden „dumm“, „einfach“, „fantasielos“ und „angepasst“.

In einer zweiten Teilbefragung wurden andere Probanden von Shaikh et al. gebeten, anzugeben, welche der 20 vorgestellten Schriftarten sie am ehesten verwenden würden für bestimmte Dokumenttypen im Web. Einen wirklichen Achtungserfolg erzielte hier die *Times*: Diese war die bevorzugte Schrift der Nutzer in 14 von 25 Verwendungs-Kategorien, in zwei weiteren Kategorien war sie die durchschnittlich zweite Wahl. Unter anderem empfanden die Teilnehmer die *Times* als am angemessensten für „Web Texts“, „Web Headlines“, „Technical Docs“, „E-Books“, „E-Magazines“, „E-Scientific Articles“ und „Online News“. Von den übrigen gängigen Web-Schriftarten landete die *Verdana* fünf Mal auf Rang zwei in der Präferenz (unter anderem „Instant Messaging“ und „Math Docs“), die *Arial* war durchschnittlich

⁵¹³ Vgl. Shaikh et al. 2006; vgl. ferner Shaikh et al. 2005

dritte Wahl bei „Web Headlines“, die *Georgia* eroberte insgesamt einen Spitzenrang („Online Brochures“).

3.4.6.9 *Times* versus Schmuckschrift

In einem Experiment mit 25 Teilnehmern (13 Männer, 12 Frauen) im Alter von 19 bis 63 Jahren untersuchten Morrison und Noyes 2003 die Lesbarkeiten der *Times* und einer sehr „verschnörkelten“ Schmuckschrift namens *Gigi* am Bildschirm⁵¹⁴. Beide Schriften wurden in 12 Punkt Größe am Monitor präsentiert; die Teilnehmenden lasen zwei Texte von je 140 Wörtern in jeweils einer der beiden Schriftarten und dokumentierten das Ende ihres Rezeptionsprozesses jeweils, indem sie vernehmlich „Stop“ riefen. In diesem Moment wurde die Stoppuhr angehalten und damit die Lesezeit ermittelt.

Dabei wurden für die Texte, die in der *Times* gesetzt waren, durchschnittlich 22,06 Sekunden Lesezeit ermittelt, für die *Gigi* 26,99 Sekunden. Die Differenz war statistisch signifikant. Ferner hatten Morrison und Noyes in jedem Text zehn inhaltlich inkohärente „Ersatzwörter“ platziert, die die Probanden ebenfalls durch lautes Aussprechen während des Leseprozesses zu identifizieren hatten; aus der Anzahl identifizierter Wörter folgerten die Forscher das inhaltliche Textverständnis seitens der Probanden, das für die in *Times* gesetzten Texte (8,80 gefundene Wörter im Schnitt) wiederum statistisch signifikant besser ausfiel als für die in *Gigi* dargebotenen Passagen (6,68).

Befragt nach ihrer individuellen Einschätzung der beiden verglichenen Schriftarten, erklärten alle Probanden, sie würden für den Eigengebrauch die *Times* verwenden. 68 Prozent erklärten, sie hielten die Verwendung der *Times* auf Webseiten für angemessen, 32 Prozent bevorzugten die *Gigi*. Als „attraktiver“ empfanden 80 Prozent der Teilnehmer die *Gigi*, 20 Prozent die *Times*.

3.4.6.10 Zwei neue Schriftarten im Test

Chaparro et al. untersuchten 2006 an 10 Probanden⁵¹⁵ (5 Frauen, 5 Männer) im Alter zwischen 18 und 35 Jahren die Identifizierbarkeit von Buchstaben, Satzzeichen und mathematischen Symbolen auf dem Bildschirm in drei Schriftarten. Die Buchstaben schienen einzeln und nacheinander auf, gemessen wurde jeweils die Zeitspanne, die vom Erscheinen bis zur verbalen Identifikation des Zeichens durch den Teilnehmenden verstrich.

Chaparro et al. maßen die Reaktionszeiten für die Zeichen in drei Schriftarten, von denen zwei noch gar nicht offiziell auf dem „Markt“ sind: *Cambria* und *Constantia*, kriert für die

⁵¹⁴ Vgl. Morrison/Noyes 2003

⁵¹⁵ Chaparro et al. 2006

Einpassung in den Standardschriftensatz des anstehenden neuen *Windows*-Version *Vista* (siehe Seite 113). Die Schriftarten wurden in ihrer Wirkung verglichen mit der *Times*.

In der Identifizierbarkeit von Buchstaben fanden die Forscher dabei keine Unterschiede zwischen den Schriftarten. Mathematische Zeichen hingegen und Satzzeichen scheinen in der *Cambria* deutlich besser erkennbar zu sein als in den beiden Vergleichsschriften.

3.4.6.11 Schriftglättung mit „ClearType“

Dillon et al. untersuchten 2006 die Auswirkungen der Glättungs-Technologie „ClearType“ aus dem Hause *Microsoft* auf Lesegeschwindigkeit, Textverständnis und individuell erfahrene Ermüdung von Probanden⁵¹⁶. Es ging also darum, ob Bildschirmschriften in „geglätteter“ Form besser rezipierbar seien als ungeglättete.

Dillon et al. fanden, dass die 26 Probanden die vorgelegten Texte im Schnitt 8 Prozent schneller lasen, wenn „ClearType“ aktiviert war. Allerdings fiel die Lesegenauigkeit (überprüft durch inhaltliche Fragen zum Testtext) bei den ungeglätteten Texten sehr leicht besser aus. Befragt wurden die Teilnehmenden zudem nach der individuell empfundenen „Ermüdung“ nach dem Lesen der „ClearType“- beziehungsweise der ungeglätteten Texte; hier stellten die Forscher keine Unterschiede fest. Eine Messwiederholung mit 25 weiteren Teilnehmern ergab vergleichbare Resultate.

3.4.6.12 Ein älterer Schriftglättungs-Test

In einem zweiten Teil der bereits vorgestellten Studie von Boyarski et al.⁵¹⁷ befragten die Forscher 16 weitere Teilnehmer nach ihren visuellen Eindrücken von Schriften, die mithilfe des Schriftglättungssystem von *Microsoft* auf den Bildschirm gebracht wurden; sie baten dabei um Vergleiche zu ungeglättet aufbereiteten sowie zu Schriften, die mit dem Konkurrenzsystem der Firma *Adobe* geglättet waren. Insgesamt gaben die Probanden tendenziell eher den geglätteten Fonts den Vorzug vor den ungeglätteten Schriften; im Vergleich von *Adobe*- und *Microsoft*-Glättungsmodus war praktisch kein Unterschied in den Urteilen feststellbar.

⁵¹⁶ Dillon et al. 2006. S. 12ff.; vgl. ferner Dillon et al. 2004, S. 1ff.

⁵¹⁷ Vgl. Boyarski et al. 1998, S. 87ff.

3.5 Schriftgröße

3.5.1 Die Schriftgröße als typografisches Phänomen

3.5.1.1 Der Punkt: Die wackelige Maßeinheit der Typografen

Die typografische Größenmessung widersetzt sich bis heute ebenso hartnäckig wie erfolgreich dem metrischen System – daran hat auch die vermeintliche digitale Revolution nur wenig ändern können. Fakt ist: Die Maßeinheit für Schriftgröße, Zeilenabstand, oft auch immer noch für Zeilenbreiten ist und bleibt der „Punkt“ oder dessen nächsthöhere Einheit „Pica“. Was nicht weiter tragisch wäre – wenn die Einheit, analog dem metrischen System, eindeutig definiert wäre. Doch der „Punkt“ kennt mindestens vier metrische Entsprechungen. Eher fünf.

Die Maßeinheit „Punkt“ ist zunächst eine print-spezifische. Sie geht zurück auf den französischen Typografen Pierre Simon Fournier (1712 – 1768), der 1737 als Vereinheitlichung der zu diesem Zeitpunkt über 100 europäischen fußbasierten Maßsysteme ein System aus dem „Cicero“ als Haupt- und dem „Punkt“ als Stückelungseinheit vorschlug⁵¹⁸, wobei ein Cicero aus 12 Punkt bestand. Fourniers Punkt maß umgerechnet 0,34882 Millimeter, krankte aber an einem zentralen Problem: Er stand genauso als Solitär-Einheit, wie es im Grunde alle anderen Maßsysteme jener Zeit taten. Da half auch die vermeintlich logische Herleitung der neuen Einheit nicht viel (es handelte sich bei Fourniers Punkt um ein 864-stel der Fußlänge des Königs von Frankreich⁵¹⁹): Fourniers Punkt war nur eine weitere Maßeinheit unter vielen, ohne nachvollziehbare Koppelung an ein vernünftiges Referenzmaß.

Diesen Makel suchte Fourniers Landsmann, der Schriftsetzer Francois-Ambroise Didot, zu tilgen, indem er Ende des 18. Jahrhunderts seinen „Didot-Punkt“ deutlich nachvollziehbarer an den damaligen französischen Fuß koppelte⁵²⁰: 1 französischer Fuß beinhaltete in Didots System 12 französische Zoll, 1 Zoll 12 Linien, 1 Linie 12 Punkte – und 2 Punkte (im Plural) definierte Didot schließlich als einen typografischen Punkt (p)⁵²¹. Umgerechnet betrug Didots Punkt 0,376065 Millimeter. Er wurde 1784 in Frankreich zur Vorschrift⁵²², die US-amerikanische „Type Founder’s Association“ folgte 1886, die britische 1898⁵²³.

Spätestens mit Einführung und Verbreitung des metrischen Systems zunächst im nachrevolutionären Frankreich, später in ganz Europa bot sich die Chance, auch das typografische Maßsystem in ein metrisches Dezimalsystem zu überführen. Didots Sohn und Schriftkünstler Fir-

⁵¹⁸ Vgl. Bergner 1990, S. 25

⁵¹⁹ Vgl. Gulbins/Kahrman 1992, S. 50

⁵²⁰ Vgl. McLean 1992, S. 21

⁵²¹ Vgl. Rommen 1989, S. 229f.

⁵²² Vgl. Kiehn/Titzmann 1998, S. 41

⁵²³ Vgl. McLean 1992, S. 73

min höchstpersönlich machte sich 1811 dafür stark. Jedoch: „Die Reform scheiterte (...) an den mit dieser Umstellung verbundenen gewaltigen Kosten“⁵²⁴.

Doch nicht nur dies. Unübersichtlicher wurde die Szenerie ein gutes Jahrhundert nach Didot dem Älteren durch den sogenannten „Pica-Punkt“ oder „point“ (pt), der 1886 mit der Zeilengussmaschine der Firma Linotype auch nach Europa kam⁵²⁵ und mit einer Größe von 0,351459 Millimetern (das entsprach und entspricht in etwa, aber nicht genau einem 72-stel des amerikanischen Zoll) um immerhin 6,5 Prozent unter dem Didot-Punkt lag⁵²⁶. Eine 14-Punkt-Schrift im Didot-System entspricht also rein rechnerisch nahezu exakt einer 15-Punkt-Schrift im Pica-System. Dieser Dualismus währte wiederum ein gutes Jahrhundert.

Das Jahr 1973 brachte dann eine vermeintliche Vereinfachung des Wirrwarrs, zumindest auf europäischer Ebene: die – nun „typografischer Punkt“ genannte – kleinste Einheit wurde (durch die damalige Europäische Gemeinschaft) auf exakte 0,375 Millimeter bestimmt⁵²⁷. Im Grunde sollte die 73er-Reform sogar noch weiter gehen: Offiziell wurde mit der Verordnung das metrische Maß nun auch in der Welt der Typografie Standard, der „Punkt“ also hoch offiziell zur Hilfseinheit degradiert. Spätestens in den 1980er Jahren durfte dieser Ansatz aber als gescheitert betrachtet werden⁵²⁸ – die Typografen blieben mehrheitlich im Punkt-System⁵²⁹.

Ein Grund für diesen neuerlichen Reform-Fehlschlag dürfte im Aufkommen der ersten leistungsfähigen Rechner mit optischen Gestaltungsfunktionalitäten zu finden sein, die Anfang, Mitte der 1980er Jahre auf den Markt traten. Mit ihnen kam nicht das metrische System (warum auch, Marktführer *Apple* und seine Computer eroberten immerhin von den USA aus den europäischen Markt, von einem angelsächsischen Staat aus also mit angelsächsischem Maßsystem), sondern – nochmals ein neu bemessener „Punkt“, der heute als der maßgebende angesehen werden muss. Hier tritt also nach Fournier-, Didot-, Pica- und typografischem Punkt die fünfte „Währungseinheit“ auf die Rampe – der sogenannte „DTP-Punkt“ (DTP = „desktop publishing“, sinngemäß: computergestütztes Publizieren, als Wortschöpfung vermutlich zurückgehend auf Seybold⁵³⁰) oder „dot“, der per definitionem einem exakten 72-stel des amerikanischen Zoll entspricht und damit 0,352778 mm misst⁵³¹. Diese – und nicht die Pica-Einheit, wie Kiehn und Tietzmann behaupten⁵³² – ist das Standardmaß nicht nur für heutige Bildschirmausgaben, sondern auch relevante Einheit in populären Satzprogrammen wie *QuarkXPress* oder *InDesign*⁵³³.

⁵²⁴ Rommen 1989, S. 229

⁵²⁵ Vgl. Bergner 1990, S. 25

⁵²⁶ Vgl. Gulbins/Kahrman 1992, S. 336f.

⁵²⁷ Vgl. Meissner 1992, S. 33

⁵²⁸ Vgl. Gulbins/Kahrman 1992, S. 53f.

⁵²⁹ Vgl. Böhringer et al. 2003, S. 101

⁵³⁰ Vgl. Seybold 1987, S. 149

⁵³¹ Vgl. Neutzling 2002, S. 34

⁵³² Vgl. Kiehn/Tietzmann 1998, S. 42

⁵³³ Vgl. Bollwage 2001, S. 88

3.5.1.2 Unverwüstlich: der Bleikegel

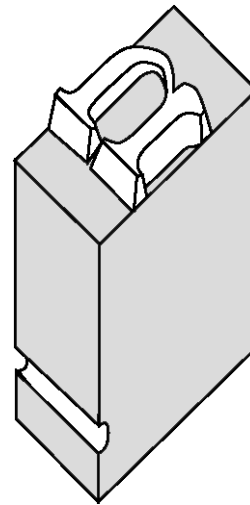
Es ist absehbar, dass sich der DTP-Punkt als Maßeinheit langfristig als Standard durchsetzen wird (wenn er es im Jahre 2006 nicht längst ist). Nicht nur die bereits erwähnten Satzprogramme wie *InDesign* und *QuarkXPress*, sondern auch Bildbearbeitungssoftware wie *Photoshop* oder Zeichentools wie *Illustrator* basieren sämtlich auf dieser Maßeinheit und sind inzwischen weltweit so verbreitet, dass ein Zurück beispielsweise in die „Didot“-Zeiten mehr als unwahrscheinlich ist.

Doch auch die Annahme des DTP-Punkts als zeitgenössischem Quasi-Standard wird, und das ist der Wermutstropfen, die Verlässlichkeit der Größenbestimmung von Schriften kaum erhöhen. Denn ein weiterer Anachronismus lebt fort im DTP-Zeitalter: der Kegel.

Obgleich nämlich der Bleisatz längst Geschichte ist, bemisst sich die Größe einer Schrift bis heute nicht etwa an der Höhe der Versalien (also der Großbuchstaben), auch nicht an der Höhe der Gemeinen (der Kleinbuchstaben oder Minuskeln also, wie es Rehe schon 1981 ersehnte⁵³⁴), auch nicht an der vertikalen Differenz zwischen maximaler Ober- und Unterlänge einer Schrift (wie Kommer und Mersin irrtümlich behaupten⁵³⁵), sondern an der – inzwischen fast immer nur noch virtuellen – „Kegelhöhe“⁵³⁶ oder auch Kegel-„Konsole“⁵³⁷ (vergleiche Abbildung 3.19).

Der Kegel war und ist im Bleisatz der metallische Quader oder auch „Stempel“, auf dessen einer verkürzten Seite die einzelne Letter als Prägung aufgebracht ist⁵³⁸; dabei war und ist diese Letter aber selbst bei den ausgesprochen seltenen und auch nur in einigen Alphabeten vorhandenen Lettern, die sowohl Ober- als auch Unterlängen aufweisen (also beispielsweise dem kursiven „f“ in der Schriftart *Times*, siehe Abb. 3.20), genau so hoch wie der Kegel als solcher⁵³⁹. Jeder Kegel hält also bereits einen schriftimmanenten Abstand (Schriftsetzer sprechen hier auch von der „Achselfläche“) nach oben und unten bereit, quasi einen „eingebauten“ Zeilenabstand⁵⁴⁰. In diesem Sinne ist es in fast allen Fällen nicht richtig, wenn Balzert behauptet, dass bei „Zeilen, die ohne Durchschuss gesetzt sind, (...) (sich) die Unter- und Oberlängen zweier Zeilen (berühren)“⁵⁴¹. Selbst wer also in einem Satzprogramm eine *Times* in 12 Punkt setzt mit 12 Punkt Zeilenabstand, wird immer noch vertikalen Weißraum vorfin-

Der Bleikegel



Skizze eines Bleikegels mit der Letter „n“.

ABBILDUNG 3.19

⁵³⁴ Vgl. Rehe 1981, S. 30

⁵³⁵ Vgl. Kommer/Mersin 2002, S. 193

⁵³⁶ Vgl. Bergner 1990, S. 28

⁵³⁷ Vgl. Luidl 1989, S. 62

⁵³⁸ Vgl. McLean 1992, S. 58

⁵³⁹ Vgl. Köhler 2002, S. 86

⁵⁴⁰ Vgl. Luidl 1989, S. 38f.

⁵⁴¹ Balzert 2004, S. 174

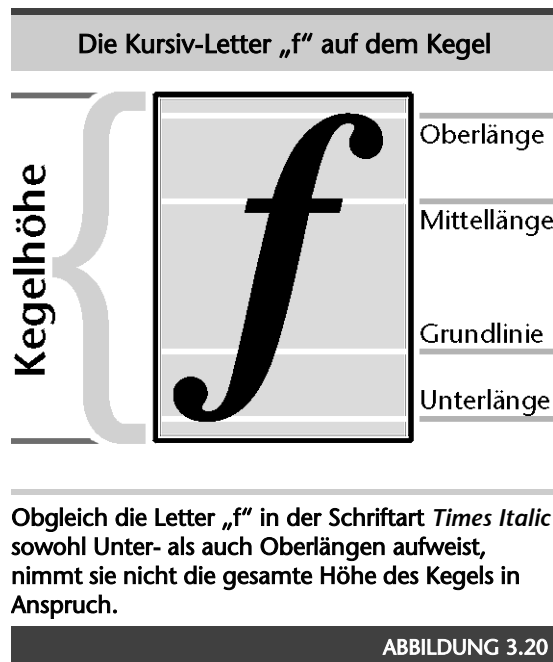
den zwischen der Unterlänge eines „g“ in der oberen und der Oberlänge eines „A“ in der darunterliegenden Zeile.

Es entbehrt also nicht eines gewissen Anachronismus', wenn sogar Neuschöpfungen wie die Schriftart *Verdana*, die explizit für die Bildschirm-Ausgabe konzipiert wurden, also höchstens aus Nostalgie je auf einen Metall-Kegel geprägt werden dürften, einem virtuellen Kegel aufsitzen. Das tun sie nichtsdestotrotz.

Nun mag die Altertümlichkeit der Schriftgrößen-Bemessung auf Kegel-Grundlage noch als Schrulle der Schriftsetzer-Gilde durchgehen. Was die Angelegenheit jedoch stark erschwert, ist die Tatsache, dass die Relation beispielsweise

von Großbuchstaben-Lettern auf Bleikegeln zur Kegelhöhe selbst nicht normiert ist. Die Folge: Die Letter „A“ kann auf den Kegeln einer Schriftart zwei Drittel, auf den numerisch gleich großen Kegeln einer anderen Schriftart aber gute drei Viertel der Konsolen-Höhe beanspruchen⁵⁴². Dies führt dazu, dass die effektive Schriftgröße beispielsweise der *Times*, deren Versalhöhe in großer Auflösung 66,2 % der Kegelhöhe beträgt und deren x-Höhe 47,4 % der Kegelhöhe, bei nominell gleicher Schriftgröße deutlich kleiner ausfällt als die *Arial*, deren vergleichbare Versalhöhe 71,6 % und deren x-Höhe 54,4 % der Kegelhöhe beträgt.

Das heißt: Die Großbuchstaben der *Arial* sind, bei numerisch absolut identischer Schriftgröße, über 8 Prozent größer als die der *Times*, die Minuskeln (also die Kleinbuchstaben, an denen die x-Höhe gemessen wird, die auch gelegentlich als „n-Höhe“ bezeichnet wird⁵⁴³) gar um nahezu 15 % größer. Da die Höhe der Mittellängen für die „optische Größe“ einer Schrift eine bestimmendere Rolle spielt als die Versalhöhe (siehe nächstes Kapitel: „Die optische Größe“), muss also eine *Times*, die in etwa so groß wirken soll wie eine in 40 Punkt gesetzte *Arial*, etwa 12 Prozent größer, mithin in etwa



⁵⁴² Vgl. Gulbins/Kahrman 1992, S. 52f.

⁵⁴³ Vgl. Gulbins/Kahrman 1992, S. 349

45 Punkt gesetzt werden (vgl. Abb. 3.21). Poulton sah sich in einer Studie 1924 gar gezwungen, zur optischen Größen-Angleichung zweier zu vergleichender Schriftarten diese Schriften in Größen einzurichten, die sich um sage und schreibe 2,5 Punkt unterschieden⁵⁴⁴.

Angaben, gar Empfehlungen zur zu verwendenden Schriftgröße müssen daher immer nur von eingeschränktem Erkenntniswert bleiben, wenn sie die verwendete Schriftart und deren Eigenschaften verschweigen⁵⁴⁵.

3.5.1.3 Auch das noch: die optische Größe

Selbst in deutschsprachigen Texten, denen – im internationalen Vergleich – meist überproportional viele Hauptwörter und damit relativ zahlreiche Großbuchstaben zu eigen sind⁵⁴⁶, dominieren in aller Regel die Minuskeln. Das Gesamtbild einer Schrift wird damit vorrangig geprägt von den Kleinbuchstaben⁵⁴⁷: Sie machen den „Grauwert“ eines Textes aus, bestimmen seine optische „Schwere“ mit⁵⁴⁸. Wie hoch dagegen die Versalien als „Leuchttürme“ aus dem Textgesamt herausragen, spielt eher eine nachgeordnete Rolle.

Laut Rehe determiniert die x-Höhe die sogenannte „optischen Größe“ einer Schrift⁵⁴⁹, die mithin als weiterer beachtenswürdiger Faktor neben die nominale Schriftgröße (also die Kegelhöhe) und die physische Größe der Letter auf dem Kegel tritt. Schriften mit kleiner x-Höhe wirken insgesamt kleiner, Schriften mit großer x-Höhe größer⁵⁵⁰. Gleichzeitig fordern Schriften mit kleiner x-Höhe weniger Zeilenabstand, da sie gleichsam über einen eingebauten „Puffer“, immanenten Weißraum also zwischen x-Marke und Kegeloberkante verfügen⁵⁵¹.

3.5.2 Manipulation von Schriftgrößen in HTML-basierten Browsern

Wie aber bekommt nun der Webdesigner eine gewählte Schrift in der gewünschten Größe auf den Bildschirm des Web-Users? Die Antwort mag an dieser Stelle der Diskussion kaum noch zu überraschen: In absolut zuverlässiger Form kann er oder sie das genauso wenig, wie er eine fixe Schriftart für die Seite zu definieren vermag.

Denn selbst die vermeintlich eindeutige Schriftgrößen-Anweisung eines Webdesigners per CSS, dass eine Schriftart in „12pt“, also in 12 Punkt Größe, im Browser zu erscheinen habe, reicht in ihrer Verbindlichkeit nicht über den Rang einer Empfehlung hinaus, die kundige

⁵⁴⁴ Vgl. Poulton 1924, S. 358ff.

⁵⁴⁵ Vgl. Lynch/Horton 1999, S. 88

⁵⁴⁶ Vgl. Friedl 1998, S. 24

⁵⁴⁷ Vgl. Balzert 2004, S. 181

⁵⁴⁸ Vgl. Kommer/Mersin 2002, S. 176

⁵⁴⁹ Vgl. Rehe 1981, S. 30; vgl. ferner Lynch/Horton 1999, S. 88

⁵⁵⁰ Vgl. Kiehn/Titzmann 1998, S. 43

⁵⁵¹ Vgl. Poulton 1955, S. 360

Webnutzer schlicht missachten und leicht unterlaufen können – indem sie im Menü ihres Browsers die Schriftart als „größer“ oder „kleiner“ definieren. In diesem Sinne können Webgestalter bis heute keine absoluten Größen setzen, sondern höchstens Größen-Hierarchien, die das wechselseitige Verhältnis unterschiedlich groß definierter Textelemente auf einer Webseite regeln. Lediglich eine Ausnahme existiert: der *Internet Explorer*. In diesem Browser nämlich sind Größenangaben, die in der Einheit „Pixel“ (und nur in dieser Einheit!) definiert werden, durch den Benutzer nicht mehr manipulierbar⁵⁵². Ob dies Strategie oder lediglich ein weiterer der vielen „Bugs“ (Fehler) des *Microsoft*-Browsers ist, sei dahingestellt – offizieller W3C-Standard ist es in jedem Falle nicht.

Doch warum eigentlich nicht? Das Stichwort lautet: „Barrierefreiheit“. Tatsächlich würde eine Webseite mit fix definierten Schriftgrößen Menschen mit Sehbehinderung potenziell ausschließen vom Konsum. Denn: Diesen Nutzern die Möglichkeit vorzuenthalten, die Schriftgröße ein, zwei Stufen heraufzusetzen, kommt einem faktischen Ausschluss dieser Personengruppe gleich⁵⁵³. Das W3-Consortium hat – bereits im Jahre 1999 – eine „Recommendation“ namens „Web Content Accessibility Guidelines 1.0“ herausgegeben⁵⁵⁴, in dem sich Hinweise und Ratschläge finden zur behindertengerechten Aufbereitung von Web-Informationen. Die Verwendung von durch Nutzer manipulierbaren Größen-Anweisungen mithilfe von CSS findet sich in der W3C-Empfehlung in Abschnitt („Guideline“) unter „Checkpoint“ 3.4.

Doch nicht nur unter dem Gesichtspunkt der Barrierefreiheit spricht einiges dafür, Nutzern Verfügungsgewalt über die Größeneinstellungen auf Webseiten zuzubilligen. Hinzu tritt, dass der klassische Computer-Monitor längst nicht mehr die einzige Plattform darstellt, auf der Webseiten zu begutachten wären. Immer mehr Nutzer „surfen“ auch via Handy oder über PDA – mithilfe von Ausgabemedien also, die auf völlig anderen Monitor-Formaten basieren⁵⁵⁵ (Mobiltelefone beispielsweise haben meistens hochformatige Displays, Computerbildschirme dagegen querformatige⁵⁵⁶). Vor allem aber sind die Sichtflächen dieser Endgeräte schlichtweg kleiner, basieren auf Pixel-Matrizen kleinerer Einheitenmenge; was also auf einem 17-Zoll-Monitor als angemessene Schriftgröße durchgehen kann, wirkt auf Handy-Displays potenziell völlig überdimensioniert. Auch unter diesem Aspekt ist es zweifellos sinnvoll, Nutzern die Manipulation von Schriftgrößen anheimzustellen. Es bedarf an dieser Stelle wohl keiner extensiveren Darlegung mehr, dass Designer von Webseiten sich ob dieser neuerlichen Kontrollbeschneidung die Haare raufen.

⁵⁵² Crederholm 2006, S. 38

⁵⁵³ Vgl. Lazar 2005, S. 148

⁵⁵⁴ Vgl. World Wide Web Consortium (W3C) 1999

⁵⁵⁵ Vgl. Seibert/Hoffmann 2006, S. 32

⁵⁵⁶ Vgl. Peck 2000, S. 168ff.

3.5.2.1 Die Definitionseinheiten in CSS

Dass in Bezug auf die Definition bereits der schlichten print-basierten Einheit „Punkt“ seit Jahrhunderten ein großer Wirrwarr herrscht, wurde weiter oben bereits dargelegt. Dass CSS eine ungleich größere Menge als definierbare Einheiten kennt, macht das Geschäft des Webdesigns auch in diesem Feld nicht einfacher.

Dabei ist zu unterscheiden zwischen zwei Hauptgruppen von Maß- oder Längeneinheiten, die CSS-basierte Browser zu interpretieren in der Lage sind: absolute und relative. Dass die vermeintlich „absoluten“ Angaben tatsächlich sehr „wackelig“ sind, wurde bereits angedeutet und soll weiter unten vertieft werden. Zunächst jedoch seien die verfügbaren Längeneinheiten einmal vorgestellt.

Dabei sind die als „absolut“ beschriebenen Maßeinheiten, grob umrissen, jene, die aus dem Prä-Computerzeitalter überkommen sind: nämlich das amerikanische Zoll, also der Inch (abgekürzt in CSS: „in“), Zentimeter („cm“), Millimeter („mm“), Punkt („pt“) sowie Pica als nächsthöhere Einheit im Punkt-System („pc“; ein Pica entspricht dabei, wie gezeigt, 12 Punkt). Derlei absolute Maßeinheiten sind freilich nach Crederholm „nur bei Print sinnvoll oder wenn die physischen Dimensionen und Eigenschaften des Browsers und/oder des Geräts bekannt sind“⁵⁵⁷.

Wo die eben vorgestellten „absoluten“ Einheiten vermeintlich exakte Größen-Vorgaben definieren, orientieren sich die „relativen“ CSS-Einheiten stets an einer virtuellen Richtgröße, die im Prinzip von niemandem bestimmbar ist außer vom betrachtenden Benutzer selbst. Ausgangspunkt aller Größenangaben auf einer Webseite ist in diesem System der relativen Größenangaben nämlich die vom Nutzer in seinem Browser voreingestellte „Standard“-Schriftgröße. Ausgehend von diesem „höchsten“ Definitionselement werden dann die übrigen Größen auf der Seite festgelegt.

So gilt beispielsweise die vom Nutzer definierte „Standard“-Schriftgröße im CSS-System der relativen Längeneinheiten per se als „1 em“⁵⁵⁸ – gleichgültig, ob der Nutzer nun „18 Punkt“ als Standardschriftgröße definiert hat oder „12 Punkt“. Definiert nun ein Webdesigner eine Überschrift in der Größe „1.5 em“, so bekommt der erstgenannte Nutzer die Schrift in der Größe (1.5 * 18 Punkt) = 27 Punkt auf seinem Monitor zu sehen, der zweitgenannte dieselbe Überschrift in (1.5 * 12 Punkt) = 18 Punkt. Auf demselben Prinzip beruht im Übrigen nicht nur die hier beispielhaft angeführte Einheit „em“, sondern auch die Einheit „ex“, die definiert ist als die Höhe des Buchstabens „x“ in der vom Nutzer in seinem Browser eingestellten „Standard“-Schriftgröße. Dass es in CSS prinzipiell auch möglich ist, sich nicht nur auf die „Standardeinstellung“ zu beziehen, sondern „1.5 em“ beispielsweise auch vom vorhergehenden Element abhängig zu machen, sei hier nur der Vollständigkeit halber erwähnt; so ist

⁵⁵⁷ Crederholm 2006, S. 20

⁵⁵⁸ Vgl. Champeon/Fox 1999, S. 195f.

es beispielsweise möglich, zu definieren, dass ein Fließtext „0.7 em“ der Größe einnehmen soll, die die ihm voranstehende Überschrift aufweist – hier ist der Bezug zur Standardgröße des Browsers also nur noch über Umwege hergestellt. Die Tricks und Kniffe dieses Prinzips der „Vererbung“ sind für die in dieser Arbeit diskutierten Fragestellungen jedoch nicht sehr bedeutsam und sollen daher nicht weiter verfolgt werden⁵⁵⁹.

Ebenfalls zu den „relativen“ Größenangaben in CSS zählen Prozentangaben. Sie basieren auf demselben Prinzip wie die oben beschriebenen Einheiten „em“ und „ex“: Hat ein Benutzer als Standard-Größe etwa „14 Punkt“ definiert, erscheint eine in der Größe „150 %“ definierte Schrift auf dessen Monitor in $(14 \text{ Punkt} * 150 \%) = 21 \text{ Punkt}$.

Ebenfalls ein relatives Einheitensystem schlägt zudem das W3C vor. Dieses System der „Schlüsselwörter“ definiert die sieben „Größen“ „xx-small“, „x-small“, „small“, „medium“, „large“, „x-large“ sowie „xx-large“, wobei die jeweils nächsthöhere Einheit definiert ist als die 1,5-fache Größe des jeweils vorhergehenden Elements⁵⁶⁰. Das Problem: Als Ausgangsgröße dieses Hierarchiesystems gilt zwar wiederum verlässlich die vom Nutzer voreingestellte Standard-Schriftgröße. Jedoch interpretiert der *Internet Explorer* diese Standardschriftgröße als „small“, alle übrigen Browser dagegen als „medium“⁵⁶¹. Hinzu tritt, dass in diesem System lediglich sieben Größenabstufungen möglich sind, während die Einheiten „em“, „ex“ und „%“ deutlich feinere Differenzierungen in der Dimensionierung erlauben. Das siebenstufige System des W3C findet dementsprechend und augenscheinlich auch kaum Verbreitung auf zeitgenössischen Webseiten.

3.5.2.2 Der „Pixel“ – eine neue Maßeinheit?

Eine Maßeinheit ist im vorigen Kapitel noch nicht betrachtet worden: die Maßeinheit „Pixel“, abgekürzt „px“. Dies hat verschiedene Gründe.

Interessant ist zunächst, dass sich die Literatur keineswegs einig zeigt in einer durchaus wichtigen Grundsatzfrage: Ist der „Pixel“ den „absoluten“ CSS-Einheiten zuzurechnen (wie es beispielsweise Neutzling tut⁵⁶²) oder den „relativen“ (diese Meinung vertritt zum Beispiel Crederholm⁵⁶³)? Steht der „Pixel“ als Einheit also dem klassischen typografischen „Punkt“ näher oder der CSS-spezifischen, relativen Einheit „em“?

Diese Frage ist mehr als rein akademischer Natur. Nicht nur deswegen, weil sich der „Pixel“ im Webdesign augenscheinlich nach wie vor wohl der größten Beliebtheit erfreut. Ist der „Pixel“ also eine echte Alternativ-Einheit im Rahmen von CSS?

⁵⁵⁹ Vgl. ausführlicher Champeon/Fox 1999, S. 198ff.

⁵⁶⁰ Vgl. World Wide Web Consortium (W3C) 2006

⁵⁶¹ Vgl. Crederholm 2006, S. 19ff.

⁵⁶² Vgl. Neutzling 2002, S. 37

⁵⁶³ Vgl. Crederholm 2006, S. 19f.

Zunächst ist eine Feststellung von zentraler Bedeutung: Der Pixel ist die einzige Längeneinheit im CSS-Universum, die wirklich Monitor-spezifisch ist – in dem Sinne zumindest, als sie Quantitäten genau in jenem „Aggregatzustand“ definiert, in dem sie letztlich auf dem Bildschirm zu sehen sind: als Konstruktion von Bildquadraten, als Kollage von „picture elements“ also – eben „Pixeln“. In diesem Verständnis ist der „Pixel“ also eindeutig die „bildschirmgerechteste“, eindeutigste Einheit. In gewissem Sinne muss der „Pixel“ daher auch als „absolut“ angesehen werden: Ein Pixel auf einem Bildschirm ist und bleibt ein Pixel.

Und dennoch ist der Pixel, anders als beispielsweise der print-orientierte Punkt, eindeutig kein absolutes Maß⁵⁶⁴. Dies wird im Folgenden nachgewiesen; zum besseren Verständnis dessen sei allerdings noch kurz eine definitorische Setzung fixiert. Ich werde im Folgenden sowohl von der „systemischen Bildschirm-Auflösung“ sprechen als auch von der „physischen Bildschirm-Auflösung“. Unter der „systemischen Auflösung“ verstehe ich dabei die Anweisung des Nutzers an „sein“ Betriebssystem, in wie viele Bildquadrate die Monitorfläche zu zerlegen sei; die heute verbreitetste „systemische Auflösung“ dürfte die horizontale Aufteilung der Monitorfläche in 1024 Bildpunkte und die vertikale Aufteilung in 768 Bildpunkte darstellen – die „berühmte“ Auflösung „1024 x 768“. Diese systemische Auflösung kann vom Nutzer manipuliert werden.

Die „physische Auflösung“ geht noch einen Schritt weiter: Hier wird die „systemische Auflösung“ in Bezug gesetzt zur physischen Größe des Monitors. Erst auf diese Weise wird die Maßeinheit „Pixel“ wirklich greifbar und – wenn man so will – dann auch in gewissem Sinne „absolut“.

Diese Feststellung ist alles andere als unbedeutend. Denn sie ist Grundlage für den Nachweis, dass ein sich hartnäckig haltendes Gerücht schlichtweg falsch ist: Das Gerücht nämlich, Bildschirme wiesen prinzipiell eine physische Auflösung von 72 ppi auf⁵⁶⁵. „ppi“ steht dabei, wie bereits dargelegt, für „pixel per inch“, „(Bild-)Punkte pro Zoll“ also, und bezeichnet die Menge an Pixeln, welche auf einem bestimmten Monitor in einer bestimmten „systemischen Auflösung“ für einen einzelnen Bildpunkt, also ein „Pixel“ zur Verfügung stehen. Und der „72-dpi-Mythos“ behauptet, wie erwähnt, diese Pixel-Dichte sei auf allen Monitoren gleich.

Wäre dies tatsächlich so, wäre die Rechnung in der Tat klar und die Welt der Bildschirmgestaltung als Ganze einfacher zu handhaben: 1 amerikanischer Zoll (= inch) entspricht 25,40 Millimetern, mithin entspräche auf einem 72-dpi-Bildschirm ein Pixel ($25,40 / 72$) = 0,352778 Millimetern (und damit wiederum genau dem DTP-Punkt). Der Pixel wäre in dieser Rechnung also absolut identisch mit dem typografischen Punkt. Hier ist tatsächlich, wie Weinman doziert, „ein Punkt gleich einem Pixel“⁵⁶⁶. Aber eben nur in diesem, einen Beispiel.

⁵⁶⁴ Vgl. Spona 2006, S. 115

⁵⁶⁵ Vgl. z. B. Bayer 2003, S. 12; vgl. ferner Weinman 2000, S. 305

⁵⁶⁶ Weinman 2000, S. 305

Der ominöse Wert „72 ppi“ gilt eben nicht absolut; in der „Windows-Welt“ war er sogar nie richtig. Ein simpler Selbstversuch entlarvt die Unhaltbarkeit der These.

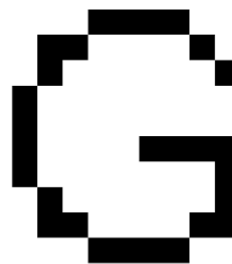
Die vorliegende Arbeit nämlich wurde überwiegend an einem *Apple* PowerBook G4 verfasst, dessen Bildschirmfläche (grob handvermessen) 36,7 Zentimeter = 14,45 inch breit ist. Bei einer systemischen Auflösung von 1440 Pixeln, die das Notebook in der Breite darstellt, ergibt sich eine Pixelbreite von $(14,45 / 1440) = 0,0100$ inch = 0,2540 Millimetern für diesen speziellen Rechner; das heißt, die 1440 Bildquadrate, für die der Versuchsrechner in der Horizontalen Platz bietet, sind jeweils grob einen Viertelmillimeter breit. Das aber entspricht rechnerisch einer physischen Auflösung von 99,7 ppi. Durch Ungenauigkeiten beim Vermessen der Bildschirmbreite kann die Diskrepanz von fast 40 % zwischen den beiden bislang errechneten Werten sicherlich nicht erklärt werden.

Nun sind Besitzer von *Macintosh*-Rechnern traditionell dem Misstrauen von Nutzern von PCs ausgesetzt, auf denen ein *Microsoft*-Betriebssystem installiert ist. Daher sei das Beispiel wiederholt und – vor allem – erweitert am Beispiel eines 17-Zoll-PC-Monitors, der sich im Haushalt des Verfassers fand, dessen Monitorfläche 32,1 Zentimeter = 12,64 inch in der Breite misst und der in einer systemischen Auflösung von 1024 x 768 Pixeln betrieben wird: Auf diesem Rechner in dieser systemischen Einstellung ist ein Pixel $(12,64 / 1024) = 0,0123$ inch = 0,31 Millimeter groß – und das wiederum entspricht einer physischen Auflösung von 81,0 dpi. Aber nur in dieser Auflösung! Stuft man nämlich – was in heutigen Betriebssystemen kein Problem darstellt – auf demselben Monitor die systemische Auflösung herab auf 800 x 600 Pixel, so wachsen die Pixel auf eine Größe von $(12,64 / 800) = 0,0158$ inch = 0,401 Millimeter – und die relative Auflösung schrumpft prompt auf erbärmliche 63,3 dpi⁵⁶⁷.

Was folgt aus dieser Kalkulation? Ein „G“, eingerichtet in einer Versalhöhe von 10 Pixeln (siehe Abb. 3.22), ist tatsächlich (fast) immer 10 Pixel hoch. Wie hoch die Type allerdings tatsächlich in – absoluten – Millimetern auf dem Monitor ausfällt, hängt, wie eben gezeigt, einerseits von der Größe des Bildschirms ab und andererseits von der vom Betrachter eingestellten systemischen Auflösung. Auf einer 24,2 Zentimeter hohen Monitorfläche, auf der die Auflösung 1024 x 768 implementiert ist, nähme dieses „G“ $(10 / 768 * 242) = 3,15$ Millimeter Höhe ein, auf einer 22,8 Zentimeter hohen Bildschirmfläche mit der Auflösung 1440 x 900 Pixeln dagegen nur $(10 / 900 * 228) = 2,55$ Millimeter.

Der Pixel ist also eindeutig ein relatives Maß. So relativ, dass es beispielsweise unmöglich ist, ein Zentimeter-Maßband einzuscannen und in eine HTML-Seite einzubetten mit dem An-

10 Pixel Versalhöhe



So sieht ein „G“ (*Arial*) aus, wenn man ihm 10 Pixel Platz in der Vertikalen einräumt.

ABBILDUNG 3.22

⁵⁶⁷ Vgl. ergänzend Kesper et al. 2004, S. 48

spruch, dass Betrachter der Seite am Bildschirm mithilfe des Maßband-Scans verlässliche Messungen durchführen könnten. Das Maßband auf dem Bildschirm entspräche seinem Original in Höhe und Breite nur dann, wenn die Auflösung des Bildes exakt gleich der physischen Auflösung des Bildschirms wäre. Diese Konstellation aber dürfte, wie gezeigt, nur selten und dann in aller Regel zufällig zustande kommen.

Woher rührt also der „72-dpi-Mythos“? Er dürfte aus der überkommenen Epoche stammen, in der Computer und Bildschirme der Firma *Apple Macintosh* zumindest den Markt des Desktop Publishing beherrschten. So wies beispielsweise ein 11 Zoll breiter Monitor bei einer Auflösung von 800 x 600 Pixeln tatsächlich eine Pixeldichte von $(800 / 11) = 72,72$ Pixel pro Zoll (gleich Inch) auf. Damit war die 100 %-Darstellung der am (auf 72 dpi basierenden) *Apple*-Rechner entworfenen Papierseite nahezu identisch in der Größe mit dem später zu druckenden Produkt. Man konnte also ein DIN A 4-Blatt neben den Monitor halten und einschätzen, wie das fertige, später zu druckende Produkt aussehen würde. In Zeiten des Nadeldruckers kein Nachteil.

Das Betriebssystem *Windows* der Firma *Microsoft* basiert jedoch auf der Annahme einer Bildschirmauflösung von 96 dpi. Da aber – wie beschrieben – der DTP-Punkt per Definition einem 72-stel des amerikanischen Zoll entspricht, ergibt sich in *Windows*-Umgebungen folgendes Problem: Bestimmt der Gestalter einer Webseite eine Schriftgröße mit beispielsweise 10 Punkt, berechnen *Windows*-Systeme die Pixel-Höhe wie folgt: 10 Punkt entsprechen, da 1 Punkt im DTP-Punkt-System ein Zweiundsiebzigstel Inch beträgt, zehn Zweiundsiebzigstel Inch; das entspricht einer darzustellenden Schrifthöhe von etwa 0,1389 Inch. Stehen nun, wie in den *Windows*-Betriebssystemen, 96 Pixel zur Verfügung, um einen Inch darzustellen, so werden der definierten 10-Punkt-Schrift $(96 \times 0,1389) = 13,333$ Pixel eingeräumt.

Mit dem Betriebssystem OS X hat im Übrigen auch die Firma *Apple Macintosh* in Sachen Schriftgrößen-Berechnung vor dem großen Rivalen *Microsoft* resigniert – und verrechnet inzwischen Schriftgrößen virtuell auch auf 96-dpi-Basis. Das Diktum, dass auf dem PC „der Text generell größer dargestellt (wird) als auf einem *Apple Macintosh*“⁵⁶⁸, besitzt also nur noch so lange Gültigkeit, wie *Apple*-Rechner mit dem Betriebssystem 9 und älter betrieben werden. Und diese Ära dürfte sich zügig ihrem Ende zuneigen.

Was folgt aus dieser Erkenntnis? Nun, zum einen, dass die „Währungseinheiten“ Pixel und Punkt nicht mehr die wechselseitige Entsprechung besitzen, die sie eine Zeitlang in der Monitorarstellung beanspruchen durften. Ob man für eine Webseite eine Schriftgröße von 12 Punkt definiert oder eine von 12 Pixeln, determiniert in den gängigen Betriebssystemen einen Größen-Unterschied von 33 Prozent. Oder, anders gesagt: eine Schriftgrößen-Angabe in Punkt muss immer ein Drittel über der Pixel-Angabe liegen, um eine größenmäßige Entsprechung am Bildschirm zu generieren.

⁵⁶⁸ Köhler 2002, S. 274.; vgl. ferner Kommer/Mersin 2002, S. 193

Warum aber unterstellen Betriebssysteme überhaupt eine Auflösung, in diesem Falle eine von 96 dpi? Warum werden Schriftgrößen an Bildschirmen nicht einfach in der einen, nahe liegenden Einheit eingefordert und dargestellt: Genügte in der Welt der Monitore nicht eine Selbstbeschränkung auf den zwar unwägbaren, aber immerhin planbaren „Pixel“? Ich habe in meinen Recherchen keine eindeutige Antwort auf diese Frage erhalten, verfechte aber inzwischen die Theorie, dass die Welt des Bildschirms zwar in Bezug auf Online-Medien eine weitgehend in sich geschlossene bildet – dass aber nach wie vor der Rechner immer noch oft genug nur Mittel zum „Materialisierungszweck“ ist: meistens ist dieser Zweck der „Ausdruck“ eines Dokuments. Eine bestmögliche Annäherung von Bildschirmdarstellung und Papier-Print ist aus diesem Grunde immer noch erwünscht, die Kreation eines „Verrechnungsfaktors“ zwischen Bildschirmpixel und Ausgabeprodukt unumgänglich. Für die isolierte Welt des Bildschirms würde die Maßeinheit „Pixel“ tatsächlich ausreichen; für die Kommunikation mit der Nicht-Monitor-Welt aber bedarf es eines „Umtauschkurses“. Und das sind zurzeit eben jene vermuteten 96 dpi (oder ppi).

3.5.3 Empfehlungen und Vorgaben aus der Literatur

3.5.3.1 Die Frage nach der Maßeinheit

Angesichts der Fülle von Alternativen, die das Konzept der „Cascading Style Sheets“ alleine in puncto „Größeneinheit“ vorhält, überrascht es wenig, dass die Empfehlungen und Vorgaben der Literatur, gelinde gesagt, uneinheitlich sind, was die vermeintlich bestmögliche Maßeinheit im Web angeht. Im wesentlichen sind freilich zwei recht eindeutige „Lager“ zu identifizieren, deren Standpunkte, implizit oder ausdrücklich, von der Grundeinschätzung darüber beeinflusst sind, wie erwünscht der „mündige Nutzer“ ist.

Das Lager der „Designer“ – so möchte ich die Angehörigen der einen „Denkschule“ ungebührlich verkürzend, aber griffig und kurz bezeichnen – wünscht bestmögliche Kontrolle in dem Sinne, als sie darauf zielen, ihre Seitenentwürfe möglichst unverstellt und möglichst wenig manipulierbar durch die Nutzenden auf die Bildschirme zu bringen. Eine Abhängigkeit von der „Standardgröße“ im Browser des Nutzers wird in diesem Lager eher als lästige Einschränkung denn als Chance begriffen. Das Lager der „Designer“ ist also das Lager derjenigen, die „absolute“ Vorgaben und damit Maßeinheiten favorisieren.

Das Lager der „Usability-Verfechter“ dagegen hält fest am Ideal des selbstbestimmten Nutzers. Wenn dieser eine größere Schrift wünscht – dann soll er sie herstellen!

Es sei allerdings nochmals darauf verwiesen, dass die Gegensätze dieser vermeintlich unversöhnlichen Lager eher ideologischer als praktischer Natur sind. Reduziert man die tatsächlichen Kontrollmechanismen auf ihren Kern, so bleibt nicht viel mehr übrig als dies: Die „Usability-Verfechter“ unterwerfen sich von vorneherein der „Standard-Schriftgröße“ des Nutzers

als Ausgangswert, die „Designer“ versuchen dagegen zunächst eine klare, eindeutige Definition, beispielsweise in Form einer vermeintlich absoluten Größenangabe „14 pt“ – diese aber ist wiederum genauso manipulierbar wie die relative Definition „1.5 em“. Mit der einen Ausnahme, die bereits angesprochen wurde: Im *Internet Explorer* ist eine auf Basis der Einheit „Pixel“ definierte Schriftgröße nicht relativierbar über das Browser-Menü⁵⁶⁹.

Wie also verteilen sich die Lager? Lamprecht empfiehlt eindeutig, „mit absoluten Größen (zu) arbeiten. Dabei sollten Sie Maße verwenden, die im Zusammenhang mit Schriften nützlich sind, wie zum Beispiel das Maß ‚Punkt‘“⁵⁷⁰. Neutzling empfiehlt dagegen eher den Pixel als Einheit: „Vergessen Sie die Größenangabe Punkt (pt), wenn Sie Schriftgrößen für die Bildschirmausgabe einstellen wollen. (...) Definieren Sie dagegen Ihre Schriftgrößen in Pixeln, sind Sie auf der sicheren Seite, weil die Einheit Pixel sozusagen absolut ist“⁵⁷¹. Weinman favorisiert ebenfalls die Einheit Pixel „zum Festlegen von Text (...), da Sie auf diese Weise die zuverlässigsten Ergebnisse erhalten“⁵⁷². Auch Balzert verfißt diese These⁵⁷³.

Zu den bekannten Verfechtern eines „nutzerorientierten“ Webdesigns zählt, wie bereits mehrfach dargelegt, der US-Amerikaner Jakob Nielsen. Sein Apodiktum überrascht daher wenig: „Verwenden Sie keine absoluten Fontgrößen. Spezifizieren Sie den Text stattdessen in einer relativen Größe zum Ausgangsfont. (...) Es ist schon ärgerlich, wenn (...) Sie auf eine Schaltfläche klicken, um einen Text größer zu machen, und nichts passiert, weil die Fontgrößen absolut definiert wurden“⁵⁷⁴. Diese Vorgabe wiederholt er später gemeinsam mit Tahir⁵⁷⁵. Auch Kesper et al. meinen, absolute Maßeinheiten seien für die Angabe von Schriftgrößen nicht zu empfehlen⁵⁷⁶. Das sieht auch Crederholm so⁵⁷⁷, ebenso wie Seibert und Hoffmann⁵⁷⁸ und Hellbusch; dieser bezieht zusätzlich zum Bekenntnis zur „relativen“ Größeneinheit darüberhinaus noch eindeutig Stellung, welche Einheit dies sein sollte: nämlich die Einheit „em“⁵⁷⁹.

Augenfällig ist in der Gesamtschau zweierlei: Erstens, dass die Verfechter „relativer“ Größenangaben sich merklich in der Mehrzahl befinden – und, zweitens die dünne Quellenlage insgesamt. Es erstaunt doch, dass die durchaus basale Frage „absolute oder relative Größenangabe“ vielen Autoren schlicht keine Zeile wert zu sein scheint.

⁵⁶⁹ Vgl. Maurice/Rex 2006, S. 62

⁵⁷⁰ Lamprecht 2002, S. 159

⁵⁷¹ Neutzling 2002, S. 37

⁵⁷² Weinman 2000, S. 312

⁵⁷³ Balzert 2004, S. 181

⁵⁷⁴ Nielsen 2000a-, S. 84

⁵⁷⁵ Vgl. Nielsen/Tahir 2002, S. 51

⁵⁷⁶ Vgl. Kesper et al. 2004, S. 86f.

⁵⁷⁷ Vgl. Crederholm 2006, S. 38

⁵⁷⁸ Vgl. Seibert/Hoffmann 2006, S. 217ff.

⁵⁷⁹ Vgl. Hellbusch 2005, S. 148

3.5.3.2 Jenseits der Maßeinheit: Wie groß soll's denn nun sein?

Wenn im Folgenden Empfehlungen aus der Literatur zusammengefasst werden, die den Anspruch erheben, Webdesignern handhabbare Daten zur Einrichtung der Schriftgröße am Monitor mit auf den Weg zu geben, so muss dies mit der Einschränkung geschehen, dass fast alle „verordneten“ Maße an folgenden Grundproblem kranken:

- Fast nie ist das Punkt-System genannt, auf dessen Basis die Autoren argumentieren. Dieser Aspekt ist freilich am ehesten verschmerzbar, weil davon auszugehen ist, dass die Zitierten sich mehrheitlich auf den DTP-Punkt beziehen. Verlässlich ist die Information aber dennoch nicht.
- Die Autoren benennen fast nie, auf welches Computer-Betriebssystem sie sich beziehen. Das aber ist, wie gezeigt, eine entscheidende Information: wer eine 12-Punkt-*Times* als erstrebenswerte Schrift empfiehlt und dies vor dem Jahr 2001 tut, in dem sich das *Macintosh*-Betriebssystem OS X durchzusetzen begann, meint höchstwahrscheinlich 12 Punkt für *Microsoft*-PCs und 16 Punkt für *Macintosh*-Rechner. Wahrscheinlich, wohlgermerkt. Aber bestimmt nicht sicher. Und meint ein Autor, wenn er „Schriftgrößen nicht unter 10 Punkt“ für Webseiten anmahnt, 10 Punkt im alten *Apple*-System (= 10 Pixel Kegelhöhe) oder 10 Punkt auf 96-dpi-Basis (=13,333 Pixel)? Das Problem wird im Übrigen seine Bedeutung behalten auch im Zusammenhang mit den vorzustellenden empirischen Erkenntnissen: Waren die Teilnehmer eines Experiments, in dem 12- und 14-Punkt-Schriften verglichen wurden, mit Schriften in 12 und 14 Pixeln Höhe konfrontiert oder mit solchen in $(12 \times 1,333) = 15$ Pixeln und $(14 \times 1,333) = 18,667$ Pixeln? Die Unterschiede sind beträchtlich!
- Der stark variierende Charakter verschiedener Schriftarten wird sehr häufig unterschlagen in den Empfehlungen zur Schriftgröße. Eine 13-Punkt-*Times* aber entspricht, nimmt man die x-Höhe zum Maßstab, wie gezeigt in etwa einer 11-Punkt-*Verdana*. Was fängt man also in der Praxis an mit Empfehlung der Qualität, am Bildschirm seien möglichst keine Schriftgröße unter 12 Punkt zu verwenden? Nun, man nimmt sie vorläufig hin.

Wo also liegen die empfohlenen Margen? Beginnen wir mit Balzert: „Aus Gründen der Lesbarkeit sollten Sie bei Websites mit größeren Schriften arbeiten als bei Druckmedien“⁵⁸⁰, erläutert die Autorin; genauere Angaben finden sich bei ihr nicht. In der Tendenz liegt Balzert allerdings auf einer Linie mit ihren Kolleginnen und Kollegen.

Denn es fällt auf, dass die meisten Autoren in ihrem Empfehlungen vor allem darauf Wert legen, dass Schriften im Web nicht „zu klein“ ausfallen dürften; es dominieren folglich Ratschläge bezüglich einzuhaltender Minima, während Aussagen zu einer wie auch immer bemessenen „optimalen“ Größe fast unauffindbar sind. In der Benennung der Untergrenze allerdings existiert unter den Autoren keine übermäßig ausgeprägte Varianz. Bürgel und Neumann befinden sich am untersten Limit, wenn sie vorgeben: „Schriften unter 10 pt sind

⁵⁸⁰ Balzert 2004, S. 181

so grob aufgelöst, dass sie kaum lesbar sind“⁵⁸¹. Kommer und Mersin nennen den selben Wert als Minimum, flankieren ihn allerdings mit der differenzierenden Konkretisierung, dass dieses Minimum für Serifenschriften am besten noch einen Punkt größer, also bei „11 Punkt“ angesetzt werden sollte⁵⁸². Auch Neutzling gibt als Richtlinie vor, es sei dafür Sorge zu tragen, „dass der Betrachter nur Schriften ab etwa 10 Pixeln Schriftgröße zu sehen bekommt“⁵⁸³ (Neutzling unterstellt im zitierten Werk eine Äquivalenz von „Punkt“ und „Pixel“ als Größenangabe). Erben erläutert gleichfalls, dass Größen zwischen etwa 6 pt und 9 pt am Bildschirm fast nie angemessen seien. „Für einen Bildschirmtext sollten die Schriftgrade daher mindestens 10 pt betragen, besser aber zwischen 11 pt und 14 pt liegen“⁵⁸⁴. Auch Böhringer et al. raten zu mindestens 10 pt⁵⁸⁵.

Etwas höher setzen Kiehn und Tietzmann die untere Schwelle an. Sie raten grundsätzlich von Schriftgrößen unter 12 Punkt ab – und dies ausdrücklich in Bezug auf Schriftarten, „die von ihrer Form her“ prinzipiell bildschirmgeeignet seien. Es überrascht angenehm, dass die Autorinnen über eine bloße Dimensionsbenennung hinaus ausdrücklich Beziehungen zwischen Schriftgestalt und -größe konstruieren: Schriften mit kleinen Mittellängen und feinen Strichstärken – die Kiehn und Tietzmann als weniger monitorgerecht klassifizieren – benötigten danach sogar „deutlich mehr als 12 Punkt“⁵⁸⁶. Auch Lazar empfiehlt, insbesondere mit Blick auf ältere Nutzer mit eventuell eingeschränkter Sehkraft, Schriften nicht kleiner als in 12 Punkt zu setzen⁵⁸⁷.

Noch ein wenig höher setzt Stocksmeier das Größen-Minimum am Bildschirm an. Er empfiehlt allgemein: „Schriftgrößen unterhalb etwa 0,5 cm sind zu vermeiden“⁵⁸⁸ – was in etwa 14 Punkt im typografischen Maßsystem entspricht. Damit liegt er bereits merklich über Nielsen und Tahir, die zwar, wie gezeigt, prinzipiell die Schriftgrößenbestimmung im Ermessen der Betrachter am sinnvollsten aufgehoben sehen, allgemein aber die Verwendung der Schriftgröße 12 pt als Basiseinstellung empfehlen⁵⁸⁹. Da es die beiden bei dieser Angabe belassen, ist davon auszugehen, dass es sich in ihrem Verständnis um ein anzustrebendes Optimum handeln.

Erstaunlich selten finden sich Angaben zur anzustrebenden Schriftgröße in Form einer Marge. Köhler bildet eine der wenigen Ausnahmen; er rät „wegen des deutlich größeren Leseabstandes“ (im Vergleich zum Lesen vom Papier) dazu, „die Schriftgröße ca. 30 % größer (zu) wählen als bei gedruckten Dokumenten“⁵⁹⁰ und empfiehlt Schriftgrößen „von 12 bis 14

⁵⁸¹ Bürgel/Neumann 2001, S. 109

⁵⁸² Vgl. Kommer/Mersin 2002, S. 193

⁵⁸³ Neutzling 2002, S. 108

⁵⁸⁴ Erben 1998, S. 86

⁵⁸⁵ Vgl. Böhringer et al. 2003, S. 129

⁵⁸⁶ Kiehn/Tietzmann 1998, S. 43

⁵⁸⁷ Vgl. Lazar 2005, S. 149

⁵⁸⁸ Stocksmeier 2002, S. 95

⁵⁸⁹ Vgl. Nielsen/Tahir 2002, S. 53

⁵⁹⁰ Köhler 2002, S. 271

Punkt⁵⁹¹. Auch Hellbusch macht den Abstand zwischen Auge und Bildschirm als Einflussfaktor aus und eröffnet eine komplexere Rechnung: „Bei einem Abstand von 50 cm beträgt (die Schriftgröße, M.L.) ca. 3 mm, bei einem Abstand von 70 cm ca. 4,5 mm“⁵⁹². 3 mm entsprechen im typografischen System 8,5 Punkt, 4,5 Millimeter knappen 13 Punkt – dies also darf als Hellbuschs Größenempfehlung gelten.

Ebenfalls einer Nachkalkulation bedürfen die Empfehlungen Bürgels und Neumanns. Sie berechnen, dass der Leseabstand am Bildschirm mit 45 bis 55 Zentimetern gut um die Hälfte über dem Leseabstand bei Printerzeugnissen liege. Da die Autoren eine Print-Schriftgröße von 4 bis 5 mm für angemessen halten, kommen sie für den Bildschirm auf einen erstrebenswerten Wert von 6 bis 8 Millimetern⁵⁹³ – ins typografische System übersetzt entspricht dies einem bemerkenswerten „Korridor“ zwischen 17 und 23 Punkt.

3.5.4 Empirische Erkenntnisse zur Schriftgröße am Bildschirm

3.5.4.1 Schriftgrößen und Schriftarten am Bildschirm

Redelius untersuchte 1998 in einem Experiment mit 50 Teilnehmenden (21 Frauen, 29 Männer) im Durchschnittsalter von 32,1 Jahren die Lesbarkeit von vier Schriftarten in jeweils drei Schriftgrößen am Bildschirm. Er verglich dabei die Schriftarten *Times*, *Candida*, *Frutiger* und *Olympia* und damit nur eine der oben als typische „Web-Schrift“ klassifizierte. Jede dieser Schriftarten legte Redelius ungefähr in 10, 11 und 12 Punkt Größe vor. Das Wort „ungefähr“ ist dabei bewusst gewählt; hoch anzurechnen ist Redelius nämlich, dass er sich bei der Bestimmung der zu vergleichenden Schriftgrößen nicht am nominalen Punkt-Wert orientierte; um den unterschiedlichen Schriftgestalten seiner Testfonts gerecht zu werden, setzte er die vier Schriftarten jeweils so groß, dass die Spanne zwischen Ober- und Unterlänge über alle Vergleichsschriften einer Größen-Gruppe identisch war⁵⁹⁴.

Letztlich fielen Redelius' Ergebnisse überraschend wenig variabel aus. In der jeweils größten Schriftgröße bewegten sich die erhobenen Lesezeiten im denkbar schmalen Korridor zwischen 149,57 und 158,98 Sekunden (Differenz: 6,3 Prozent). Ausgeprägter war die Spanne zwischen bestem und schlechtestem Ergebnis bei den beiden kleineren Schriftgrößen; dies freilich war der Tatsache geschuldet, dass die Lesezeiten der *Times* bei jeder der beiden Verkleinerungen geradezu dramatisch anwuchsen, während die übrigen drei Schriftarten relativ nahe beieinander verharrten in Bezug auf die Lesegeschwindigkeiten.

Insgesamt kommt Redelius zum Schluss, es habe sich kein Hinweis darauf ergeben, dass „Bildschirmschriften, die Serifen aufweisen, grundsätzlich besser oder schlechter am Monitor

⁵⁹¹ Köhler 2002, S. 271

⁵⁹² Hellbusch 2005, S. 92

⁵⁹³ Vgl. Bürgel/Neumann 2001, S. 109

⁵⁹⁴ Vgl. Redelius 1998

lesbar sind als Groteskschriften“⁵⁹⁵. Die *Times* habe sich aber in Größen unter 12,17 pt als signifikant schlechter lesbar als alle übrigen getesteten Schriftarten erwiesen. Dies gelte auch für die *Olympia*, jedoch nicht für die *Frutiger* und die *Candida*, die offenkundig robuster gegen Verkleinerungen seien⁵⁹⁶.

3.5.4.2 Tempomessung mit Anti-Aliasing

Bernard et al. verglichen 2000 in einem Experiment mit 35 Versuchsteilnehmern die Lesegeschwindigkeiten der Schriftarten *Arial* und *Times New Roman* am Bildschirm⁵⁹⁷. Jede der beiden Schriftarten wurde dabei doppelt manipuliert in der Darstellung: beide Schriften wurden jeweils in 10 Punkt und in 12 Punkt präsentiert, zusätzlich erfolgte die Schriftdarstellung einmal „geglättet“ und einmal im „1-Bit-Modus“.

Die Ergebnisse der Lesegeschwindigkeits-Messungen waren erstaunlich uneinheitlich. In zweier der vier Anordnungen wurde, betrachtet man zunächst nur den Faktor „Schriftart“, die *Arial* schneller rezipiert, in den beiden anderen die *Times New Roman*. In den beiden geglätteten Varianten lag einmal die *Times New Roman* vorne, einmal die *Arial*. In der Schriftgröße 10 Punkt lag einmal (geglättet) die *Times New Roman* vor der *Arial*, im anderen Falle war es umgekehrt.

Auffällig schlecht schnitt die geglättete *Arial* in 10 Punkt ab (399 Sekunden Lesezeit). Lässt man diesen Ausreißer-Wert einmal beiseite, bewegten sich allerdings die übrigen sieben Konstellationen in Lesezeit-Bereichen von 365 bis 382 Sekunden – ein denkbar schmaler Korridor. Entsprechend fanden Bernard et al. auch keine einzige statistisch signifikante Differenz der Konstellationen untereinander. Die geringsten Lesezeiten fanden Bernard et al. für die ungeglättete *Times* in 12 Punkt Größe (365 Sekunden), gefolgt von der ungeglätteten 10-Punkt-*Arial* (369 Sekunden) und der geglätteten 10-Punkt-*Times* (370 Sekunden).

Insgesamt jedoch fanden Bernard et al. im wechselseitigen Vergleich der getesteten Konstellationen keinerlei signifikante Unterschiede, so weit es die Lesegeschwindigkeiten betraf.

Befragt nach ihrer subjektiven Einschätzung, welche der acht Konstellationen gut oder weniger gut lesbar sei, setzten die Versuchsteilnehmer in einer Sieben-Punkte-Wertungsskala die 12-Punkt-*Arial* in ungeglätteter Aufbereitung auf Rang eins (5,7 Punkte im Schnitt), hauchdünn vor der ebenfalls ungeglätteten 12-Punkt-*Times New Roman* und der geglätteten 12-Punkt-*Arial* (jeweils 5,6). Alle 12-Punkt-Varianten schnitten im Urteil der Probanden um einen guten halben Wertungspunkt besser ab als die 10-Punkt-Varianten, unabhängig vom Glättungsmodus. Im direkten Vergleich der Schriftarten lagen *Times* und *Arial* praktisch über alle Größen-/Modus-Kombinationen hinweg gleichauf in der Bewertung. Signifikante Unter-

⁵⁹⁵ Redelius 1998, S. 120

⁵⁹⁶ Vgl. Redelius 1998, S. 122ff.

⁵⁹⁷ Vgl. Bernard/Mills 2000

schiede fanden die Forscher jedoch auch unter den Bedingungen dieses Messkriteriums „subjektiv erlebte Lesbarkeit“ nicht.

Ein weitgehend identisches Bild ergab die Befragung, welche der acht Konstellation die Teilnehmenden als „scharf“ einschätzten – eine Erfassung, die Bernard et al. vermutlich unter dem Eindruck des häufigen Vorwurfs installierten, geglättete Schriften wirkten tendenziell „matschig“. Wiederum lagen alle 12-Punkt-Konstellationen vor den 10-Punkt-Varianten; die *Arial* schnitt jeweils ein wenig besser ab als die *Times*. Zwischen geglätteten und ungeglätteten Schrifanordnungen erkannten die Teilnehmenden aber überraschenderweise so gut wie gar keine Unterschiede in der „Schärfe“.

In ihrem Fazit geben sich Bernard et al. unentschlossen; dies überrascht wenig angesichts der Tatsache, dass dieses Experiment wirklich statistisch bedeutsame Ergebnisse nicht erbrachte. Die Forschenden halten vorsichtig fest, dass 12-Punkt-Schriften wohl allgemein 10-Punkt-Konstellationen vorzuziehen seien, unabhängig von der Schriftart. Die *Times* als Serifenschrift sei zudem offenbar nicht schlechter lesbar als die *Arial*, wohl aber tendenziell etwas unbeliebter.

3.5.4.3 8 Schriftarten in 3 Größen: Lesbarkeit und Attraktivität

Bernard et al. führten 2002 eine weitere Studie durch, die für die Diskussion dieser Arbeit sehr wertvoll erscheint⁵⁹⁸. 60 Probanden (16 Männer, 44 Frauen) zwischen 18 und 55 Jahren bekamen dabei Texte zu lesen in acht Schriftarten, die zumindest potenziell (weil *Windows*-Standardfonts) durchaus als Web-Schriften in Frage kommen (*Century Schoolbook*, *Courier New*, *Georgia*, *Times New Roman*, *Arial*, *Comic Sans*, *Tahoma* und *Verdana*), und zwar jede dieser Schriftarten in 3 Größen: 10, 12 und 14 Punkt; jeder Teilnehmende wurde also mit insgesamt 24 Textpassagen konfrontiert. Ob die Schriften auf dem Testmonitor geglättet dargeboten wurden, geht aus den Beschreibungen der Autoren nicht hervor.

Insgesamt am besten unter allen 24 gemessenen Konstellationen schnitt die *Tahoma* in 12 Punkt Größe ab (270 Sekunden durchschnittliche Lesezeit), vor der *Times* in 12 Punkt (274 Sekunden), die wiederum vor der *Verdana* in 12 Punkt (285 Sekunden) lag. Auf Rang vier landete – als „Überraschkandidat“ – die Kinderschrift *Comic* in 12 Punkt (288 Sekunden), knapp dahinter rangierten die *Arial* in 14 Punkt (291) und die *Arial* in 12 Punkt Größe (292), gefolgt von der *Georgia* in 12 Punkt (300).

Bereits diese erste Übersicht lässt deutlich werden: Die Schriftgröße „12 Punkt“ erwies sich in fast allen Schriftarten als die am besten lesbare; in allen Schriftarten wurden die Texte in 10 Punkt Größe wie auch in 14 Punkt teilweise erheblich langsamer rezipiert, dabei in der 14-Punkt-Variante durchgehend schneller als in der 10-Punkt-Einrichtung (die einzige Ausnah-

⁵⁹⁸ Vgl. Bernard et al. 2002b

me bildete die *Schoolbook*, die in 14 Punkt am raschesten gelesen wurde). Im Schnitt wurden die 10-Punkt-Größen statistisch signifikant langsamer gelesen als die 12-Punkt-Größen.

Im Durchschnittsvergleich der Lesezeiten schnitten alle *Times*-Varianten (also die gemittelten 10-, 12- und 14-Punkt-Konstellationen) am besten ab (306 Sekunden Lesezeit), vor den *Arial*-Größen (311 Sekunden). Diese beiden Schriftarten erwiesen sich damit als am wenigsten empfindlich gegen Größenveränderungen; laut Bernard et al. waren beide zudem signifikant schneller lesbar als die Schriftarten *Courier*, *Schoolbook* und *Georgia* (jeweils in allen Größen). Die *Comic* lag diesmal auf Rang 3 (320 Sekunden), es folgte die *Verdana* (321).

Bernard et al. befragten die Teilnehmenden zudem danach, welche der aufgebotenen Konstellationen sie als „lesbarste“ empfanden. Auf einer Skala von 1 bis 6 (1 = gar nicht, 6 = sehr) bewerteten die Probanden dabei die *Tahoma* in 10 Punkt als am besten lesbar (5,3), vor der *Georgia* in 10 Punkt und der *Courier* in 12 Punkt (beide 5,1). Es folgten die *Arial* in 10 und die *Verdana* in 12 Punkt Größe (jeweils 5,0) sowie die *Courier* in 10 Punkt (4,9) und die *Times* in 10 , die *Comic* in 12 und die *Arial* in 14 Punkt (je 4,8). Über alle Schriftgrößen hinweg wurde die Schreibmaschinenschrift *Courier* als lesbarste bewertet, knapp gefolgt von der *Arial* (beide 4,9) und der *Georgia* (4,8). Es folgten *Times* und *Tahoma* (4,7) sowie – nur auf Platz 6 – die *Verdana* (4,6). Echte Signifikanzen fanden Bernard et al. hier allerdings nicht.

Damit erwiesen sich also einige Schriftkonstellationen als „lesbar empfunden“, die dieses Etikett den Geschwindigkeitsmessungen zufolge gar nicht verdienten. In ihrer Mehrheit wurden im Übrigen auch in dieser Befragung die 14-Punkt-Varianten nicht als besser, sondern als schlechter lesbar eingestuft als die jeweiligen 12-Punkt-Varianten.

Ebenfalls sehr interessant sind die Ergebnisse der zweiten Befragung, die Bernard et al. in ihre Untersuchung integrierten. Diesmal ging es um die „Attraktivität“ der dargebotenen Schriftkonstellationen, die die Teilnehmenden wiederum auf der bekannten sechsstufigen Skala zu bewerten hatten. Ganz vorne lagen gleich 3 Serifenschriften: die *Georgia* in 10 Punkt (4,5), die *Times* in 14 Punkt (4,2) sowie die *Georgia* in 14 Punkt. Es folgte die *Verdana* in 12 Punkt (3,9) sowie die *Georgia* in 12, die *Verdana* in 12 und die *Schoolbook* in 14 Punkt (3,8). Über die verschiedenen Größen hinweg lag die *Georgia* ebenfalls klar vorne (4,1), es folgten die *Times* (3,8) und die *Verdana* (3,6). Abgeschlagen: *Arial* (3,2) und *Courier* (3,0), die statistisch signifikant schlechter abschnitten als die *Georgia*. Insgesamt fällt freilich auf, dass die Noten für die „Attraktivität“ durchgehend für alle Schriftarten um einen guten Punkt schlechter ausfielen als die für die „Lesbarkeit“.

Wie sind nun diese Ergebnisse in der Gesamtschau zu bewerten? Letztlich überließen Bernard et al. dies abschließend ebenfalls ihren Probanden; sie fragten nach der allgemeinen „Font-Präferenz“ ihrer Teilnehmer und drückten diese in einem nicht erläuterten Ermittlungs-Verfahren in Prozentwerten aus (höhere Werte entsprechen dabei einer höheren Be-

vorzugung). Unter den 10-Punkt-Größen lag dabei die *Verdana* ganz eindeutig vorne (67 Prozent), gefolgt von *Georgia* (58), *Comic* und *Arial* (je 55). Unter den 12-Punkt-Größen eroberte die *Arial* die Spitze (55 Prozent), dahinter lagen *Georgia* (52), *Verdana* (50) sowie *Comic* (49). In der „Wertungskategorie 14 Punkt“ lag die *Comic* vorne (55), *Georgia* und *Arial* folgten gleichauf (52), dahinter die *Verdana* (48). Sehr schlecht schnitt die *Times* in allen drei Größen ab: in 10 und 12 Punkt bildete sie das Schlusslicht in der Präferenz (6 und 17 Prozent), in 14 Punkt lag sie mit 27 Prozent nur knapp vor der *Schoolbook*. Über alle vorgelegten Größen hinweg lag die *Verdana* (55) knapp vor der *Georgia* und der *Arial* (je 54). Die *Comic* war mit 52 Prozent die Nummer vier in der Präferenzliste. Gesamt-Schlusslicht: Die *Times* (18).

3.5.4.4 Vier Schriftarten im Test

Tullis et al. erforschten 1995 die „Lesbarkeit von Schriften in der ‚Windows‘-Umgebung“ in einem Experiment mit 15 Teilnehmern im Alter von 27 bis 45 Jahren⁵⁹⁹. Damit war diese Untersuchung zwar keine unmittelbar web-spezifische, bezog sich jedoch auf jene *Windows*-Version, mit deren Hilfe sich der *Internet Explorer* zum bestimmenden Browser weltweit aufschwingen sollte. Getestet wurden die Schriftarten *Small Fonts*, *Arial*, *MS Sans Serif* und *MS Serif* in unterschiedlichen Größen. Gemessen wurden zum einen die Lesezeiten; diese ergaben sich aus der Zeitdifferenz zwischen dem Erscheinen eines Testtextes auf dem Bildschirm und der Betätigung der „Enter“-Taste durch die Probanden, mittels dessen diese das Ende des Leseprozesses signalisierten. Um einen konzentrierten Leseprozess sicherzustellen, waren in allen Testtexten zwischen einem und fünf zufällige Fehler platziert, die die Teilnehmenden ebenfalls per Tasteneingabe zu identifizieren hatten.

Alle Schriftarten wurden am schnellsten gelesen in der jeweils größten dargebotenen Schriftgröße (*Small Fonts* in 6,75 Punkt, die übrigen Schriften in 9,75 Punkt), *MS Serif*, *MS Sans Serif* und *Arial* lagen in dieser Größe praktisch gleichauf in den Lesezeiten. Die *Arial* wurde mit schrittweiser Vergrößerung immer besser gelesen (zwischen der Größe 7,5 Punkt und der Größe 9,75 Punkt verringerte sich die Lesedauer um fast 20 Prozent!), in der Schriftgröße 8,25 Punkt lag die *MS Sans Serif* vor allen anderen Vergleichsschriften.

Auch in der Messung der Lese-Akkuratesse (erhoben anhand der jeweils identifizierten Fehler in den Texten) schnitten die Schriften in 9,75 Punkt am besten ab. Auch in diesem Vergleich wurden die Werte für die *Arial* mit wachsender Schriftgröße immer besser (in der Größe 7,5 Punkt entdeckten die Probanden in *Arial*-Texten nur 58 Prozent der Fehler, in der Schriftgröße 9,75 Punkt hingegen 84 Prozent); auch insgesamt lag die *Arial* vorne, und zwar deutlich (*MS Sans Serif*: 77 Prozent entdeckte Fehler, *MS Serif*: 73 Prozent).

⁵⁹⁹ Vgl. Tullis et al. 1995

In der abschließenden individuellen Bewertung der Lesbarkeiten der vorgestellten Konstellationen auf einer Skala von 1 bis 4 (1 = „poor“, 4 = „excellent“) votierten die Probanden eindeutig für *Arial* und *MS Sans Serif* in 9,75 Punkt Größe (beide Durchschnittswert 3,7), es folgte die *Arial* in 9 Punkt (3,3), die *MS Sans Serif* in 8,25 Punkt (3,0) sowie die *MS Serif* in 9,75 Punkt (2,7). Die *Small Fonts* schnitt in beiden dargebotenen Größen (6 und 6,75 Punkt) sehr schlecht ab (1,0 beziehungsweise 1,4).

3.5.4.5 Schriftpräferenzen von Kindern

In einer weiteren Studie erforschten Bernard et al. die Vorlieben von Kindern bezüglich der Schriftauswahl⁶⁰⁰. 10 Jungen und 17 Mädchen im Alter von 9 bis 11 Jahren wurden an einem 17-Zoll-Monitor acht Texte zum Lesen präsentiert: Je zwei Texte waren dabei in *Times New Roman* gesetzt, je zwei weitere in *Courier New*, *Arial* und *Comic Sans*. Jede Schriftart wurde in 12 und 14 Punkt Größe präsentiert.

Die Kinder wurden gebeten, verschiedene Aspekte jeder Schriftkonstellation auf einer Skala von 1 bis 7 zu bewerten (1 = gar nicht, 7 = sehr). Ergebnis: Die Kinder empfanden die *Comic Sans* nicht nur als am besten lesbar (5,5 Punkte im Schnitt), sondern auch als am attraktivsten (5,1) und als Schulbuch-Schrift wünschenswert (4,9). Am schlechtesten in allen drei Kategorien schnitt die *Times New Roman* ab (Lesbarkeit: 4,6; Attraktivität: 4,5; Schulbuch-Eignung: 3,8). Die Lesbarkeit der *Times New Roman* wurde dabei statistisch signifikant schlechter eingeschätzt als die der *Arial* und der *Comic Sans*, ihre Attraktivität statistisch signifikant schlechter als die der *Comic Sans*. *Times New Roman* und *Courier* wurden statistisch signifikant weniger in Schulbüchern gewünscht als die *Comic Sans*.

Durch die Bank schnitten im Experiment die in 14 Punkt gesetzten Schriften messbar, jedoch nicht statistisch signifikant besser ab als die 12-Punkt-Schriften.

3.5.4.6 Lesegeschwindigkeit, -effizienz und Schriftpräferenzen älterer User

In einer weiteren Studie fragten Bernard et al. nach den Lesepräferenzen älterer Computernutzer und testeten Lesegeschwindigkeiten und Leseeffizienz von vier Schriftarten in je zwei Größen. Sie gewannen 27 Versuchsteilnehmer (12 Männer und 15 Frauen) im Alter von 62 bis 83 Jahren (Durchschnittsalter 70 Jahre) für ihr Experiment⁶⁰¹. Den Teilnehmern wurden an einem 15-Zoll-Monitor jeweils acht Texte vorgelegt; verwendet wurden dabei, pro Text variierend, die Schriftarten *Arial*, *Verdana*, *Times New Roman* und *Georgia*, jede Schriftart dabei in 12 und 14 Punkt Größe.

⁶⁰⁰ Bernard et al. 2001b

⁶⁰¹ Bernard et al. 2001a

In der Messung der Lesegeschwindigkeiten wurden die 12-Punkt-Serifenschriften *Times New Roman* und *Georgia* signifikant langsamer gelesen als alle 14-Punkt-Schriften im Test. Nicht signifikant, aber als Beobachtung interessant war, dass die beiden Serifenschriften in 12 Punkt langsamer, in 14 Punkt jedoch im Schnitt schneller rezipiert wurden (252 Sekunden) als die Grotesken *Arial* und *Verdana* (262 Sekunden). Allgemein deckten sich diese Erkenntnisse mit den Urteilen der Probanden, die Bernard et al. ebenfalls erhoben: Die 14-Punkt-Schriften wurden statistisch signifikant als besser „lesbar“ bewertet als die 12-Punkt-Schriften. In puncto Schriftart dagegen wurde keiner der gezeigten Fonts signifikant als „besser lesbar“ beurteilt als einer der anderen.

Ferner stellten die Forscher fest, dass die allgemeine Beurteilung der Schriftkonstellationen („Welche dieser Schriften ziehen Sie allgemein vor?“) auch in diesem Experiment stark mit der Größe der dargebotenen Schrift korrelierte. So erwiesen sich die beiden Grotesken *Arial* und *Verdana* in 14 Punkt als statistisch signifikant stärker präferiert als alle Schriftvariationen in der Größe 12 Punkt sowie die beiden Serifenschriftarten in 14 Punkt. Letztere dagegen konnten sich nicht signifikant absetzen von den in kleinerem Schriftgrad gesetzten 12-Punkt-Grotesken.

Insgesamt fanden Bernard et al., dass die Schriftgröße unabhängig von der jeweils dargebotenen Schriftart einen Haupteffekt auf die Leseeffizienz ausübte – diese maßen Bernard et al. mittels in den Texten versteckter Störwörter, die die Teilnehmer durch verbale Nennung zu identifizieren hatten. Die in 14 Punkt dargestellten Texte wurden statistisch signifikant „effizienter“ gelesen als diejenigen in 12 Punkt.

Schließlich stellen Bernard et al. heraus, dass sich keine Unterschiede in einer der Messkategorien ergeben habe zwischen den „Online“-Schriftarten *Georgia* und *Verdana* und ihren Print-Pendants *Times* und *Arial*.

3.5.5 Zwischenfazit: Schriftarten und Schriftgrößen

An dieser Stelle der Diskussion sei ein kurzes Zwischenfazit gestattet. Ein Zwischenfazit insbesondere, das die erstaunlich differierenden Erkenntnisse von Fachliteratur einerseits und empirischer Erkenntnis andererseits betonen soll: Die vermeintliche Unterlegenheit serifentragender, in den Strichstärken ausdifferenzierter Schriftarten in der Bildschirm-Lesbarkeit, die, wie gezeigt, herrschende Unterstellung ist in einschlägigen Publikationen, wurde in der Mehrzahl der vorliegenden empirischen Untersuchungen alles andere als bestätigt. Vielmehr erwiesen sich insbesondere die serifentragenden Schriftarten *Times* und *Georgia* in den meisten Studien als mindestens äquivalent schnell lesbar wie die „groteske“ Konkurrenz, oft sogar als überlegen in diesem Kriterium. Auch auf das Textverständnis scheint keinen Einfluss auszuüben, ob die gewählte Schrift Serifen aufweist oder nicht.

Wie kann das sein? Nun, ein weiteres Faktum hilft meines Erachtens weiter in der Interpretation: Insbesondere die *Times* ist zwar offenbar durchaus eine gut „lesbare“ Schriftart – aber, und diese Tatsache darf nicht unberücksichtigt bleiben, sie ist in ihrer ästhetischen Gestalt am Monitor offenkundig auch unbeliebt. Sie wird als unattraktiv empfunden, zumindest in Schriftgrößen bis etwa 12 Punkt. Dieser Meinung kann man sich meines Erachtens durchaus anschließen; ich hege jedoch den Verdacht, dass viele der oben zitierten, vermeintlichen „Fachautoren“ den eingeschränkten visuellen „Appeal“ serifentragender Schriften am Bildschirm, die „Deformierung“ der Typen, unmittelbar in Korrelation setzen mit deren Lesbarkeit – und zumindest dieser Kausalschluss scheint nach empirischer Quellenlage unhaltbar. Kurzum: Die *Times* ist wenig gelitten in der Nutzerschaft – gut lesbar ist sie aber offenbar trotzdem am Bildschirm.

Was das Thema „Schriftgröße“ angeht, so erachte ich die vorliegenden Studien insgesamt als nicht sonderlich aussagefähig. Eine in 14 Punkt gesetzte *Times* entspricht in ihrer x-Höhe in etwa einer *Verdana* in 12 Punkt – dass also letztere Konstellation in etwa gleich gut lesbar ist wie die erstere, kann da kaum überraschen. Insgesamt scheint es, als würden Schriften in 12 Punkt Größe von den Nutzern bevorzugt, und als verringere eine weitere Vergrößerung auf 14 Punkt eher den Zuspruch sowie Lesegeschwindigkeit und Leseerfolg der Rezipientenschaft. Es bleibt jedoch festzuhalten, dass die Schriftarten *Tahoma* und *Georgia* in einem Test in 10 Punkt Größe zumindest die besten Zustimmungsraten bezüglich ihrer allgemeinen „Attraktivität“ unter den Probanden überhaupt erzielten, zum Teil deutlich vor Schriften, die zwei oder gar vier Punkt größer dargeboten wurden.

Was folgt aus diesen Feststellungen? Nun, zunächst eine gesunde Skepsis im weiteren Verlaufe dieser Arbeit gegenüber kommenden, eventuell sogar „herrschenden“ Meinungen der Fachliteratur, die sich – diese zweifellos übertreibende Unterstellung sei unterbreitet – offenkundig nur sehr ungern in der Verbreitung von Dogmen stören lässt durch substanziell zuwiderlaufende empirische Erkenntnisse. Und damit sei übergeleitet zum nächsten typografischen Faktor als Betrachtungsgegenstand: dem Zeilenabstand.

3.6 Zeilenabstand

3.6.1 Der Zeilenabstand als typografisches Phänomen

Der Zeilenabstand bemisst sich als vertikaler Abstand zwischen den Grundlinien zweier übereinander positionierten Textzeilen. Es sei an dieser Stelle ausdrücklich darauf hingewiesen, dass „Zeilenabstand“ und „Durchschuss“ keine Synonyme bilden; vielmehr bezeichnet der Durchschuss die Differenz zwischen Schriftgröße und Zeilenabstand⁶⁰². Ist also Zeilenabstand gleich Schriftgröße (also: Kegelhöhe), ist der Durchschuss null.

⁶⁰² Vgl. Gulbins/Kahrmann 1992, S. 40

3.6.2 Manipulation des Zeilenabstands in HTML-basierten Browsern

Der Zeilenabstand geriet als letzter wichtiger typografischer Faktor in die „Einflusshäre“ des Webdesigns. Die erste CSS-Version von 1996, die freilich damals noch kaum ein Browser zu interpretieren in der Lage war, sah den Befehl „line-height“ vor, der bis heute existiert. Ausgestaltbar ist er mit numerischen Angaben in einer der vielen Web-Maßeinheiten, die im Kapitel „Schriftgröße“ bereits zur Sprache kamen: von „cm“ bis „em“ ist jede Spezifikation möglich.

3.6.3 Empfehlungen und Vorgaben aus der Literatur

Auch diese Zusammenfassung herrschender Meinungen in der Literatur, diesmal zum Thema „Zeilenabstand“, möchte ich beginnen mit einer Aussage von Praxisrelevanz, die politischen Sonntagsreden in wenig nachsteht. Dies tue ich, weil das folgende Zitat erneut verdeutlicht, in welchen für Praktiker mit konkretem Wissensdurst unverwertbaren Sphären sich ein Großteil der Fachliteratur zum Komplex „Webdesign“ bewegt. Neutzling nämlich warnt in Bezug auf die Einrichtung von Zeilenabstand: „Ist er zu gering, kann das Auge den nächsten Zeilenanfang nicht finden und beginnt zu suchen. Ist er dagegen zu groß, fallen die Zeilen auseinander“⁶⁰³. Eines Ratgebers unwürdig ist diese Aussage, nicht mehr und nicht weniger.

Konkreter wird da schon Erben, der meint, der optimale Zeilenabstand liege „bei etwa 150 %, eher noch darüber“⁶⁰⁴. Die Prozentangabe bezieht sich dabei auf die Größe der Schrift, gemessen in Punkt; für eine 12-Punkt-Schrift würden damit 18 Punkt Zeilenabstand fällig, für eine 10-Punkt-Schrift 15 pt Zeilenabstand. Auf jeden Fall sollte der Zeilenabstand aber laut Erben stets „großzügiger angelegt werden als auf dem Papier“⁶⁰⁵. Auch Schweizer meint, ein Zeilenabstand von „etwa 150% der Schriftgröße (erhöht) die Lesbarkeit deutlich“⁶⁰⁶. Köhler sieht gar noch größere Margen: „Akzeptable Zeilenabstände sind hier 150 % bis sogar 200 % der jeweiligen Schriftgröße“⁶⁰⁷, und auch für Bollwage sind diese 150 % das absolute Mindestmaß⁶⁰⁸. Thissen rät ebenfalls zu „eineinhalb- bis zweizeiligem Abstand“⁶⁰⁹. Kommer und Mersin setzen dagegen deutlich geringere Margen an: „Ein geeigneter Zeilenabstand für den Bildschirm sind 120 bis 140 Prozent der Schriftgröße (...). Ab einem Zeilenabstand von 150% beginnt der Text zu zerfallen“⁶¹⁰. In ähnlichen Größenordnungen setzt Gorbach die Grenzen, wenn er empfiehlt „beispielsweise bei 14 Punkt Buchstabengröße 4 bis 6 Punkt, also 14/18 oder 14/20 Punkt“⁶¹¹. Auch Balzert bewegt sich mit ihrem Beispiel eher im

⁶⁰³ Neutzling 2002, S. 111

⁶⁰⁴ Erben 1998, S. 90

⁶⁰⁵ Erben 1998, S. 90

⁶⁰⁶ Schweizer 2003, S. 153

⁶⁰⁷ Köhler 2002, S. 271

⁶⁰⁸ Vgl. Bollwage 2001, S. 102

⁶⁰⁹ Vgl. Thissen 2003, S. 98

⁶¹⁰ Kommer/Mersin 2002, S. 175

⁶¹¹ Gorbach 2001, S. 255

Bereich von 130 % Schriftgröße als Richtwert für den Zeilenabstand: „Für die Bildschirmdarstellung ist ein größerer Zeilenabstand als beim Druck empfehlenswert, z.B. 12/16“⁶¹².

Zu Interaktionen des typografischen Faktors „Zeilenabstand“ mit anderen Faktoren finden sich kaum Aussagen in der Literatur. Lediglich Kommer und Mersin meinen, man solle „weiche, runde Schriften wie die *Georgia*“ mit etwas mehr Zeilenabstand versehen⁶¹³.

3.6.4 Empirische Erkenntnisse zum Zeilenabstand am Bildschirm

Chaparro et al. untersuchten an 20 Studierenden (10 Männer, 10 Frauen) 2004 die Wirkung von „Weißraum“ auf Webseiten auf Lesegeschwindigkeit und Nutzerzufriedenheit⁶¹⁴. Der „Weißraum“ war dabei in zweifacher Hinsicht interessant für die Forscher: Zum einen maß sie die Wirkung von Weißraum, der einen Textblock als Ganzes umschließt (was im Zusammenhang dieser Arbeit weniger von Interesse ist), zum anderen jedoch den Einfluss von Weißraum zwischen Textzeilen – dem Zeilenabstand also.

Dabei installierten Chaparro et al. zwei Zeilenabstands-Varianten: In der Schriftart *Times* gesetzter Text wurde einmal aufgeboten in von den Forschern als „suboptimal“ definierter Form (4 Millimeter Zeilenabstand), in der Vergleichsvariante in „optimaler“ Form (5 Millimeter Zeilenabstand). Leider verlieren die Autoren kein Wort über die Schriftgröße, in der sie ihre Texte präsentieren; dies erschwert die Einordnung der verglichenen Zeilenabstände doch erheblich. Die Beispiel-Abbildungen legen jedoch nahe, dass es sich bei der „suboptimalen“ Variante mehr oder minder um „Kompress“-Zeilen handelte (der Zeilenabstand also numerisch annähernd der Schriftgröße entsprach); die „optimale“ Variante lag ungefähr 3 Punkt darüber.

Erstaunlicherweise erwiesen sich im Experiment die „suboptimalen“ Zeilenabstände als besser lesbar als die vermeintlich „optimalen“ – und zwar um etwa 3 Prozent, wenn das Textgesamt von Weißraum umschlossen war, und sogar um über 8 Prozent, wenn der umschließende Raum fehlte. Auch unter dem Kriterium des Leseverständnisses – operationalisiert mithilfe von Verständnisfragen an die Teilnehmenden nach dem Leseprozess – ergab sich kein messbarer Vorteil für die Texte mit „optimalen“ Zeilenabständen (im Modus ohne umschließenden Weißraum lagen hier die in kompresssem Zeilenabstand gesetzten Texte sogar vorne).

Ein nicht unwichtiger Effekt der vergrößerten Zeilenabstände, der im Experiment festgestellt wurde, darf allerdings nicht unterschlagen werden. Die Probanden äußerten sich nämlich insgesamt deutlich „zufriedener“ und bezeichneten sich als „weniger ermüdet“ nach der Re-

⁶¹² Balzert 2004, S. 175

⁶¹³ Vgl. Kommer/Mersin 2002, S. 176

⁶¹⁴ Vgl. Chaparro et al. 2004

zeption der Texte mit vergrößertem Zeilenabstand. Abschließend befragt nach ihrer individuellen Präferenz, äußerten 47 Prozent, sie bevorzugten die Variante „umschließender Weißraum und vergrößerter Zeilenabstand“, 32 Prozent entschieden sich für die Version „kein umschließender Raum und vergrößerter Zeilenabstand“. Nur knapp über 20 Prozent empfanden dagegen die kompakt gesetzten Texte als ansprechendste Darbietungsform.

3.6.5 Zwischenfazit: Zeilenabstand

Alle erwähnten Autoren empfehlen für die Bildschirmdarstellung einen „erhöhten“ Zeilenabstand. Die lediglich eine vorgefundene diesbezügliche empirische Studie bestätigt die herrschende Ansicht teilweise; nämlich in der Hinsicht, als Rezipienten es offenbar als angenehmer empfinden, wenn der Zeilenabstand großzügiger bemessen ist. In puncto Lesegeschwindigkeit und Leseverständnis dagegen spricht nach der dünnen Erkenntnislage nichts für erhöhte „Durchschüsse“ am Bildschirm.

3.7 Zeilenbreite

3.7.1 Die Zeilenbreite als typografisches Phänomen

Rein physikalisch lässt sich die Breite von Textzeile in gängigen Einheiten wie Millimetern oder typografischem Punkt bezeichnen. Dass Angaben auf dieser Basis von geringem Aussagewert sind, wenn die Schriftgröße unberücksichtigt bleibt, liegt auf der Hand. Folglich hat sich als „Maßeinheit“ für die Zeilenbreite sinnvollerweise die Anzahl der Zeichen beziehungsweise der „Anschläge“, gelegentlich auch die Anzahl der Wörter pro Zeile herausgebildet.

Bedeutsam ist die Zeilenbreite auch für die allgemeine Lesbarkeit eines Textes insofern, als zusätzlich zu den bereits beschriebenen „Sakkaden-Sprüngen“, in denen das geübte Auge horizontal und in Etappen über die Zeichen eines Textes gleitet und dabei diesen verarbeitet, jeweils ein Sprung am Ende einer jeden Zeile in die jeweils folgende Zeile hinzutritt. Dieser Zeilensprung unterbricht den Rezeptionsprozess und dauert um so länger, je weiter das Auge vom jeweils oberen Zeilenende zum darunter liegenden Zeilenanfang wandern muss⁶¹⁵.

3.7.2 Manipulation der Zeilenbreite in HTML-basierten Webbrowsern

In den Zeiten der ersten HTML-Versionen Anfang der 1990er Jahre erübrigte sich für Webdesigner (die es seinerzeit ohnehin in sehr spärlicher Anzahl gab) nahezu jedwede Erwägung, wie breit sie die Zeilen der Texte auf ihren Websites ausgestalten sollten. Texte im Internet

⁶¹⁵ Vgl. Gulbins/Kahrmann 1992, S. 44

waren (und sind im Prinzip bis heute) elastisch in ihrer Breite; in den Browsern wird im Grundsatz schlicht der Raum zwischen linker und rechter Fensterseite mit Text gefüllt, ein Umbruch erfolgt standardmäßig, wenn der horizontale Raum gefüllt ist⁶¹⁶ – was auf einem 20-Zoll-Monitor naturgemäß in aller Regel später und an anderer Textstelle der Fall ist als auf einem 17-Zoll-Bildschirm. Lediglich das HTML-Tag „
“ (für „Break“, erzwungener Umbruch) erlaubt ein aktives Eingreifen in die Zeilenbreite. Erst die 1994 neu eingeführten HTML-Tabellen und später die CSS-Formatierungsmöglichkeiten setzten die Gestalter von Webseiten in den Stand, die Zeilenbreiten ihrer Texte ansatzweise effizient zu manipulieren.

Die Freude über diese neu gewonnenen Gestaltungspotenziale ist allerdings nicht ungeteilt. Bis heute finden sich viele, durchaus anerkannte Fürsprecher des ursprünglichen „liquid designs“ oder „variablen Layouts“⁶¹⁷, das das Aussehen einer Webseite von der Bildschirmgröße beziehungsweise der Breite des Browser-Fensters abhängig machte: Webseiten (und damit die auf ihnen dargebotenen Texte) sollten nicht mit absoluten Breiten ausgestattet werden, so argumentieren diese Kritiker, da dies den unterschiedlichen Systemvoraussetzungen und Bildschirmbreiten der Betrachter weltweit nicht gerecht werde: Besitzer großer Monitore würden die Flexibilität zu schätzen wissen, die ihnen durch die flexible Breite einer Webseite entstünde; gleichzeitig würden Nutzer mit kleineren Bildschirmen nicht ausgeschlossen von der ungestörten Betrachtung der Site⁶¹⁸.

Es ist allerdings zu fragen, ob diese Haltung angesichts immer größerer Bildschirme (und damit potenziell auch immer breiter aufgezogener Browser-Fenster) noch haltbar ist. Ein in seiner Breite völlig flexibler, mithin potenziell über die gesamte Bildschirmbreite laufender Text beispielsweise in der Schriftart *Verdana*, die in 12 Punkt gesetzt ist, erzeugt in einer zugestandenen Breite von 1000 Pixel Zeilen, die durchschnittlich mehr als 160 Zeichen Platz bieten – bei der vermutlich weit verbreiteten systemischen Bildschirmauflösung von 1024 x 768 Pixeln ist dieses Beispiel keineswegs realitätsfremd konstruiert.

Es wird auf den folgenden Seiten – auch anhand der vorliegenden empirischen Studien zu diesem Thema – zu klären sein, ob Zeilen mit Zeichenpotenzial von über 150 Anschlägen wirklich lesbar sind. Fakt ist aber, dass diverse Autoren von einer völligen Freigabe der Zeilenbreite abraten⁶¹⁹: Hierdurch entstünden potenziell „sehr lange“⁶²⁰, wenn nicht gar „Bandwurmzeilen“⁶²¹, die „größere laterale Augenbewegungen oder sogar ein leichtes Drehen des Kopfes (erfordern) (...). Dadurch erhöht sich die Gefahr, die Zeile aus den Augen zu verlieren. Auch wenn es sich nur um ein paar Zentimeter handelt, wirkt diese Bewegung auf Dauer ermüdend“⁶²².

⁶¹⁶ Vgl. McKelvey 2000, S. 26

⁶¹⁷ Vgl. Balzert 2004, S. 106

⁶¹⁸ Vgl. Nielsen/Tahir 2002, S. 40

⁶¹⁹ Vgl. z.B. Crederholm 2006, S. 227

⁶²⁰ Balzert 2004, S. 106

⁶²¹ Lankau 2000, S. 41

⁶²² Kommer/Mersin 2002, S. 173

Zwar ist es richtig, dass Nutzer die Breite ihres Browser-Fensters durchaus selbst einrichten können; wie viele dies aber tatsächlich tun, und ob es wirklich „geschätzt“ würde von Web-Nutzern, für jede neu aufgerufene Seite die Browserfenster-Breite neu zu justieren, erscheint fraglich. Zunächst hilft ein Übersichts-Blick auf die Vorgaben der Literatur sicherlich weiter.

3.7.3 Empfehlungen und Vorgaben aus der Literatur

3.7.3.1 Allgemeine Vorgaben

Kaum ein typografisches Einzelphänomen wird in der Literatur zur Textgestaltung am Bildschirm widersprüchlicher behandelt als die Zeilenbreite im Allgemeinen – und die Margen, innerhalb derer sie sich bewegen sollte, im Besonderen.

„Überlänge der Zeilen ermüdet rasch“⁶²³, doziert zunächst Pring sehr allgemein und unter konsequenter Vermeidung klarer Grenzdefinitionen, und schiebt später ähnlich definiert nach: „Kürzere Zeilen erhöhen die Lesbarkeit“⁶²⁴. Eventuell hebt Pring dabei auf die verbreitete Befürchtung ab, dass überlange Zeilen Lesende vor Schwierigkeiten beim dann erforderlichen überweiten Rückschwung in die darunter liegende Zeile stellen könnten, sie also schlicht die Folgezeile buchstäblich „aus den Augen verlieren“ könnten⁶²⁵. Ausdrücklich sagt Pring dies freilich nicht.

Zunächst wohltuend konkreter wird da schon Gorbach. Er empfiehlt 35 Anschläge pro Zeile als Richtwert einer guten Zeilenbreite – wobei sein Hinweis, dass man am Bildschirm „nicht den Printmaximen folgen sollte“⁶²⁶, weder Aufschluss gestattet über die von ihm zugrunde gelegten Maximen noch über die vermeintlichen Gründe für die Abweichung von selbigen Maximen für die Monitor-Gestaltung. Immerhin sind zumindest gewisse Kongruenzen zwischen Gorbachs Empfehlung und Erbens Setzung bezüglich der Zeilenlänge erkennbar; letzterer freilich geht noch einen deutlichen Schritt weiter, denn sein Maximum entspricht Gorbachs Optimum. So gelten laut Erben „auf dem Papier 35 bis 55 Zeichen pro Zeile als optimal. Auf dem Bildschirm sind es (dagegen) maximal 35 Zeichen, eher weniger“⁶²⁷. Insgesamt korrespondiert diese Empfehlung mit der Puschers: „35 bis 40 Zeichen sind das Maximum“⁶²⁸, erklärt dieser, wobei er es versäumt, diesen Fingerzeig mit einer empirischen Erkenntnis zu unterfüttern. Darin freilich steht er ausnahmslos allen Vor- und Nachzitierten in diesem Kapitel in nichts nach. Auch Bollwage meint ohne weitere Referenz, im Web seien „40 Zeichen in einer Zeile (...) genug“⁶²⁹. Das finden auch Böhringer et al.⁶³⁰

⁶²³ Pring 2000, S. 18

⁶²⁴ Pring 2000, S. 24

⁶²⁵ Vgl. Gulbins/Kahrman 1992, S. 44

⁶²⁶ Gorbach 2001, S. 255

⁶²⁷ Erben 1998, S. 92

⁶²⁸ Puscher 2001, S. 42

⁶²⁹ Bollwage 2001, S. 102

⁶³⁰ Vgl. Böhringer et al. 2003, S. 133

Ist also bei 35 bis 40 Anschlägen Schluss mit der Lesbarkeit? Zweifel sind gestattet. Immerhin setzt beispielsweise Bürgel seine favorisierte Zeilenbreite im Minimum genau dort an, wo Gorbach das Optimum und Bergen sogar das Maximum erkennt: er erblickt die bestmögliche Zeilenlänge bei „ca. 35 bis 70 Zeichen pro Zeile“⁶³¹. Auch Kommer und Mersin gehen nur noch am Rande konform mit den Apologeten der schmalen, in den „Dreißigern“ befindlichen Spalten, wenn sie als anzustrebende Breite 30 bis 60 Anschläge pro Zeile empfehlen⁶³². In einer insgesamt ähnlichen, jedoch etwas knapper bemessenen Marge von 30 bis 50 Buchstaben verortet Hellbusch das gesunde Maß, mit der nicht näher spezifizierten Einschränkung freilich, die Breite sei auch „abhängig von Schrifttyp und Inhalt“⁶³³. Und wo Bürgel mit seiner Untergrenze von 35 Anschlägen gerade noch in Einklang zu bringen ist mit Puschers oberem Limit in genau derselben Dimension, sprengen Kiehn und Tietzmann endgültig die Harmonie der damit wiederum als nur vermeintliche identifizierten Fachliteratur, indem sie die Latte nochmals ein paar Einheiten höher legen: „Für das Lesen am Bildschirm sollte die Zeilenlänge 50 Zeichen nicht überschreiten“⁶³⁴, lesen wir diesmal. Doch damit ist der Bieterwettbewerb noch längst nicht beendet.

Eine wiederum recht weit gespannte „Maximalbreite von etwa 40 bis 60 Zeichen/Zeile“⁶³⁵ fordert Lankau für Text auf Webseiten, und überschreitet damit im hier diskutierten Zusammenhang als zweiter nach Kommer und Mersin das fünfte Dutzend an vermeintlich erlaubten Zeichen je Zeile. Doch er steht nicht allein damit: „Der optimale Lesbarkeitswert liegt zwischen 45 und 60 Zeichen je Zeile. Mehr oder weniger stören den Lesefluss“⁶³⁶, behauptet nämlich Schweizer, und steht damit wiederum in weitgehender Entsprechung mit Neutzling. Dieser verfügt Folgendes: „Während man bei der Gestaltung von Drucksachen von etwa 55 bis 75 Zeichen ausgeht, gelten am Bildschirm 45 bis 55 Zeichen pro Zeile als optimale Werte“⁶³⁷. Lynch und Horton favorisieren Neutzlings goldene Mitte, also 50 Anschläge pro Zeile⁶³⁸. Ihre Meisterin finden alle drei Autoren wiederum in Balzert, die in ihrer Empfehlung völlig neue Dimensionen nach oben eröffnet: „Bei einem Text mit guter Lesbarkeit sollten die Zeilen mindestens 50 und maximal 75 Zeichen enthalten“⁶³⁹.

An dieser Stelle sei die Bemerkung gestattet, dass das virtuelle Zeilenbreiten-„Wettrüsten“ allmählich in seine Endphase eintritt. Doch Wirths Aufschlag darf nicht unerwähnt bleiben; Für ihn sind nämlich 60 bis 70 Zeichen pro Text „nach den Standardregeln der Ergonomie“ der anzustrebende Wert, alles darüber sei „digitale Bleiwüste“⁶⁴⁰. Mit dieser Empfehlung,

⁶³¹ Bürgel/Neumann 2001, S. 134

⁶³² Vgl. Kommer/Mersin 2002, S. 173

⁶³³ Hellbusch 2005, S. 93

⁶³⁴ Kiehn/Tietzmann 1998, S. 46

⁶³⁵ Lankau 2000, S. 370

⁶³⁶ Schweizer 2003, S. 154

⁶³⁷ Neutzling 2002, S. 110

⁶³⁸ Vgl. Lynch/Horton 1999, S. 85

⁶³⁹ Balzert 2004, S. 186

⁶⁴⁰ Vgl. Wirth 2005

dies sei als Anmerkung gestattet, ist nunmehr feierlich die doppelte Menge empfohlener Zeichen pro Zeile erreicht, die Puscher als erträgliche Obergrenze von Zeichen pro Zeile benennt (s.o.). Wirth selbst legt im Übrigen Wert auf die Feststellung, dass seine eigene Website kommdesign.de mit einer durchschnittlichen Zeilenbreite von 80 Zeichen gestaltet ist, was insofern eine Bemerkung wert ist, als er damit den Bereich seiner selbst definierten „Bleiwüste“ betritt. Immerhin liegt Wirth damit wiederum weitgehend auf einer Linie mit Thissen: Dieser empfiehlt 8 bis 10 Wörter und 60 bis 80 Buchstaben pro Zeile als lesbares Optimum⁶⁴¹.

Schließlich sollen nicht unerwähnt bleiben Dengler und Volland. Diese befassen sich wiederum ausschließlich mit der Obergrenze, die Zeilen im Web breitenmäßig nicht überschreiten sollten, und nennen dabei elf Wörter pro Zeile⁶⁴². Dies entspricht, ganz grob kalkuliert und auf deutschsprachige Texte bezogen, einer Obergrenze von etwa 80 Anschlägen.

Noch höherzahlige Vorgaben fand ich in der Literatur nicht vor. In der Diskussion der empirischen Befunde zum Thema „Zeilenbreite“ werden zwar noch erheblich größere Zeichenmengen aufscheinen, doch für die Fachliteratur gilt: Jenseits der streng numerischen Vorgaben finden sich nur wenige ergänzende Hinweise und Ratschläge zu diesem typografischen Phänomen. Neutzling rät beispielsweise zur Differenzierung nach Schriftart: „Wählen Sie eine breitlaufende Type, können die Zeilen deutlich länger sein als bei einer schmalen Schrift“⁶⁴³. Leider lässt Neutzling nicht deutlich werden, ob sich der Terminus „Breite“ in diesem Zusammenhang auf eine absolute Breite bezieht oder eine bestimmte Anzahl Anschläge. Kommer und Mersin raten, für jüngeres, lese-ungeübteres Publikum eher kürzere als längere Zeilen vorzusehen⁶⁴⁴. Es liegt der Verdacht nahe, dass sie sich in dieser Empfehlung auf das Experiment von Bernard et al. stützen, das weiter unten noch zu diskutieren sein wird⁶⁴⁵. Was sie quantitativ unter „kürzer“ verstehen, lassen die Autorinnen allerdings offen.

Mit dem Zusammenspiel von Zeilenbreite und Zeilenabstand befasst sich schließlich Erben, der schreibt: „Je größer der Zeilenabstand ist, um so länger können die Zeilen sein – und umgekehrt“⁶⁴⁶. Auch Balzert empfiehlt: „(Bei) sehr langen Zeilen sollte ein größerer Zeilenabstand gewählt werden“⁶⁴⁷.

3.7.3.2 Der Aspekt der Zeilenfall-Ästhetik

Ein wichtiger Teilaspekt des typografischen Faktors „Zeilenbreite“ (der im Übrigen auch Bedeutung hat für die Wahl der Schriftausrichtung) ist der des Zeilenfalls. In der Literatur wird diesem Aspekt überraschenderweise praktisch keine Zeile gewidmet, lediglich Neutzling

⁶⁴¹ Vgl. Thissen 2000, S. 86

⁶⁴² Vgl. Dengler/Volland 2000, S. 51

⁶⁴³ Neutzling 2002, S. 110

⁶⁴⁴ Vgl. Kommer/Mersin 2002, S. 173

⁶⁴⁵ Vgl. Bernard et al. 2002a

⁶⁴⁶ Erben 1998, S. 92; vgl. ferner Kommer/Mersin 2002, S. 175

⁶⁴⁷ Balzert 2004, S. 175

stellt recht lapidar fest: „Die Zeilen dürfen nicht zu kurz angelegt werden, sonst stehen nur noch einzelne Wörter in einer Zeile (...)“⁶⁴⁸, Kiehn und Titzmann empfehlen, die Wahl der Zeilenbreite solle auch darauf hin abgeklopft werden, ob ein „annähernd ruhiges Schriftbild mit gleichmäßigem Grauwert“⁶⁴⁹ entstehe.

Dabei ist der Zeilenfall zumindest ästhetisch durchaus von Bedeutung, insbesondere, da im Web bis heute der Rausatz dominiert (siehe unten, Kapitel 3.8). Bezeichnet wird mit dem Begriff die unterschiedliche Länge aufeinanderfolgender Zeilen und damit der „Rhythmus“ des Weißraum-Musters, das sich (meistens) an der rechten Spaltenkante durch die horizontal variierenden Zeilenenden ausformt. Differieren die Zeilenlängen stark, ergeben sich große Varianzen und damit „Löcher“ an den jeweiligen Zeilenenden, der Satz wirkt insgesamt unruhig (oder, freundlicher formuliert: lebendig).

Zweifellos ist gerade die deutsche Sprache mit ihren vielen buchstabenreichen Komposita anfällig für „nervösen“ Zeilenfall; hinzu tritt erschwerend, dass bis heute in keinem Webbrowser ein Trennprogramm implementiert ist⁶⁵⁰, das verhindern könnte, dass Wortungetüme wie „Bundesnachrichtendienst“ schlimmstenfalls bereits in der Zeilenmitte zu einem Umbruch in die Folgezeile führen.

Dass der Aspekt der Zeilenfall-Ästhetik durchaus von Bedeutung ist, fand ich in einem kleinen „Zwischen-Experiment“ heraus. Dabei kreierte ich einen „Prüftext“, den ich zusammensetzte aus Aufmacher-Beiträgen der „Spiegel online“-Rubriken „Home“, „Politik“, „Wirtschaft“, „Panorama“, „Sport“ und „Kultur“, wie sie am 09.09.2006 um 21.15 Uhr im Web anzutreffen waren. Dieser Experimentaltext enthielt 18.067 Zeichen inklusive Leerzeichen. Ich untersuchte diesen Text nun auf seinen Zeilenfall; dazu schoss ich ihn so aus, dass die Zeilen im ersten Falle maximal 25 Zeichen pro Zeile Platz boten, im zweiten Falle 40, im dritten 55, im vierten 70, im fünften 85 und im sechsten 100 Zeichen pro Zeile. In jedem dieser Fälle erhob ich, wie viele der sich ergebenden Zeilen sich in der Anzahl ihrer Zeichen in welchem Umfange an den Maximalwert annäherten. Abbildung 3.23 verdeutlicht die Ergebnisse diagrammatisch.

Es zeigte sich beispielsweise, dass nur etwa 43 Prozent der Zeilen, für die eine maximale (in der Grafik: „nominale“) Zeichenanzahl von 25 vorgesehen war, zwischen 90 und 100 Prozent dieser Zeichenanzahl (also 23 bis 25 Zeilen-Anschläge) tatsächlich auch aufwiesen. Etwa 21 Prozent der Zeilen bewegten sich dagegen im Bereich von 80 bis 90 Prozent der Maximal-Zeichenanzahl (20 bis 22 Zeichen), weitere 36 Prozent der Zeilen enthielten noch weniger Zeichen. Diese Verteilung spricht für ein ausgesprochen unruhiges Satzbild.

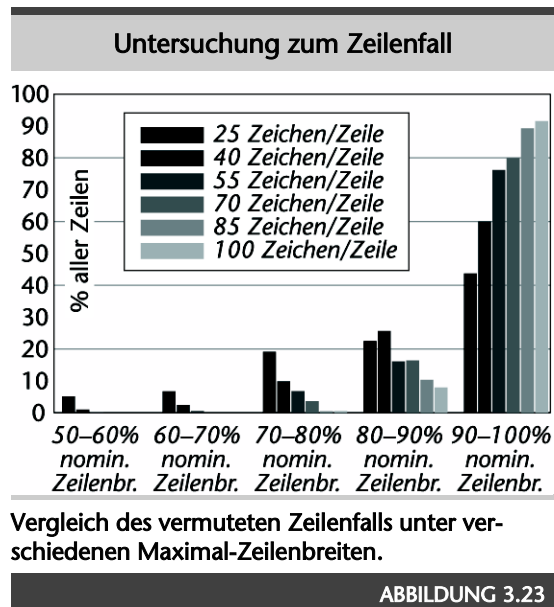
⁶⁴⁸ Erben 1998, S. 92

⁶⁴⁹ Kiehn/Titzmann 1998, S. 46

⁶⁵⁰ Vgl. u.a. Hellbusch 2005, S. 93

Auch bei zulässigen 40 Zeichen pro Zeile variierten die Breiten noch sehr stark:

Nur 60 Prozent der sich ergebenden Zeilen erreichten 36 bis 40 Anschläge, immer noch jede siebte Zeile jedoch enthielt weniger als 29 Zeichen. Ein etwas ruhigeres Satzbild versprechen dagegen Zeilen mit maximal 55 Anschlägen: In diesem Versuchsmodus lagen bereits 76 Prozent der Zeilen bei einer Zeichenzahl von 50 bis 55 Anschlägen, weitere 16 Prozent der Zeilen zwischen 44 und 49 Zeichen pro Zeile.



Es scheint also, als werde der Zeilenfall insgesamt merklich ruhiger bei einer Verbreiterung von Zeilen von 25 auf 55 Anschläge als Maximalwert pro Zeile. Weitere Verbreiterungen erbringen offenbar eine weitere Harmonisierung des Satzbildes, jedoch nicht mehr in so ausgeprägtem Umfang wie in den Bereichen der kürzeren Zeilenlängen.

Zwei abschließende Bemerkungen zu diesem Aspekt seien gestattet. Erstens: Natürlich ist die Anzahl der Zeichen pro Zeile bei Proportionalschriften, wie sie im Web dominieren, nicht absolut gleichzusetzen mit der „physischen“ Zeilenbreite, die beispielsweise in Zentimetern benennbar wäre. Näherungsweise ist hier aber sicherlich eine Kongruenz zu unterstellen, und für eine Kleinst-Untersuchung wie die eben beschriebene reicht diese Annahme allemale aus.

Zweitens: Es ist interessant, dass die durchschnittliche Anzahl von Zeichen pro Zeile in der eben vorgestellten Untersuchung fast durchgehend um etwa 4 Zeichen unter der Maximal-Zeichenzahl lag – und zwar vollkommen unabhängig vom vorgegebenen Höchstwert. Bei 25 Anschlägen als gegebenem Maximalwert etwa lag die durchschnittliche Zeichenzahl pro Zeile bei 21,1, bei 40 Anschlägen waren es 36,1, bei 55 Anschlägen 51,1, bei 70 Anschlägen 66,1. Selbst bei 100 Maximalansschlägen lag der Durchschnittswert noch bei 96,3. Sollte ich hier – quasi en passant – auf eine nützliche Faustregel gestoßen sein? Es wäre interessant, diesen Aspekt weiter zu verfolgen und eventuell durch ergänzende Studien zu erhärten. Im Rahmen dieser Arbeit freilich wird dies nicht geschehen – bleiben wir beim Thema!

3.7.4 Empirische Studien zur Zeilenbreite am Bildschirm

3.7.4.1 Die Times in verschiedenen Zeilenbreiten

Youngman und Scharff untersuchten die Lesbarkeit der Schriftart Times 1998 in verschiedenen Zeilenbreiten (4, 6 und 8 Inches, das entspricht in etwa 10, 15 und 20 Zentimetern)

und mit unterschiedlich bemessenem umgebenden Raum um den Textblock als Ganzes⁶⁵¹. Die verglichenen Textblöcke wiesen damit durchschnittlich 65, 100 und 140 Zeichen pro Zeile auf.

27 Probanden wurden aufgefordert, die dargebotenen Texte aufmerksam zu lesen. In den Texten fanden sich Aufmerksamkeitswörter, die geometrische Formen schriftlich bezeichneten (wie beispielsweise „square“, also „Quadrat“, oder „star“, also „Stern“). Unterhalb der Texte waren diese geometrischen Formen grafisch auf dem Bildschirm abgebildet und per Maus anklickbar. Die Probanden wurden aufgefordert, beim Entdecken eines Aufmerksamkeitswortes auf das entsprechende geometrische Symbol zu klicken; aus der Messung der Geschwindigkeit, in der diese Aufgabe gelöst wurde, ermittelten Youngman und Scharff die individuelle Lesegeschwindigkeit.

Diese fiel für die 4-Inch-Zeilen fast durchgehend besser aus als für die 6-Inch-Variante; die durchschnittliche gemessene Reaktionszeit war für die 4-Inch-Zeilen mit 13,6 Sekunden signifikant besser als für die 6-Inch-Zeilen (14,8 Sekunden), bewegte sich dagegen etwa auf einem Niveau mit den Ergebnissen für die breitesten gemessenen Zeilen. Zwischen schmaler und breiter Zeile fand sich entsprechend kein signifikanter Unterschied der Messzeiten.

3.7.4.2 6 Zeilenbreiten im Vergleich

Mithilfe gewisser Herleitungen ist auch die Studie von Andreyev und Martynov aus dem Jahre 2000 durchaus verwertbar für die Diskussion dieser Arbeit⁶⁵². Zwar war es das eigentliche Erkenntnisinteresse der Forscher, die Vor- und Nachteile von Mehrspaltensatz am Bildschirm zu untersuchen; doch ergeben unterschiedliche Spaltenanzahlen eben auch unterschiedliche Zeilenbreiten, so dass Schlussfolgerungen möglich sind.

Analysiert man die Versuchsanlagen von Andreyev und Martynov, so lässt sich induzieren, dass die Forscher Zeilen verglichen, in denen 35, 50, 65, 115, 160 oder 200 Zeichen im Schnitt Platz fanden. Die Forscher maßen dabei zum einen die reine Lesezeit (die insgesamt 30 Probanden signalisierten das Ende eines Lesedurchgangs jeweils per Mausklick), das Leseverständnis wurde im Anschluss an jeden Text durch inhaltliche Multiple-Choice-Fragen erhoben. Schließlich wurden die Probanden noch um ein Urteil gebeten, welche der Konstellationen sie als am angenehmsten empfanden.

In allen drei Messkategorien schnitt die Variante „35 Anschläge“ am besten ab. Lediglich in der Lesezeit erwiesen sich die „115er“-Zeilen im Schnitt als minimal (0,38 Prozent) besser lesbar. Die drittbeste Durchschnittszeit erzielten die Probanden in der doch sehr großzügigi-

⁶⁵¹ Vgl. Youngman/Scharff 1998

⁶⁵² Vgl. Andreyev/Martynov 2000

gen Zeilenbreite mit 160 Anschlägen, die viertbeste Zeit entfiel auf die „65er“-Variante. Signifikant war keiner der gemessenen Unterschiede.

Im Urteil der Teilnehmenden schnitten die drei schmalen Zeilenbreiten eindeutig am besten ab: Auf einer Skala von 1 bis 10 (1 = unzufrieden, 10 = sehr zufrieden) erzielten die schmalen Breiten Werte zwischen 6,2 und 6,9, die breiteren Zeilen dagegen nur Noten zwischen 3,6 und 5,1. Andererseits wurden im Anschluss an die „65er“-Texte die meisten Fehler in der Inhaltsabfrage festgestellt, die wenigsten in der „35er“-Version, die zweitwenigsten bei den Texten mit einer Durchschnitts-Zeilenbreite von 160 Anschlägen.

3.7.4.3 Vergleich zweier Zeilenbreiten

Auch Lam et al. zielten in ihrem Experiment im Jahre 2000 ursprünglich auf die Messung von Unterschieden zwischen Ein- und Dreispaltensatz am Bildschirm⁶⁵³. Analog zu Andreyevs und Martynovs Experiment lassen sich aber auch aus diesem Grundansatz Erkenntnisse über die Wirkungen von Zeilenbreiten herleiten; bei Lam et al. wurden Zeilen vorgelegt, die entweder 110 oder 35 Zeichen Platz boten. Die insgesamt 20 Teilnehmer (16 Männer, 4 Frauen) im Alter von 19 bis 24 Jahren wurden gebeten, die Texte zu lesen und das Ende eines Rezeptionsprozesses per Mausclick zu signalisieren. Ferner wurden auch bei Lam et al. im Anschluss Verständnisfragen gestellt, um die Leseeffizienz der Probanden zu messen.

Die Messung der Lesezeiten erbrachte ein erstaunliches Resultat: Die schmalen Zeilen wurden im Schnitt in 230,40 Sekunden durchgearbeitet, die breiten in 323,05 Sekunden – das entspricht einer Verlangsamung der Geschwindigkeiten von fast 29 Prozent, die sich natürlich auch als statistisch signifikant erwies. Absolut gleichauf lagen die Teilnehmenden dagegen in der Fehlerrate bei der Beantwortung der Verständnisfragen zum Text.

3.7.4.4 Drei Zeilenbreiten im Abgleich

Das obige Verfahren soll an dieser Stelle ein drittes Mal zur Anwendung kommen: Auch Baker maß 2005 ursprünglich die Wirkung von Spaltensatz auf die Lesbarkeit am Bildschirm⁶⁵⁴; aus seinem Untersuchungsdesign lassen sich aber wiederum auch Erkenntnisse bezüglich der Wirkung unterschiedlicher Zeilenbreiten extrahieren. Eine Analyse der vorgelegten Experimentaltexte zeigte mir, dass Baker Zeilen bereitstellte mit entweder durchschnittlich 30, 45 und 90 Zeichen.

66 Studierende im Durchschnittsalter von 22,8 Jahren nahmen an dem Experiment teil. Je 22 lasen den einen Experimentaltext, den Baker vorhielt, in jeweils einer der drei Zeilenbrei-

⁶⁵³ Vgl. Lam et al. 2000

⁶⁵⁴ Vgl. Baker 2005

ten – von diesen 22 Probanden erhielten wiederum je elf den Text in linksbündigem Flattersatz vorgelegt, die anderen elf lasen den Text in Blocksatz (dazu später mehr im Kapitel „Ausrichtung“). Die Texte mit der schmalsten Zeilenbreite wurden – über beide Ausrichtungsarten hinweg – mit durchschnittlich 245 Wörtern pro Minute gelesen, die Texte in mittlerer Breite mit 256 Wörtern pro Minute, die Texte in der breitesten Zeilenanlage mit 237 Wörtern pro Minute. Die breiteste gemessene Zeilenlänge war dabei schwach signifikant schlechter lesbar als die mittlere Breite.

Das Leseverständnis – erhoben mithilfe von Inhaltsfragen im Anschluss an den Leseprozess – variierte nur unerheblich abhängig von der Zeilenbreite. Auch die Befragung der Teilnehmer, wie „zufrieden“ sie mit der ihnen vorgelegten Zeilenbreite waren, ergab keine nennenswerten Resultate: Aus statistischer Sicht waren alle Zeilenbreiten identisch gut rezipierbar unter dem Kriterium des Leseverständnisses.^q

3.7.4.5 25 und 80 Anschläge im Vergleich

Dyson und Kipping erforschten 1997 mit 18 Teilnehmenden Lesegeschwindigkeiten und Textverständnis zweier verschiedener Zeilenbreiten⁶⁵⁵. Verglichen wurden Zeilen mit durchschnittlich 25 beziehungsweise 80 Anschlägen im Schnitt. Dabei waren die Texte mit den kürzeren Zeilen einmal so gestaltet, dass die Nutzer scrollen mussten, um die Inhalte in Gänge zu erfassen, ein weiteres Mal kam das „paging“ zum Einsatz: Der Text war auf mehrere Einzelseiten verteilt, per Tastendruck gelangten die Probanden auf die jeweilige Fortsetzungsseite (der Testtext mit der breiteren Zeilenbreite passte ohnehin auf eine einzige Bildschirmseite). Die Texte waren gesetzt in einer 10-Punkt-*Arial* mit 12 Punkt Zeilenhöhe.

Die Ergebnisse zeigen ein differenziertes Bild. Absolut am schnellsten gelesen wurden im Schnitt die kurzen Zeilen im „paging“-Modus (603 Sekunden). Im Scrolling-Modus stieg die Lesezeit der kurzen Zeilen merklich (um 9 Prozent) an auf 659 Sekunden; die breiten Zeilen wurden im Schnitt in 637 Sekunden rezipiert. Die kurzen „paging“-Zeilen waren damit signifikant schneller lesbar als die beiden anderen Darbietungsformen. Das Leseverständnis bewegte sich in allen drei Testformen auf praktisch vollkommen identischem Niveau. Befragt, welche Testform sie individuell als die „angenehmste“ empfanden (breit oder kurz), entschieden sich fast alle Probanden für die kürzeren Zeilen.

3.7.4.6 6 Zeilenbreiten im Abgleich

Dyson und Kipping untersuchten 1998 die Lesegeschwindigkeiten und das Textverständnis von sechs Texten, die in verschiedenen Zeilenbreiten gesetzt waren⁶⁵⁶. Die Zeilen boten durchschnittlich 25, 40, 55, 70, 85 oder 100 Zeichen Platz. Die Probanden erhielten die

⁶⁵⁵ Vgl. Dyson/Kipping 1997, S. 703ff.

⁶⁵⁶ Vgl. Dyson/Kipping 1998, S. 151ff.

sechs Texte nacheinander zu lesen. Da die Texte mit etwa 800 Wörtern in Ganzheit deutlich über eine Bildschirmhöhe hinausragten, erhielt je die Hälfte der Teilnehmenden die Möglichkeit, sich durch den Text vertikal zu scrollen, die andere Hälfte dagegen konnte sich seitenweise durch die Texte bewegen, die in Passagen portioniert waren, die niemals mehr als eine Bildschirmhöhe in Anspruch nahmen. Die Teilnehmenden signalisierten das jeweilige Ende eines Leseprozesses per Tastendruck. Anschließend stellten die Forscher gezielte Fragen, um zu messen, welche Textinhalte die Probanden erinnerten. 24 Teilnehmende lieferten schließlich verwertbare Ergebnisse.

Zunächst zeigte sich, dass die Teilnehmenden, die sich seitenweise durch die Texte arbeiten konnten, durchweg um 3 bis 7 Prozent bessere Lesezeiten erzielten als die „Scroller“; insgesamt aber erwies sich der Einfluss der „Bewegungsmethode“ nicht als signifikant beeinflussender Faktor auf die Lesezeiten. Unter den „Scrollern“ – auf die hier das Augenmerk konzentriert werden soll, da diese Teilnehmenden meines Erachtens „web-ähnlicher“ lasen als die Vergleichsgruppe – wurden Texte in der Zeilenlänge „100 Zeichen pro Zeile“ durchschnittlich in 239 Sekunden bearbeitet, mit jeder Verkürzung der Zeilenbreite stiegen die Lesezeiten an, wenn auch jeweils nur in sehr kleinen Margen (85 Anschläge: 250 Sekunden; 70 Anschläge: 253 Sekunden; 55 und 40 Anschläge: 254 Sekunden; 25 Anschläge: 256 Sekunden). Die kürzesten Zeilen wurden damit um 6,5 Prozent langsamer gelesen als die breitesten; diese Differenz erwies sich als statistisch signifikant, auch in der Vergleichsgruppe derjenigen, die seitenweise durch die Testtexte gegangen waren. Ansonsten fanden Dyson und Kipping keine Auffälligkeiten. Auch in den Kontrollabfragen zum Leseverständnis ermittelten die Forscher keinerlei Einflüsse der Zeilenbreiten auf die Inhaltserinnerung.

Abschließend wurden die Teilnehmer befragt, welche der dargebotenen Zeilenbreiten sie als „am besten lesbar“ empfänden. Über die beiden Bewegungsmuster hinweg („scrolling“ oder „paging“) favorisierten die Probanden die „55er“-Zeilen. Es folgten die Zeilenbreiten „40“ und „70“, wobei die „Scroller“ die „40er“-Breite bevorzugten, die Vergleichsgruppe die „70er“-Breite. Als am schlechtesten lesbar werteten beide Gruppen (in dieser Reihenfolge) die „Extremwerte“ „100“, „25“ und „85“.

In einem zweiten Telexperiment überprüften Dyson und Kipping diese Resultate, diesmal nur mit den drei Zeilenbreiten „25“, „55“ und „100“. Das Ergebnis fiel wiederum vielschichtiger aus: Im „paging“-Modus ergaben sich praktisch gar keine Unterschiede in den Lesezeiten der drei Zeilenbreiten-Varianten. Im „scrolling“-Modus dagegen wurden die kurzen Zeilen erheblich langsamer rezipiert (331 Sekunden im Schnitt) als die mittlere (304 Sekunden) und die längste (282 Sekunden). Insgesamt fanden Dyson und Kipping, dass die Zeilenbreite als Einflussfaktor einen statistisch signifikanten Effekt auf die Lesezeiten ausübte; der Unterschied zwischen kürzester und längster Zeile im „scrolling“-Modus war ebenfalls signifikant. In der Überprüfung des Leseverständnisses ergaben sich wiederum keinerlei Auffälligkeiten

zwischen den verschiedenen Testbreiten. Insgesamt bevorzugten die Teilnehmenden auch in diesem Experiment die „55er“-Breite als die subjektiv „lesbarste“.

3.7.4.7 Zeilenbreiten und Leseverständnis

Dyson und Haselgrove untersuchten 2001 die Wirkung unterschiedlich bemessener Zeilenabstände auf das Leseverständnis⁶⁵⁷. 36 Teilnehmer lasen verschiedene Texte aus dem „National Geographic“, die in einer 10-Punkt-Arial mit 12 Punkt Zeilenabstand sowie Zeilenbreiten von 25, 55 oder 100 Zeichen aufbereitet waren; die eine Hälfte der Teilnehmer wurde gebeten, den Text „sehr rasch“ zu lesen, die andere Hälfte war angewiesen, einen möglichst „normalen“ Rezeptionsprozess durchzuführen. Im Anschluss wurden den Probanden Verständnisfragen zu den Texten gestellt, aus denen Dyson und Haselgrove den individuellen Lernerfolg folgerten.

Die Forscher stellten zunächst fest, dass die „Schnell-Leser“ durchweg und in jeder Zeilenbreite im Schnitt weniger Verständnisfragen korrekt beantworteten als die „Normal-Leser“. Die Lesegeschwindigkeit wurde als statistisch signifikanter Einflussfaktor auf das Textverständnis ermittelt. Ebenso der Faktor „Zeilenbreite“: Auch dieser erwies sich als signifikant Einfluss nehmend auf das Leseverständnis. Die meisten richtigen Antworten bei den „Normal-Lesern“ fanden sie in den Befragungen, die auf die „55er“-Zeilenbreiten folgten (2,3 von 3,1 Punkten), am zweitbesten fiel die Zeilenbreite mit 25 Anschlägen aus (2,15), die schlechtesten Werte wurden in den langen Zeilen gemessen (2,1). Bei den „Schnell-Lesern“ generierten die Zeilenbreiten „25“ und „55“ praktisch identische Messresultate (je 1,8), die langen Zeilen waren auch hier die uneffektivsten (1,7).

Ferner fanden Dyson und Haselgrove heraus, dass die zu schnellem Lesen animierten Probanden die Texte umso rascher durcharbeiteten, je breiter die Zeile gesetzt war (25 Anschläge: 0,79 Wörter pro Sekunde; 55 Anschläge: 0,85; 100 Anschläge: 0,87); die Befunde erwiesen, dass die Zeilenbreite einen signifikanten Einfluss auf die Lesegeschwindigkeit ausübte. Bei den „Normal-Lesern“ dagegen fanden Dyson und Haselgrove die größte Geschwindigkeit bei 55 Anschlägen pro Zeile.

3.7.4.8 Zeilenbreiten – wie sie auf Kinder und Erwachsene wirken

Bernard et al. erforschten 2002 die Lesegeschwindigkeiten und das Leseverständnis von 20 Erwachsenen im Alter von 18 bis 61 Jahren sowie 20 Kindern im Alter von 9 bis 11 Jahren am Bildschirm unter den Bedingungen unterschiedlicher Zeilenbreiten⁶⁵⁸. Die Probanden bekamen nacheinander Texte in drei verschiedenen Zeilenbreiten vorgelegt: 132 Zeichen pro

⁶⁵⁷ Vgl. Dyson/Haselgrove 2001. S. 585ff.

⁶⁵⁸ Vgl. Bernard et al. 2002a

Zeile („volle Breite“), 76 Zeichen pro Zeile („mittlere Breite“) und 45 Zeichen pro Zeile („enge Breite“). Gesetzt waren die Texte in *Arial* in 12 Punkt Größe; die Erwachsenen hatten sich mit psychologischen Themen auseinanderzusetzen, die Kinder lasen kurze Märchen.

Die Ergebnisse überraschen: Bei den Erwachsenen lagen die erhobenen Durchschnittszeiten zwischen 363 Sekunden für die „mittlere Breite“ und 370 Sekunden für die „volle Breite“ („enge Breite“: 366 Sekunden) – angesichts dieser schmalen Marge von gerade einmal sieben Sekunden verwundert es kaum, dass Bernard et al. keine signifikanten Unterschiede zwischen den Testbreiten feststellen konnten. Ähnlich fiel das Ergebnis für die Kinder aus: Ganz leicht am besten war die mittlere Lesezeit bei Texten, die in der „engen Breite“ (266 Sekunden) dargeboten worden waren, es folgten jedoch dichtauf die „volle Breite“ (276) und die „mittlere Breite“ (279).

Ebenfalls keine signifikanten Befunde ergab die Befragung der Testpersonen zu den Textinhalten; die Resultate dieser Verständniskontrollen wurden mit den Lesezeiten verrechnet. Bei den Erwachsenen lag nunmehr die „volle Breite“ (verrechnete 425 Sekunden) sehr leicht vor der „engen Breite“ (443) und klar vor der mittleren Testbreite (463 Sekunden). Diese „enge Breite“ lag bei den Kindern wiederum nach der Verständnisabfrage vorne (330 Sekunden), vor der „mittleren Breite“ (359) und der größten Zeilenlänge (363 Sekunden).

Befragt nach ihrem individuellen Urteil, erklärten die Erwachsenen je zur Hälfte, dass ihnen die „mittlere“ oder die „enge Breite“ am ehesten zusage, die größte gemessene Zeilenbreite favorisierte niemand. Die Kinder sprachen sich mit deutlicher Mehrheit für die schmalen Zeilenbreiten aus.

Die Autoren mutmaßen in ihrer Zusammenfassung, dass sich potenziell bessere Lesezeiten für die kürzeren Zeilenbreiten im Test nicht im Ergebnis niederschlagen konnten, weil sie durch erhöhte Scrolltätigkeiten (natürlich fielen die kurzzeiligen Texte zeilenzahlmäßig größer aus) nivelliert worden seien.

3.7.4.9 Vier Zeilenbreiten im Vergleich

Shaikh untersuchte 2005 an 20 Probanden die Auswirkungen variierteter Zeilenbreiten auf Lesegeschwindigkeit, Textverständnis und Nutzerzufriedenheit⁶⁵⁹. Er präsentierte den Probanden Nachrichtentexte von jeweils etwa 375 Wörtern Länge in den Zeilenbreiten 35 (Zeichen pro Zeile), 55, 75 und 95. Die Texte waren aufbereitet in der Schriftart *Arial* in 10 Punkt Größe, der Zeilenabstand betrug 12 Punkt. Die Testtexte wurden jeweils auf mehrere (offline abgerufene) Webseiten verteilt; die User konnten sich über einen „Weiter“-Button bis zum jeweiligen Textende durcharbeiten. Gemessen wurde die Dauer zwischen Erscheinen der je-

⁶⁵⁹ Vgl. Shaikh 2005

weils ersten Seite eines Textes bis zum Mausklick auf den Button „Weiter“ auf der jeweils letzten Seite; diese Zeit wurde als „Lesezeit“ gewertet.

Die meisten Wörter pro Minute lasen die Teilnehmenden in Shaikhs Experiment in den Texten, die in Zeilen mit durchschnittlich 95 Zeichen pro Zeile aufbereitet waren (178,82 Wörter). Diese Konstellation erwies sich damit als statistisch signifikant besser lesbar als die übrigen drei. Je schmaler die Zeilen ausfielen, desto schlechter fielen im Übrigen die Werte aus – wenn auch unerheblich differierend: Bei 75 Zeichen pro Zeile wurden 169,44 Wörter pro Minute rezipiert, bei 55 Zeichen pro Zeile waren es 167,38, bei 35 Zeichen pro Zeile 167,21.

Um sicher zu stellen, dass die Probanden die Texte auch sorgfältig gelesen hatten, konfrontierte Shaikh sie im Anschluss an den Leseprozess mit inhaltlichen Fragen zum rezipierten Text. Die erhobenen Werte aus dieser Befragung multiplizierte er mit den gemessenen Lesegeschwindigkeiten; übrig blieb ein statistisch signifikant besserer Wert für die Zeilenbreite „95“ gegenüber der Breite „35“. Insgesamt zeigte sich, dass die Zeilenbreite keinen signifikanten Einfluss auf das „Gesamt-Verstehen“ der Textinhalte ausübte; egal, welche Zeilenbreite vorlag, die Teilnehmer erfassten die Texte inhaltlich äquivalent.

Ausgesprochen widersprüchlich fielen die Ergebnisse aus, als Shaikh abschließend seine Probanden nach der von ihnen bevorzugten Zeilenbreite befragte. 30 Prozent erklärten die „35er“-Zeilen zum Favoriten, ebenfalls 30 Prozent die „95er“-Variante. Befragt nach der in ihren Augen schlechtesten Variante wiederum nannten 45 Prozent die „35er“-Version und 55 Prozent die „95er“-Breite – diese Breiten waren also paradoxerweise die beliebtesten wie die unbeliebtesten zugleich.

3.7.4.10 Textverständnis in zwei Zeilenbreiten

McMullan et al. untersuchten 2002, ob die Menge aus einem Bildschirmtext entnommener Informationen durch die Zeilenbreite beeinflusst wird⁶⁶⁰. 67 Teilnehmer wurden mit verschiedenen populärwissenschaftlichen Texten konfrontiert, die durchschnittlich 55 Zeichen pro Zeile aufwiesen oder 115 Zeichen pro Zeile. Mithilfe von Verständnisfragen ermittelten die Forscher im Anschluss an die Lesedurchgänge den individuellen Lernerfolg der Probanden.

Die Zeilenbreite erwies sich dabei als statistisch nicht signifikanter Einflussfaktor auf das Textverständnis. Sie vermerkten durchschnittlich 63 Prozent korrekte Antworten, nachdem Texte mit 115 Zeichen pro Zeile vorgelegt worden waren, 58 Prozent der Antworten waren korrekt nach der Rezeption der kürzeren Zeilen.

⁶⁶⁰ Vgl. McMullin et al. 2002, S. 19ff.

3.7.5 Zwischenfazit: Zeilenbreiten

Sowohl die Einlassungen der Fachliteratur als auch die Ergebnisse bisheriger empirischer Bemühungen zum Phänomen „Zeilenbreite“ ergeben ein ausgesprochen diffuses Bild. In der Literatur wird die empfehlenswerte Zeilenbreite mal auf „maximal 35“, mal auf „optimal 70“ Anschläge pro Zeile taxiert, eine echte Mehrheitsmeinung ist aus dem Gesamt der Meinungen praktisch nicht extrahierbar.

Ähnliches gilt für die empirische Forschung. Wiewohl durchaus eine Reihe von Ergebnissen vorliegt, variieren die Messwerte so stark – ja, noch viel stärker als in der Fachliteratur –, dass auch hier kaum eine Art gemeinsamer Linie entwickelbar scheint. Nur eine Erkenntnis ist so gut wie allen experimentellen Erhebungen gemein: Die Zeilenbreite scheint durchaus Einfluss zu nehmen auf die Lesegeschwindigkeit; für das Leseverständnis hingegen scheint die Anzahl der Zeichen pro Zeile nahezu irrelevant – selbst bei weit über 100 Anschlägen pro Zeile, Werten also, die für die Autoren der zitierten Literatur augenscheinlich jenseits jeder Diskussionswürdigkeit liegen.

3.8 Schriftausrichtung

3.8.1 Die Schriftausrichtung als typografisches Phänomen

Es ist in der Regel ein Produkt des Zufalls, wenn übereinander liegende Zeilen exakt dieselbe physische Breite erzielen – dieser Aspekt wurde bereits im Zusammenhang mit der „Zeilenfall-Ästhetik“ in Kapitel 3.7.3.2 angesprochen. An der vorliegenden Arbeit ist dies gut zu belegen: Die jeweils rechten Kanten der Textzeilen bilden in der vertikalen Grobbetrachtung ein unregelmäßiges „Fließmuster“, entfalten eine Schriftausrichtung (oder auch: einen Satzausschluss⁶⁶¹) in der Art des „Rausatzes“⁶⁶². Dieser ist meist linksbündig ausgelegt, das heißt, die linke Satzkannte bildet – wie in dieser Arbeit – die fixe horizontale Startmarke jeder Einzelzeile. Als Varianten existieren noch rechtsbündiger sowie zentrierter Rausatz.

Der Begriff des „Rausatzes“ findet sich bis heute relativ selten in der Literatur – es dominiert, jenseits der dezidiert typografischen Fachlektüre zumindest, die Bezeichnung „Flattersatz“ als vermeintliches Synonym. Dabei bezeichnet Flattersatz zwar durchaus eine ähnliche, aber grundsätzlich eben nicht identische Ausrichtungs-Variante. Im Gegensatz zum Rausatz wird im Flattersatz grundsätzlich auf Wort- und Silbentrennungen verzichtet⁶⁶³, was die physischen Zeilenlängen notwendigerweise stärker differieren und mithin den „Fallrhythmus“ der Zeilen deutlich unruhiger ausfallen lässt⁶⁶⁴.

⁶⁶¹ Vgl. Gulbins/Kahrmann 1992, S. 42

⁶⁶² Vgl. Bollwage 2001, S. 48

⁶⁶³ Vgl. Bergner 1990, S. 146

⁶⁶⁴ Vgl. Meissner 1992, S. 47

Im klassischen Buch- und Zeitungsdruck gilt Flattersatz wie Rausatz welcher Achsenorientierung auch immer gemeinhin als unerwünscht, zumindest in Fließtexten; durch gezielte Vergrößerung der Wortabstände innerhalb einer Zeile werden in diesen Genres die Zeilen auf identische Breite manipuliert, es entsteht „Blockatz“ (in angelsächsischen Ländern tritt zur Verbreiterung des Wortabstandes auch häufig die Vergrößerung der Buchstabenabstände, der Laufweite; im deutschsprachigen Raum ist dies eher unüblich).

3.8.2 Manipulation der Schriftausrichtung in HTML-basierten Browsern

Zeitgenössischen Browsern fehlt, was im digital basierten Schriftsatz sonst längst bare Selbstverständlichkeit ist: eine Silbentrennungs-Funktion. In diesem Sinne kennen HTML und CSS prinzipiell nur vier Ausrichtungs-Varianten, die ausnahmsweise einmal in sämtlichen Browsern problemlos und mit kontrollierbarem Erfolg umsetzbar sind: Blockatz sowie Flattersatz links- und rechtsbündig sowie zentriert. Rausatz ist im Web – eben angesichts und wegen der fehlenden Silbentrennfunktion – nicht realisierbar. Der entsprechende HTML-Befehl lautet „text-align“ und kennt die Ausprägungen „left“, „right“, „center“ sowie „justify“ („gerichtet“, konkret also: Blockatz).

3.8.3 Empfehlungen und Vorgaben aus der Literatur

Seltene Harmonie entfaltet die gesichtete Literatur, was Empfehlungen und Vorgaben zum Thema „Schriftausrichtung“ angeht: Praktisch durch die Bank wird linksbündiger Flattersatz empfohlen auf Webseiten.

Neutzling begründet seine Präferenz so: „Die linke Satzkannte (...) bildet eine optische Führungslinie, an die das Auge am Ende einer Zeile automatisch zurückspringen kann“⁶⁶⁵. Auch Nielsen fordert, und zwar gewohnt apodiktisch und argumentativ bar: „Auf jeder Website ist der Text linksbündig zu formatieren“⁶⁶⁶. Als Befürworter des linksbündigen Rausatzes geben sich auch Thissen⁶⁶⁷ sowie Kommer und Mersin⁶⁶⁸, gleichfalls Hellbusch⁶⁶⁹, Kiehn und Titzmann⁶⁷⁰ und Balzert⁶⁷¹. Auch für Bollwage muss es im Web „natürlich Flattersatz“ sein⁶⁷², ebenso für Schäffer⁶⁷³. Mehr oder minder ausdrücklich warnt auch die Mehrzahl der Autoren davor, die stabile Achse spiegelverkehrt anzusetzen; rechtsbündiger Satz verlangsame das Lesen: „Da das Auge nicht immer an die gleiche Stelle springen kann, beginnt es zu suchen,

⁶⁶⁵ Neutzling 2002, S. 112

⁶⁶⁶ Nielsen 2000a-, S. 126

⁶⁶⁷ Vgl. Thissen 2003, S. 98

⁶⁶⁸ Vgl. Kommer/Mersin 2002, S. 170

⁶⁶⁹ Vgl. Hellbusch 2005, S. 93

⁶⁷⁰ Vgl. Kiehn/Titzmann 1998, S. 47

⁶⁷¹ Vgl. Balzert 2004, S. 183

⁶⁷² Vgl. Bollwage 2001, S. 102

⁶⁷³ Vgl. Schäffer 2001, S. 192

der Lesefluss wird unterbrochen, das Lesen des Textes wird anstrengend“⁶⁷⁴. Dengler und Volland verbannen die dritte denkbare Flattersatz-Version gleich mit; ihnen zufolge ist auch „zentrierter Fließsatz am Bildschirm (...) tabu“⁶⁷⁵.

Einzig Lamprecht nimmt mit seiner Meinung eine Außenseiterstellung ein. Seiner Ansicht nach wird nämlich „ die Lesearbeit der Augen am besten durch Blocksatz unterstützt“⁶⁷⁶. Argumente oder gar empirische Belege für seine Exklusiv-Ansicht liefert Lamprecht nicht. So bleibt das Feld jenen Gegnern des Blocksatzes überlassen, die erkannt haben, dass in diesem Satz-Modus das „Fehlen von Silbentrennprogrammen in den Browsern (...) zu sehr großen Löchern im Satzbild“⁶⁷⁷ führt – insbesondere, wenn ein Nutzer die Schriftgröße noch über das voreingestellte Maß hinaus anhebt und schlimmstenfalls nur noch ein Wort in einer Zeile Platz findet: Dann sei „eine große Lücke kaum von einem Zeilenumbruch zu unterscheiden“⁶⁷⁸. Freilich sinkt die Gefahr allzu ausgreifender Löcher im Blocksatz-Zeilenbild, je mehr Zeichen die Zeilen zu beherbergen in der Lage sind (siehe auch das Kapitel und die Kurz-Untersuchung zum Thema „Zeilenfall“ in Kapitel 3.7.3.2), weshalb Schweizer (bei aller eindeutigen Präferenz für Flattersatz) „in Ausnahmen“ auch Blocksatz genehmigt⁶⁷⁹.

3.8.4 Empirische Erkenntnisse zur Schriftausrichtung am Bildschirm

In der bereits vorgestellten Studie von Baker⁶⁸⁰ verglich dieser, wie gezeigt, drei verschiedene Zeilenbreiten auf Lesegeschwindigkeit, Textverständnis und Nutzerzufriedenheit. Dabei setzte er jede Zeilenbreiten-Variante einmal in linksbündigem Flattersatz und einmal in Blocksatz. Interessanterweise erwies sich die Variante „Zeilenbreite 45 Zeichen pro Zeile/Blocksatz“ als die durchschnittlich am besten (schnellsten) lesbare im gesamten Experiment. Allerdings ist auch hervorzuheben, dass die beiden anderen gemessenen Zeilenbreiten im Flattersatz merklich bessere Lesezeiten induzierten als im Blocksatz. Baker befasste sich eingehender mit diesem Resultat und kam nach einer nachträglichen Analyse der erhobenen Einzelzeiten zu dem Schluss, dass „gute“ und „schlechte“ Leser unterschiedlich auf die Ausrichtungsvarianten reagierten: „Während schnelle Leser die besten Werte in der Zwei-Spalten-Variante (Zeilenbreite 45 Anschläge, M.L.) im Blocksatz erzielten, profitierten langsame Leser eher vom Ein-Spalten-Layout (Zeilenbreite 30 Anschläge, M.L.) in Flattersatz. Dies legt nahe, dass Nutzer die Möglichkeit haben sollten, Webseiten auf ihre Leseerfordernisse hin anzupassen“⁶⁸¹. (Übersetzung M.L.).

⁶⁷⁴ Neutzling 2002, S. 112

⁶⁷⁵ Dengler/Volland 2000, S. 50

⁶⁷⁶ Lamprecht 2002, S. 38

⁶⁷⁷ Vgl. Lynch/Horton 1999, S. 82f.; vgl. ferner Lankau 2000, S. 368; vgl. ferner Seibert/Hoffmann 2006, S. 277; vgl. ferner Schäffer 2001, S. 192; vgl. ferner Balzert 2004, S. 183; vgl. ferner Neutzling 2002, S. 113

⁶⁷⁸ Hellbusch 2005, S. 93

⁶⁷⁹ Vgl. Schweizer 2003, S. 154

⁶⁸⁰ Vgl. Baker 2005

⁶⁸¹ Vgl. Baker 2005

3.8.5 Zwischenfazit: Schriftausrichtung

Die Einigkeit in der Fachliteratur zum Thema „Schriftausrichtung“ wurde bereits hervorgehoben: Linksbündiger Flattersatz ist hier eindeutiger Favorit. Bakers empirische Studie konnte diese Präferenz weder eindeutig belegen noch verwerfen.

3.9 Auszeichnungsvarianten

3.9.1 Auszeichnungsvarianten als typografische Phänomene

Unterstellt man die Theorie als richtig, dass Lesende im Web sich mehrheitlich den Angeboten von Seiten „scannend“ und „skimmend“ nähern – und die Richtigkeit dieser These wird hier angenommen –, so stellt sich die Frage, ob und inwiefern diese Prozesse durch typografische Setzungen unterstützbar, eventuell gar lenkbar sind. Genauer formuliert: Wie lassen sich entscheidende Textpassagen – Vorspanne etwa, Schlüsselwörter oder Zwischenüberschriften – in einer Form visuell über den Gesamttext erheben, die sie rasch identifizierbar gestaltet?

Typografische Hervorhebung lebt prinzipiell vom Kontrast. Zwei Instrumente, die durchaus auch zur Generierung kontrastiver Typografie bereitstehen, wurden bereits ausführlich vorgestellt: Schriftart und Schriftgröße. Schriftarten sind gezielt mischbar und damit Textpassagen und Einzelwörter bereits via Schriftgestalt voneinander abgrenzbar; bedeutsame Passagen können beispielsweise über eine Vergrößerung der Typen optisch als solche qualifiziert werden⁶⁸². Ein weiterer, sehr wichtiger Faktor der Textauszeichnung wird weiter unten noch ausführlich zur Sprache kommen: die Schriftfarbe. Doch es existieren noch weitere typografische Techniken und Varianten der Hervorhebung, von denen drei in diesem Kapitel gesonderter Betrachtung unterworfen werden sollen: Schriftfettung, Kursivstellung und Unterstreichung. Ferner wird noch eingegangen auf zwei weitere Hervorhebungs-Varianten: Versalschreibung und Kapitälchen. Keine tiefer gehende Analyse werde ich dagegen verwenden auf weitere denkbare Textauszeichnungs-Methoden, die vor allem der Digitalsatz ermöglicht hat und die eindeutig nicht zu den klassischen typografischen Auszeichnungs-Varianten zu zählen sind: beispielsweise die, virtuelle Schatten hinter Lettern zu setzen oder die Schrift zu „konturieren“; diese Methoden gelten ohnehin als nicht sehr seriös – und umsetzbar in HTML und CSS sind sie gleichfalls nicht.

Es sei betont, dass all diese Auszeichnungsmethoden durchaus sehr unterschiedlicher „physischer“ Natur sind. Natürlich ist beispielsweise eine *Times* in Kursivsatz immer noch eine *Times*; die Diskussion von Schriftschnitt-Varianten wäre also auch im Kapitel „Schriftart“ durchaus sinnvoll untergebracht gewesen. Eine unterstrichene *Times* aber ist eben nicht im Schnitt variiert, sondern um ein visuelles Element erweitert – und Versalsatz ist eher als orthografi-

⁶⁸² Vgl. Lazar 2005, S. 149

sche Auszeichnung zu verstehen denn als originär typografische. Gemeinsam jedoch ist diesen Elementen die Funktion: das visuelle Herausstellen von Passagen. Dieser gemeinsame Charakter rechtfertigt ihre Behandlung in diesem gesonderten Kapitel.

3.9.2 Manipulation der Auszeichnung in HTML-basierten Browsern

Alle der in diesem Teilkapitel interessierenden Auszeichnungsvarianten sind mithilfe von CSS für Webseiten definierbar. Kursive Passagen werden in CSS dabei über den Befehl „font-style:italic“ erzeugt. Diese Funktion wird von allen bekannten Browsern unterstützt. Unterstrichenen Text erzeugt der Befehl „text-decoration“ mit der Attributszuweisung „underline“. Weitere Attribute für „text-decoration“ sind übrigens „overline“, „line-through“ (durchgestrichen) und das unsägliche „blink“ (blinkender Text) – diese Spezialausprägungen sollen hier allerdings nicht weiter behandelt werden, da sie augenscheinlich kaum genutzt werden im zeitgenössischen Web-Journalismus – und sich ihre Verwendung auch keineswegs aufdrängt als Alternative.

Die Schrift-Fettung wird über die Befehlsbezeichnung „font-weight“ generiert, die dafür das Attribut „bold“ (eben: fett) vorsieht. Auch dieser Befehl wird von allen wichtigen Browsern korrekt visualisiert.

Bei der Darstellung von auf Basis von CSS angeforderten gefetteten oder kursiven Typen kommen in Browsern dabei zwei Varianten der Generierung zum Einsatz. Die erste: Es wird ermittelt, ob auf dem Rechner des Nutzers ein eigener, speziell gezeichneter Zeichensatz existiert für den gewünschten Schnitt, und dieser wird verwendet zur typografischen Darstellung. Gerade höherwertige Schriften zeichnen sich dadurch aus, dass vor allem ihre kursiven Varianten mehr sind als bloße „Kippungen“ des Ursprungsschnitts nach rechts – kursive Buchstaben sind in wertigen Fonts vielmehr eigene gestalterische Schöpfungen mit hoch individuellen, vom Normalschnitt merklich differierenden Charakteristika. Abbildung 3.24 zeigt das am Beispiel der *Times*: Allein, dass das „f“ im Kursivschnitt plötzlich eine Unterlänge entfaltet und das „a“ vom offenen in den geschlossenen Zustand wechselt, beweist, dass hier mehr geschieht im Übergang von „normal“ zu „italic“, als dass die Letterneigung variiert wird.

Fall zwei freilich kann gleichfalls eintreten: der nämlich, dass ein separater Zeichensatz auf dem Rechner nicht vorliegt im via CSS angeforderten Schnitt. Die *Arial Black* beispielsweise, eine extrafette Variante der Schriftart *Arial*, gibt es als Schriftdatei einzig und allein im Originalschnitt. Definiert nun jedoch der Gestalter einer Webseite eben diese Schriftart dennoch



in kursivem Schnitt (eben mittels des Befehls „font-style:italic“), generieren die meisten Browser sogenannte „unechte Schnitte“. Im beschriebenen Falle würde also die *Arial Black* in ihrer Originalform schlicht um eine bestimmte Gradzahl nach rechts „geneigt“; es entsteht eine „Pseudo-Kursive“. Ja, selbst eine Fettung der ohnehin schon massigen *Arial Black* ist im Web möglich („font-weight:bold“): die daraus generierte „Pseudo-Bold“ ist dann aber nicht mehr als eine um einige Pixel „Veckte“ Original-*Arial Black*, ein Wechselbalg. Dass diese Manipulationen der Attraktivität des Schriftbildes meistens alles andere als dienlich sind, liegt auf der Hand.

In diesem Sinne ist es auch nicht weiter bedauerlich, dass die Attribute „bolder“ und „lighter“, die der CSS-Befehlssatz prinzipiell ebenfalls vorhält als Detailabstufungen des Normalbuchstabens, bislang von keinem der marktbeherrschenden Browser korrekt interpretiert werden; es kämen meistens wohl nur suboptimale, unechte Schnitte heraus. Gleiches gilt für die ebenfalls grundsätzlich mögliche Attributisierung der CSS-Spezifikation „font-weight“ in numerischen Ausprägungen von „100“ bis „900“. Der niedrigste Wert bezeichnet dabei, zumindest theoretisch, eine sehr dünnbalkige Zeichnung der Schriftart, „900“ eine besonders fette. Nur sehr wenige Schriftarten überhaupt existieren allerdings in einer derart feinen Fetten-Gliederung; in kleineren Schriftgraden würde sich eine derartig ausdifferenzierte Formgenese im groben Pixel-Muster zeitgenössischer Bildschirme ohnehin kaum niederschlagen – in größeren Schnitten dagegen würden absehbar wiederum nur unechte Schnittvarianten auf dem Bildschirm erscheinen.

Einen Text in Großbuchstaben zu erzeugen, ist prinzipiell natürlich – und nicht nur im Web – auch ohne originär typografische Manipulation möglich: Man tippt eben die gewünschte Passage einfach in Versalien. Freilich generieren sich viele Webseiten heute aus teilweise von sehr vielen unterschiedlichen, auch verlags- und redaktionsfremden Autoren bestückten Datenbanken; um deren Einträge nicht einzeln auf korrekten Versalsatz durchsuchen zu müssen, lohnt es sich also, bestimmte Textbereiche schlicht mit dem CSS-Befehl „text-transform:uppercase“ zu versehen. Kapitalchen schließlich sind erzeugbar über „font-variant:small-caps“.

Im übrigen sind die eben beschriebenen Auszeichnungsvarianten im Prinzip auch wechselseitig kombinierbar – abgesehen natürlich von Versalsatz, der nicht gleichzeitig Kapitalchen aufweisen kann.

3.9.3 Empfehlungen und Vorgaben aus der Literatur

Schweizer gibt sich als Anhänger von typografischer Auszeichnung: „Die Verwendung unterschiedlicher Schriftschnitte verbessert die Lesbarkeit deutlich, vor allem im Web, wo Besu-

cher Texte in der Mehrzahl ‚querlesen‘⁶⁸³. Schäffer favorisiert dabei eindeutig die Textfettung: „Wirkt im normalen Schriftsatz der häufige Gebrauch von Fett schnell aufdringlich und wenig schön, bietet sich diese Auszeichnung im Web sehr gut an, da fette Texte nicht so stark hervortreten wie bei gedruckten Texten“⁶⁸⁴.

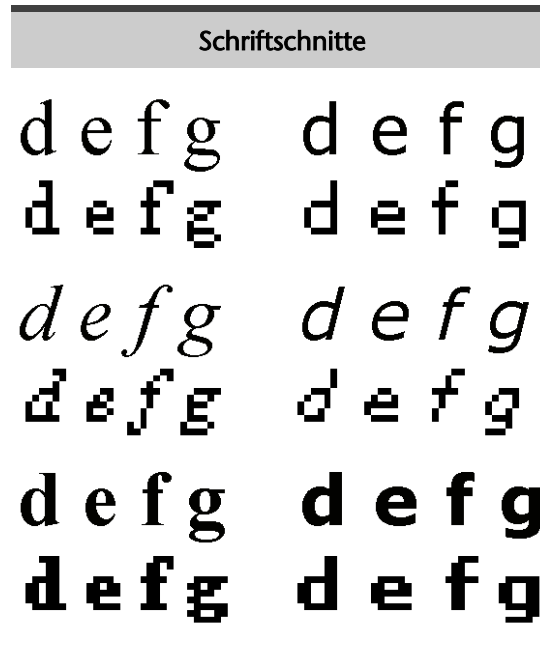
Lankau mahnt allerdings, Schriftschnitte zur Auszeichnung sparsam einzusetzen:

„Ein ganzer fetter Absatz schreit zwar laut. Aber warum sollten Sie jemanden anschreien“⁶⁸⁵. Auch George rät zu maßvollem Einsatz von fetten Lettern; zu viele davon ließen Texte rasch „anzeigenhaft“ wirken⁶⁸⁶. Peck führt ein weiteres Argument gegen ein Übermaß an „Fett“ im Text an: Allzu viel „Hervorgehobenes“ sei ab einem gewissen Punkt schlichtweg nicht mehr als solches identifizierbar⁶⁸⁷. Auf diesen Aspekt weisen auch Lynch und Horton hin⁶⁸⁸.

Eine auffällige Skepsis schlägt dem Kursivsatz in der Literatur entgegen – und dies nicht ganz zu Unrecht, wie ich meine. Denn Abbildung 3.25 demonstriert

anschaulich, dass das bildschirmimmanente Problem der kleinen Auflösung bei kursiv gesetzten Buchstaben noch augenfälligere visuelle Korruptionen der Gestaltdarstellung zeitigt: Kursive Schriften weisen meistens ein noch höheres Maß an Rundungen und geschwungenen Ausläufen auf als ihre verwandten Normalschnitte; die Gefahr der „Verpixelung“, der „Treppenbildung“ ist in dieser Schnittvariante also potenziell noch ausgeprägter⁶⁸⁹.

Schweizer mahnt daher, zumindest in kleineren Schriftgrößen auf Kursivsatz zu verzichten⁶⁹⁰, worin er mit Peck übereinstimmt⁶⁹¹. Auch Balzert meint, Kursive seien am Bildschirm schlechter lesbar als Normalschnitte⁶⁹². Lankau vertritt dieselbe Ansicht⁶⁹³, genau wie Böhringer et



Kleinbuchstaben in *Times* (links) und *Verdana* (rechts) in Normal-, Kursiv- und Fettsatz, jeweils hoch einerseits und bildschirmmännlich andererseits aufgerastert.

ABBILDUNG 3.25

⁶⁸³ Schweizer 2003, S. 152

⁶⁸⁴ Schäffer 2001, S. 187

⁶⁸⁵ Lankau 2000, S. 368

⁶⁸⁶ Vgl. George 1995, S. 63

⁶⁸⁷ Vgl. Peck 2000, S. 163

⁶⁸⁸ Vgl. Lynch/Horton 1999, S. 90

⁶⁸⁹ Hellbusch 2005, S. 93

⁶⁹⁰ Schweizer 2003, S. 152

⁶⁹¹ Vgl. Peck 2000, S. 163

⁶⁹² Balzert 2004, S. 188

⁶⁹³ Vgl. Lankau 2000, S. 366

al., die schreiben, kursive Schriften wirkten „auf Monitoren stufig“; die Lesbarkeit sei daher beeinträchtigt⁶⁹⁴. Es sei allerdings an dieser Stelle daran erinnert, dass die vermeintlich mangelnde Darstellungsqualität viele Autoren auch zum Urteil verleitete, serifentragende Schriften seien „schlechter lesbar“ am Monitor als Groteske – und sich diese Herleitung der Lesbarkeitsqualität aus der Pixelaufbereitungs-Qualität empirisch in den meisten Studien nicht als haltbar erwiesen hat. Lynch und Horton schließlich halten Kursivsatz – sehr sparsam eingesetzt – für vertretbar⁶⁹⁵.

Die Auszeichnungsvariante der Unterstreichung halten alle Autoren, die ich mit Auslassungen zu diesem Aspekt identifizierte, für ungeeignet auf Webseiten⁶⁹⁶. Und zwar berechtigterweise, wie ich finde: Denn unterstrichene Wörter, Sätze, Passagen signalisierten bereits im allerersten Browser *WorldWideWeb* von Tim Berners-Lee, dass sie Hyperlinks seien, also „anklickbar“. Auf vielen modernen Webseiten ist das zwar nicht mehr der Fall (CSS erlaubt auch die „Deaktivierung“ des Unterstrichs), als gleichwohl lebendige Konvention jedoch wird die Assoziation der Unterstreichung mit einem Link wohl noch einige Jahre fortbestehen im Web. Unterstrichene Textpassagen heben nicht hervor, wenn sie unverlinkt bleiben – sie verstören, irritieren, ja verärgern stattdessen potenziell Betrachtende⁶⁹⁷.

Lamprecht hält Versalsatz, also Wörter oder Passagen, die durchgehend in Großbuchstaben gesetzt sind, für ein weiteres geeignetes Mittel der Auszeichnung⁶⁹⁸. Balzert ist nicht dieser Meinung; sie hält darüber hinaus auch Kapitälchensatz für eine ungeeignete Herausstellungsmethodik⁶⁹⁹.

Wiederum Lamprecht verweist schließlich darauf, dass eine einzige Auszeichnungsmethode in aller Regel ausreicht, um Textbereiche als besondere kenntlich zu machen. Eine Kombination von Methodiken sei in aller Regel nicht sinnvoll⁷⁰⁰ – ein erstaunlich selten anzutreffender Hinweis, wiewohl gerade der schauerliche Dreiklang „fett-kursiv-unterstrichen“ zumindest auf semiprofessionellen Webseiten nach wie vor in bedauerlicher Blüte steht.

Kommer und Mersin machen schließlich noch auf eine Interaktion zwischen den typografischen Faktoren „Schriftschnitt“ und „Zeilenabstand“ aufmerksam: „Breite oder magere Schriften benötigen einen weiten weiten Zeilenabstand; schmale und fette Schriften können enger gesetzt werden. (...) Je leichter die Schrift, (...) desto mehr Zeilenzwischenraum ist für eine gute Lesbarkeit und ein ausgewogenes Schriftbild erforderlich“⁷⁰¹.

⁶⁹⁴ Vgl. Böhringer et al. 2003, S. 131

⁶⁹⁵ Vgl. Lynch/Horton 1999, S. 90

⁶⁹⁶ Vgl. Schäffer 2001, S. 187; vgl. ferner Balzert 2004, S. 187; vgl. ferner Peck 2000, S. 163

⁶⁹⁷ Vgl. Lynch/Horton 1999, S. 90

⁶⁹⁸ Vgl. Lamprecht 2002, S. 39

⁶⁹⁹ Vgl. Balzert 2004, S. 187

⁷⁰⁰ Lamprecht 2002, S. 39

⁷⁰¹ Kommer/Mersin 2002, S. 175

3.9.4 Empirische Erkenntnisse zur Textauszeichnung am Bildschirm

3.9.4.1 Kursivsatz: *Times*, *Arial* und *Courier*

In der bereits erwähnten Studie von Hill und Scharff verglichen die Forscher auch die Reaktionszeiten der Probanden unter den Bedingungen von Normal- und Kursivsatz. Die verglichenen Schriftarten *Times*, *Arial* und *Courier* wurden in beiden Schnittvarianten aufgeboten⁷⁰². Dabei zeigte sich, dass die Reaktionszeiten auf die kursiv gesetzten Texte durchgehend etwa 9 Prozent (und damit statistisch signifikant) schlechter ausfielen als die für die normal gesetzten. Allerdings war auch zu beobachten, dass sich dieser Effekt fast ausschließlich im Positivsatz (dunkle Schrift auf hellem Grund) zeigte; im Negativsatz dagegen fielen die Kursivschriften sogar teilweise etwas besser aus als ihre normal geschnittenen Pendanten. Zum Thema Schriftfarbe jedoch später mehr.

3.9.4.2 Kursivsatz: *Verdana*

Boyarski et al. untersuchten in ihrer bereits angesprochenen Studie aus dem Jahre 1998 auch die *Verdana* im normalen einerseits und im kursiven Schnitt andererseits in Bezug auf Lesegeschwindigkeit und Verständnisleistung⁷⁰³. Sie fanden dabei einen fast vollständigen Gleichstand der Ergebnisse in beiden Konstellationen und Messkriterien; signifikante Unterschiede zwischen kursivem und Normalsatz stellten sie gar nicht fest.

3.9.4.3 Textfettung

Dyson und Gregory überprüften 2002 mit 30 Probanden die Wirkung von Textfettungen und farblicher Textauszeichnung am Bildschirm auf das Textverständnis⁷⁰⁴. Diese Studie soll im Rahmen dieses Kapitels behandelt werden, obgleich der Faktor „Schriftfarbe“ erst weiter unten zur Detail-Diskussion ansteht.

Jeder Proband erhielt im Experiment fünf Texte zu lesen, gesetzt in einer 10-Punkt-*Verdana* mit 12 Punkt Zeilenabstand. Einer der Texte enthielt keinerlei typografische Hervorhebungen, je zwei waren mit Passagen ausgestattet, die entweder durch Rotfärbung der Lettern oder Fettdruck ausgezeichnet waren. Innerhalb dieser jeweils zwei Texte in identischer typografischer Auszeichnung differenzierten Dyson und Gregory nochmals, indem sie in jeweils einem Text ganze Sätze typografisch hervorhoben, im jeweils anderen lediglich einzelne bedeutsame Wörter. Typografisch ausgezeichnet waren zum Teil Sätze oder Wörter, die für die Beantwortung von inhaltlichen Fragen zu den gelesenen Texten relevant waren, zum Teil jedoch nur für die Befragung irrelevante inhaltliche Details; diese Befragung schloss sich an je-

⁷⁰² Vgl. Hill/Scharff 1997

⁷⁰³ Vgl. Boyarski et al. 1998. S. 87ff.

⁷⁰⁴ Vgl. Dyson/Gregory 2002, S. 326ff.

den Leseprozess an. Auf diese Tatsache der absehbaren Relevanz der ausgezeichneten Textpassagen für die Inhalts-Abfrage wurde wiederum die eine Hälfte der Teilnehmenden vor Beginn des Experiments hingewiesen, die andere Hälfte begann den Leseprozess ohne derlei Instruktion.

Das Ergebnis kann hier sehr kurz subsummiert werden: Weder bei den Lesezeiten noch beim Textverständnis fanden die Forscherinnen insgesamt einen merklichen Einfluss auch nur einer der Hervorhebungsmethodiken. Im Detail freilich ergaben sich dennoch einige interessante Erkenntnisse: Die Teilnehmer, die darüber informiert waren, dass die hervorgehobenen Textstellen sich teilweise auf die anschließenden Fragen beziehen würden, erzielten leicht bessere Ergebnisse als die Vergleichsgruppe. Auch stellte sich heraus, dass die Hervorhebung von Einzel-Wörtern bessere Erinnerungseffekte erbrachte als die Auszeichnung ganzer Sätze.

In der Diskussion ihrer Ergebnisse gestehen Dyson und Gregory jedoch zu, dass sie selbst ihre Ergebnisse zumindest für überprüfenswert halten. Unter anderem fanden die Autorinnen nämlich durchaus irritierende Zusammenhänge zwischen Befragungsinhalten und Hervorhebungsmethoden, Korrelationen also, die bei näherer Untersuchung eindeutig konfundierende Variablen identifizierten, die die Reliabilität der Erhebung durchaus insgesamt infrage stellen, wie auch die Autoren hervorheben: „Unsere Ergebnisse legen nahe, dass weitere Untersuchungen zum Gebrauch typografischer Hervorhebungen nützlich wären“⁷⁰⁵.

3.9.5 Zwischenfazit: Textauszeichnungen

Ein sparsamer Einsatz von Fettsatz wird in der Literatur empfohlen, Kursivsatz erfährt eher eine Ablehnung – so lassen sich die Empfehlungen aus der Literatur kurz zusammenfassen. Zumindest die Erkenntnisse von Hill und Scharff legen die Vermutung nahe, dass von Kursivsatz im Web tatsächlich eher abgesehen werden sollte. Allerdings fanden Boyarski et al., wie beschrieben, keine Differenzen in der Lesbarkeit der *Verdana* zwischen deren Normal- und Kursivschnitt. Das Experiment von Dyson und Kipping schließlich erbrachte keinen weiterreichenden Erkenntnisgewinn; mit gutem Willen wäre der Studie höchstens entnehmbar, dass Fettsatz die Lesbarkeit scheinbar nicht substantiell reduziert.

3.10 Laufweite

3.10.1 Die Laufweite als typografisches Phänomen

Als Laufweite wird der horizontale Abstand zwischen jeweils zwei Zeichen innerhalb eines Textes verstanden. Jede Schriftart besitzt eine vom Hersteller definierte, quasi originäre Lauf-

⁷⁰⁵ Vgl. Dyson/Gregory 2002, S. 326ff.

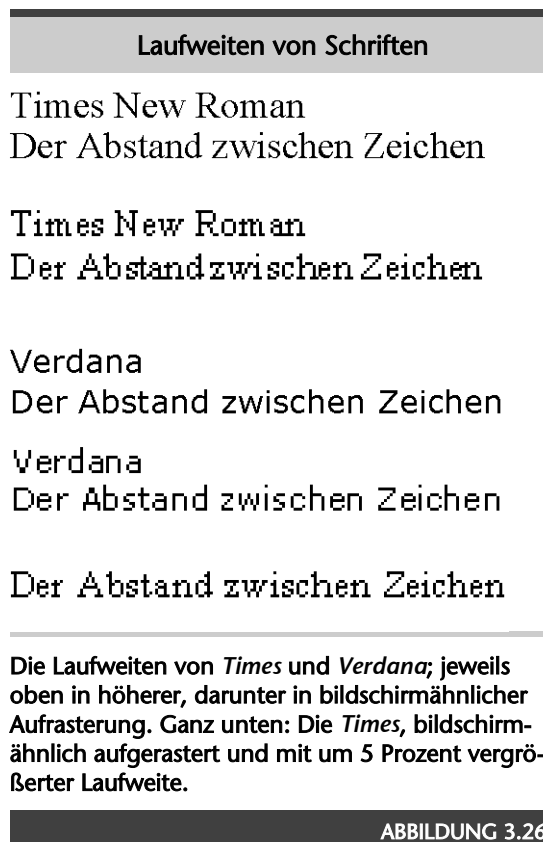
weite, die so genannte „Standard-Laufweite“. Zusätzlich ist die Laufweite in gängigen Satzprogrammen jedoch manipulierbar: Die Laufweite kann verkleinert („Engsatz“) oder verbreitert werden („Textsperrung“).

Für die Darstellung von Schriften am Bildschirm ist bereits die Standard-Laufweite von Bedeutung. Die bereits beschriebene „Verpixelung“ von Schriften am Bildschirm nimmt nämlich nicht nur Einfluss auf die Darstellung der einzelnen Lettern, sondern auch auf die horizontalen Distanzen der Einzelzeichen. Wo nämlich im höchauflösenden Druck der Abstand zwischen zwei Buchstaben in Bruchteilen von Millimetern fein austariert werden kann, steht am Monitor wiederum nur die sperrige „Einheit“ Pixel bereit.

Dies führt vor allem bei tendenziell mit enger Standard-Laufweite geschnittenen Schriften wie der *Times New Roman* dazu, dass – wie in Abbildung 3.26 erkennbar – in kleiner Auflösung manche Zeichenab-

stände wie der zwischen dem „i“ und dem „s“ im Wort „zwischen“ übermäßig groß (nämlich exakt ein Pixel breit) ausfallen, während andere Letter-Paare wie das „st“ in „Abstand“ und das „en“ in „Zeichen“ ohne Puffer aneinanderstoßen und damit künstliche Ligaturen (also Letterverschmelzungen) und sehr schwere Pixel-„Cluster“ bilden⁷⁰⁶. Das allgemeine Schriftbild der *Times*, in höherer Auflösung deutlich heller als beispielsweise das der *Verdana*, gerät dadurch zu einem schwarzlastigen, sehr „klumpigen“ Block.

Doch die Ligaturenbildung ist in der Konsequenz nicht nur aus ästhetischen Erwägungen heraus problematisch. Die Verschmelzung von Zeichen (eine „typografische Todsünde“, wie Böhringer et al. bemerken⁷⁰⁷) schafft potenziell neue, unbekannte Muster – was die Lesegeschwindigkeit durchaus beeinträchtigen dürfte: Schließlich erschwert dies absehbar den Identifikationsprozess der (nach der Merkmals-Theorie) relevanten Detektoren des jeweiligen Einzelzeichens. Diese Gefahr ist bei der *Verdana* kleiner, wie ebenfalls Abbildung 3.26 zeigt: Nicht nur das Gesamt-Gewicht des Schriftbildes bleibt in dieser Schriftart auch bei verkleinerter Auflösung recht konstant, praktisch alle Zeichen weisen zudem einen horizontalen Ab-



⁷⁰⁶ Vgl. Hagge 1994, S. 114

⁷⁰⁷ Vgl. Böhringer et al. 2003, S. 185

stand von mindestens einem Pixel auf und bleiben damit nicht nur als Solitär-Lettern erkennbar, sondern auch als potenziell bekannte Zeichen-Ketten.

Das Beispiel in der untersten Zeile von Abbildung 3.26, in dem nachträglich der Buchstabenabstand auf durchgehend einen Pixel gesetzt wurde, zeigt, dass eine variierte, leicht vergrößerte Laufweite auch der *Times* sehr zugute kommen kann.

3.10.2 Manipulation der Laufweite auf HTML-basierten Webseiten

Eine Manipulation der Laufweite von Texten ist über die Definition spezifischer Style Sheets prinzipiell möglich. Das Grundproblem löst diese Funktion allerdings nicht. Denn eine Vergrößerung der Laufweite um beispielsweise 1 Pixel wird auf den jeweiligen gesamten Textblock angewendet; dies aber führt dazu, dass die ohnehin oft überdimensionierten horizontalen Abstände in problematischen Schriften wie der *Times* zusätzlich vergrößert werden.

3.10.3 Empfehlungen und Vorgaben aus der Literatur

Allgemein wird der Laufweite in der Literatur eher wenig Aufmerksamkeit gewidmet. Gorbach empfiehlt für Textschrift auf dem Bildschirm „eine etwas weitere Laufweite als auf dem Papier“⁷⁰⁸, spezifiziert diese Angabe allerdings nicht tiefer gehend. Laut Schweizer erreicht man bereits „mit sehr kleinen Änderungen eine deutliche Verbesserung der Lesbarkeit“⁷⁰⁹. Konkrete Werte nennt auch Schweizer jedoch nicht. Als einziger empfiehlt Bollwage konkrete Laufweiten-Maße fürs Web: „Dem normalen Buchstabenabstand sollten je nach Schriftart mindestens 5 bis 10% eines Gevierts hinzugefügt werden. Bei kleineren Graden mehr und bei größeren weniger“⁷¹⁰.

3.10.4 Empirische Studien zur Laufweite von Bildschirmschriften

Zur Laufweite von Schriften am Bildschirm liegen nach meinen Recherchen keine empirischen Studien vor.

3.11 Schriftfarbe und Farbgrund

3.11.1 Schriftfarbe und Farbgrund als typografische Phänomene

Das Feld ist schier unüberschaubar. Die Kategorisierung und wissenschaftliche Erklärung des Phänomens Farbe, die zahllosen potenziellen Funktionen von Farbe, die psychologische (Heil-)Wirkung und die neurobiologische Rezeption ist und war schon Gegenstand zahlloser

⁷⁰⁸ Gorbach 2001, S. 255

⁷⁰⁹ Schweizer 2003, S. 153

⁷¹⁰ Bollwage 2001, S. 102

Forschungsvorhaben, theoretischer Erörterungen und praktischer Handreichungen⁷¹¹. Bedarf es der Erwähnung, dass menschliche Seherfahrung, also auch die Wahrnehmung von Buchstaben und anderen Zeichen, nicht nur den Gesetzen der Gestalttheorie folgend ohne Farbe schlicht ausgeschlossen ist? Bartel widmete gar ein ganzes Werk nur dem Thema „Farbe im Webdesign“⁷¹², Erben verwandte zwei Drittel seiner Darlegungen nur auf dieses Feld⁷¹³.

Trotz dieser also potenziellen Weite des Themas soll in diesem Kapitel fortgeführt werden, was auch bislang galt: die strikte Beschränkung der Diskussion des Aspekts auf mesotypografische Dimensionen sowie browserspezifische Besonderheiten und Probleme.

3.11.2 Manipulation von Farbe in HTML-basierten Browsern

Zunächst ist ein wichtiger Aspekt hervorzuheben: Farbe im Web ist entstofflicht, pigmentfrei und damit im Grunde: umsonst zu haben⁷¹⁴. Das Hinzunehmen beispielsweise einer Schmuckfarbe in ein ansonsten in Grautönen gehaltenes Dokument ist damit im Web – anders als im Papierdruck – eine nur mehr gestalterische, keine ökonomische Grundsatzentscheidung mehr. Dass gerade viele Hobby-Webdesigner diese Erkenntnis zum Anlass nehmen, ihre Webseiten mit wirklich allen verfügbaren Tönen des Regenbogens zu verunstalten, sei nur kurz angedeutet, aber nicht weiter verfolgt – denn dies ist in erster Linie ein makrotypografisches Problem.

Web-Farben sind Lichtfarben, direkte Farben; erzeugt also im sogenannten additiven Farbmodus mit den Grundfarben Rot, Grün und Blau („RGB-Modus“ oder auch „RGB-Farbraum“)⁷¹⁵. Was qua dieser Fachtermini ein recht komplexes Vorgehen vermuten ließe, ist im Web vergleichsweise wirklich anwenderfreundlich konzipiert. Denn grundsätzlich sind der Farbgestaltung in Browsern nicht nur praktisch kaum Grenzen gesetzt, die HTML-/CSS-basierte Farbdefinition ist zudem relativ simpel konstruiert. Der CSS-Befehlssatz kennt die Spezifizierung „color“ (in eben dieser US-englischen Ausformung ohne „u“ als vorletzter Letter), deren Ausprägung entweder in Form einer vordefinierten Verbalbezeichnung beschreibbar ist (zum Beispiel als „green“ oder „light blue“, aber auch als „aquamarine“ oder gar „burnt orange“) oder in Form eines sechsstelligen Hexadezimal-Codes (die Farbe „Rot“ wird hier beispielsweise verschlüsselt als „#ff0000“). Beide Varianten werden von allen gängigen Browsern interpretiert⁷¹⁶.

⁷¹¹ Vgl. z. B. Bloomer 1976, S. 120ff.; vgl. ferner Heller 2002; vgl. ferner Braun 1987, S. 27ff.; vgl. ferner Zuffo 1993, S. 101ff.; vgl. ferner Zimbardo/Gerrig 2004, S. 126ff.; vgl. ferner Böhringer et al. 2003, S. 409ff.

⁷¹² Vgl. Bartel 2003

⁷¹³ Vgl. Erben 1998

⁷¹⁴ Vgl. Bollwage 2001, S. 103

⁷¹⁵ Vgl. Weinman 2000, S. 131

⁷¹⁶ Vgl. McKelvey 1999, S. 42ff.

Dennoch sind die Kontrollmöglichkeiten von Gestaltern über Farben im Web fast noch eingeschränkter, als dies in Bezug auf Schriftarten der Fall ist. „Denn jeder Bildschirm hat eine andere Farbtemperatur, einen anderen Kontrast, eine andere Helligkeit und ein anderes Umgebungslicht“⁷¹⁷. Die Folge: Eine vermeintlich fix definierte Farbe kann an unterschiedlichen Monitoren frappierend differierende Farbeindrücke zeitigen – ein Phänomen, das aus dem Fernsehfachhandel bekannt ist, wo nebeneinander drapierte TV-Geräte ebenfalls gerne völlig unterschiedliche Farbklimata aus identischen Sendeangeboten erzeugen. Die Geräte sind unterschiedlich „kalibriert“, wie es im Expertenterminus heißt; und bei Computerbildschirmen ist es nicht viel anders. Schlimmstenfalls kippt dann an manchen Monitoren ein definiertes Dunkelrot gefährlich in den erdigen Muff eines Braunähnlichen.

Ein weiteres Problem ergibt sich aus der Eigenart des Monitors als „Selbststrahler“. Richtig ist zwar, dass Farben auf Bildschirmen „weitaus leuchtender und lebendiger (wirken) als im Vierfarbdruck, der weit weniger Farben darstellen kann“⁷¹⁸. Gerade im Schriftsatz am Monitor kann der „strahlende“ Charakter des Mediums aber auch unerwünschte Nebenwirkungen auslösen – insbesondere, wenn der Hintergrund eines dunklen Textblocks sehr hell und damit eben lichtinsensiv definiert ist. Hier besteht die Gefahr des „Überstrahlens“⁷¹⁹: Die Konturen der Vordergrundobjekte (im Schriftsatz also: der Buchstaben) werden visuell korumpiert. Folge: die Lettern wirken „dünner“⁷²⁰ und damit als Gesamt auch kleiner⁷²¹. Hinzu tritt, dass ein hoher Anteil heller Farben auf dem Bildschirm allgemein dafür sorgt, dass sehr intensives Licht auf das Auge der Betrachtenden einwirkt, zudem das Bildschirmflimmern mit der Strahlungsintensität wächst – und im Zweifel damit die Augen ermüdet⁷²².

3.11.3 Empfehlungen und Vorgaben aus der Literatur

Insbesondere das Problem des Überstrahlens spielt eine hervorgehobene Rolle in der Literatur. Zwar betonen alle Autoren weitgehend unisono, dass Farbgestaltung im Web prinzipiell kontrastreich ausfallen sollte⁷²³; Schrift und Farbgrund sollten also farblich eindeutig separiert sein. Nielsen findet sich jedoch in einer faktischen Alleinstellung in seiner Ableitung, der augenfälligste Kontrast sei der empfehlenswerteste, „damit die Schrift so lesbar ist wie möglich“⁷²⁴: Schwarzer Text auf weißem Grund sei daher allgemein „der einfachste und normalerweise der beste“⁷²⁵. Höchstens dunkelblau oder dunkelgrau seien als Schriftfarben alternativ akzeptabel. Nielsen wiederholt diese Forderung später ausdrücklich⁷²⁶.

⁷¹⁷ Thissen 2003, S. 94

⁷¹⁸ Pring 2000, S. 98

⁷¹⁹ Vgl. Köhler 2002, S. 272

⁷²⁰ Vgl. Bollwage 2001, S. 103

⁷²¹ Neutzling 2002, S. 20

⁷²² Hellbusch 2005, S. 82

⁷²³ Vgl. Lankau 2000, S. 368; vgl. ferner Stocksmeier 2002, S. 114; vgl. ferner Neutzling 2002, S. 20

⁷²⁴ Nielsen/Tahir 2002, Übersetzung M.L.

⁷²⁵ Nielsen/Tahir 2002, S. 51. Übersetzung M.L.

⁷²⁶ Vgl. Nielsen 2000a-, S. 126f.

Dies ändert freilich nichts daran, dass Nielsen mit seinem Ratschlag weitgehend allein bleibt. Einen schwachen Verbündete findet er lediglich in Lazar, der „dunklen Text (wie zum Beispiel schwarz) auf hellem Grund“ propagiert⁷²⁷. Ebenso in der Minderheit sind freilich auch Autoren wie Bürgel und Neumann, die Niensens Apodiktum schlicht umkehren und im Grunde ein Zurück propagieren zu den satt-dunklen Hintergrundfarben, der „Negativpolarität“⁷²⁸ also der 1980er Jahre: „Sehr gut geeignet sind Schriften in schwachprozentigem Schwarz auf schwarzem oder dunkelgrauem Grund“⁷²⁹. Auch Erben argumentiert in diese Richtung, denn: „Ist der Bildschirmhintergrund schwarz, wird das Flimmern optimal reduziert“⁷³⁰, Pring weist zudem darauf hin, dass zumindest serifenlose Schriften in Invers-Darstellung (also weiß auf schwarz) am Bildschirm durchaus erheblich an visueller Qualität und Schärfe gewinnen können⁷³¹. Allerdings macht Schäffer auf ein wichtiges, wenn auch rein praktisches Problem heller Schrift im Browser aufmerksam: „Der Browser berücksichtigt (...) den Hintergrund beim Ausdruck nicht und helle Buchstaben würden dann auf dem weißen Papier kaum zu lesen sein“⁷³².

In der Gesamtschau jedoch scheint die Überzeugung vorzuherrschen „dass dunkle Schrift auf hellem Hintergrund grundsätzlich besser zu lesen ist als umgekehrt“⁷³³. Darüberhinaus überwiegt in der Literatur erkennbar die Ansicht, dass Farbkontraste am Monitor allgemein schwächer ausfallen sollten als auf Papier – vermieden werden sollten Kombinationen „wie Schwarz-Weiß, Bunt-Unbunt oder der Komplementärkontrast (Ergänzungsfarben wie beispielsweise Rot-Grün oder Gelb-Violett, M.L.)“⁷³⁴. Insbesondere reinweißer Hintergrund sollte alternativen Tönen weichen, meint Hellbusch: Das kann durch einen leichten Graustich, aber auch mit jeder anderen Farbe erzielt werden“⁷³⁵. Auch Köhler empfiehlt eine „helle Hintergrundfarbe (nicht Weiß)“ für reinschwarzen Text⁷³⁶, meint aber auch, dass die Stellschraube gleichzeitig auch an der anderen Seite angesetzt werden sollte: den Buchstaben im Vordergrund. Diese sollten eher in einem „dunklen Grau“ denn Reinschwarz erscheinen. Auch Thissen hält „mittlere Kontraste“ für die besten, herstellbar „beispielsweise durch den Einsatz von ungesättigten Farben und dezenten Pastelltönen“⁷³⁷. Balzert ist derselben Ansicht: „Wählen Sie dunkle oder gesättigte Farben für die Schrift und Pastellfarben für den Hintergrund“⁷³⁸. Für Lankau ist sogar ein mutigerer Griff in die virtuelle Farbkiste zumindest in Bezug auf die Buchstaben denkbar; der „Einsatz von farbigen Schriften und dezenten Hinter-

⁷²⁷ Vgl. Lazar 2005, S. 148

⁷²⁸ Vgl. Ziefle 2002, S. 59

⁷²⁹ Bürgel/Neumann 2001, S. 111

⁷³⁰ Erben 1998, S. 24

⁷³¹ Vgl. Pring 2000, S. 15

⁷³² Schäffer 2001, S. 190

⁷³³ Kesper et al. 2004, S. 681

⁷³⁴ Neutzling 2002, S. 20

⁷³⁵ Hellbusch 2005, S. 82

⁷³⁶ Vgl. Köhler 2002, S. 272

⁷³⁷ Thissen 2003, S. 100

⁷³⁸ Balzert 2004, S. 188

gründen (kann) durchaus eine Alternative sein⁷³⁹. Die bundesdeutsche Bildschirmarbeitsverordnung von 1996 schließlich schreibt vor, dass Texte und Grafiken am Bildschirm auch dann erkennbar und verständlich sein müssen, wenn sie ohne Farbe betrachtet werden⁷⁴⁰ – dies dürfte weniger eine Reminiszenz sein an den guten alten Schwarzweiß-Bildschirm, der heute schwerer zu finden sein dürfte als ein Schwarzweiß-Fernseher, sondern trägt merklich der Tatsache Rechnung, dass eine durchaus qualifizierte Minderheit von Menschen farbenblind oder zumindest farbfahlsichtig ist⁷⁴¹. Letztlich geht diese Vorschrift aber völlig konform mit den oben zitierten Aussagen: Schriftfarbe und Schriftgrund müssen ausreichend kontrastieren. Von schwarzer Schrift auf weißem Grund spricht auch die Verordnung nicht; dunkelgraue Schrift auf hellgrauem Grund entfaltet auch im Schwarzweiß-Modus definitiv ausreichend Kontrast, um die Lettern optisch klar vom Hintergrund abzugrenzen.

Analysiert man nur kurz und überblicksweise die Gesamtgestaltung der wohl meisten journalistisch orientierten Seiten, die heute im Web des Aufrufs harren, so ist allerdings zu konstatieren, dass den zitierten Empfehlungen der Literatur praktisch nirgendwo Folge geleistet wird. Zumindest Fließtexte sind bei fast allen Anbietern Vollscharz auf reinweißem Grund gesetzt. Sollte der Grund darin liegen, dass diese Konstellation schlicht „eher nach Print aussieht“⁷⁴², damit in der Erwartung der Seitengestaltenden potenziell „seriöser“ wirkt? Oder liegen hier wirklich fast alle der oben zitierten Autoren falsch mit ihren Ratschlägen?

Ziefle kann ohnehin nicht nur Nachteile in der „positiven Polarität“ erblicken, von dunklen Zeichen auf hellem Grund also: „Wenn auch bei dieser Darstellungsart mehr Flimmerwirkung in Kauf genommen werden muss, werden auf der anderen Seite Bildschirmreflexionen und Spiegelungen weniger störend wahrgenommen, es erfolgt eine Angleichung der Leuchtdichten der Vorlage (Papier) und der Bildschirmanzeige, was insgesamt zu einer Verringerung des belastenden ständigen Wechsels zwischen Hell- und Dunkeladaption führt“⁷⁴³.

Wie dem auch sei: Ein weiterer Aspekt überrascht. Der nämlich, dass die potenziell inhaltsstützende, die ordnende, aber auch die emotionale Wirkung von Farbe nur sehr selten thematisiert wird in der Literatur – zumindest im spezifischen Bezug auf Schrift. Thissen meint zwar: „Textfarben sollten nicht als Selbstzweck oder Dekoration eingesetzt werden, sondern Aussagen unterstreichen“⁷⁴⁴. Doch vielleicht liegen die meisten seiner Kolleginnen und Kollegen auf einer Linie mit Weinman, die sagt: „Ich glaube nicht daran, dass jemand die Wirkung einer Farbe auf andere vorhersagen kann. Dies ist für meine Begriffe völlig subjektiv, egal, was in den Büchern steht“⁷⁴⁵. Immerhin rät Lazar dazu, von blauer Schrift jenseits von Verlinkungen abzusehen – nicht aus ästhetischen Gründen allerdings, sondern funktionalen:

⁷³⁹ Lankau 2000, S. 368

⁷⁴⁰ Vgl. BArbV

⁷⁴¹ Hellbusch 2005, S. 11

⁷⁴² Nielsen/Tahir 2002, S. 51. Übersetzung M.L.

⁷⁴³ Ziefle 2002, S. 61

⁷⁴⁴ Thissen 2003, S. 100

⁷⁴⁵ Weinman 2000, S. 170

„Obwohl keine spezielle kulturelle Bedeutung dahinter steht, hält sich der alte Standard für Textfarbe, dass blauer Text einen Link bezeichnet und violetter oder roter Text einen bereits besuchten Link“⁷⁴⁶. Blauer, jedoch unverlinkter Text könne daher leicht verwirren.

Der Vollständigkeit halber sei schließlich noch auf ein Phänomen eingegangen, das zumindest auf professionell gestalteten Webseiten kaum noch nachzuweisen ist: Text vor „gekachelten“, sich wiederholenden Hintergrundbildern. Im Prinzip gilt für diese „Bild-Fonds“ natürlich fort, was bereits für „plane“ Hintergrundfarben von Texten festgestellt wurde: Es ist auf einen ausreichenden Text-Grund-Kontrast zu achten⁷⁴⁷ – und zwar in allen Bereichen des inkriminierten Bildes, das daher am besten von vorneherein auf schwache Texturierung und dezente Muster, also möglichst geringe interne Helligkeitskontraste⁷⁴⁸ zu überprüfen und nötigenfalls zu verwerfen ist⁷⁴⁹ – oder alternativ durch zusätzliches Weichzeichnen in einen visuellen „Milchglaszustand“ übersetzt werden sollte⁷⁵⁰.

Insgesamt überwiegt jedoch die vernehmliche Meinung, von bildlichen Hintergründen von Schrift sei prinzipiell besser abzusehen. Denn diese produzierten eine visuelle Unruhe, die „vor allem am Zeilenanfang wie ein lautes Rauschen mit der Textinformation konkurriert“⁷⁵¹. Auch Pring stellt fest: „Der Lesbarkeit dienen die Muster (...) so gut wie nie“⁷⁵². Auch Matthai et al. halten sich nicht mit Optimierungsratschlägen auf: In Bezug auf Hintergrundbilder propagieren sie unumwunden strikten Verzicht⁷⁵³, genau wie Lazar es tut⁷⁵⁴.

3.11.4 Empirische Erkenntnisse zu Schriftfarbe am Bildschirm

3.11.4.1 Rote Schrift auf weißem Grund

In der bereits diskutierten Studie von Dyson und Gregory⁷⁵⁵ fand sich, wie erläutert, kein Einfluss der Schriftfarbenvariation auf die Lesbarkeit; rote Schrift auf weißem Grund erzielte keine merkbaren Verbesserungen oder Einbußen in Lesegeschwindigkeit oder Textverständnis im Vergleich zu schwarzer Schrift auf weißem Grund.

⁷⁴⁶ Lazar 2005, S. 146 (Übersetzung M.L.)

⁷⁴⁷ Vgl. Kiehn/Titzmann 1998, S. 74

⁷⁴⁸ Vgl. Nielsen 2000a-, S. 126

⁷⁴⁹ Kommer/Mersin 2002, S. 195

⁷⁵⁰ Vgl. Thissen 2003, S. 99

⁷⁵¹ Wirth 2005

⁷⁵² Pring 2000, S. 164

⁷⁵³ Vgl. Kesper et al. 2004, S. 681

⁷⁵⁴ Vgl. Lazar 2005, S. 144f.

⁷⁵⁵ Vgl. Dyson/Gregory 2002, S. 326ff.

3.11.4.2 Verschiedene Farbkombinationen im Vergleich

Hill und Scharff testeten in ihrem bereits vorgestellten Experiment⁷⁵⁶ die Leseleistungen ihrer Probanden unter den Bedingungen verschiedener Schriftfarbe/Farbgrund-Kombinationen: schwarze Schrift auf hellgrauem Grund, schwarze Schrift auf weißem Grund, grüne Schrift auf gelbem Grund, rote Schrift auf grünem Grund, weiße Schrift auf dunkelblauem Grund sowie gelbe Schrift auf dunkelblauem Grund.

Die besten Messergebnisse zeigten sich – über die drei Schriftarten *Times*, *Arial* und *Courier* hinweg – in der Kombination „grün auf gelb“. Nicht viel schlechter schnitten die Kombinationen „Schwarz auf Weiß“ sowie „Schwarz auf Grau“ ab. Diese drei Konstellationen erwiesen sich als signifikant besser lesbar (im Schnitt um etwa 9 Prozent) als alle übrigen Konstellationen, in denen die Schriftfarbe einen geringeren Helligkeitswert aufwies als der Grund. Den schlechtesten Durchschnitts-Messwert fanden die Forscher für weiße Schrift auf schwarzem Grund. Interessant war der Befund, dass kursiv gesetzte Texte weitaus „unempfindlicher“ auf Farbvariationen reagieren als normal geschnittene; wiewohl die Ergebnisse im Kursivsatz insgesamt schlechter ausfielen als für die Normalschriften, übten im Detailabgleich der Kursiven die Variationen der Farben keinen merklichen Einfluss auf die Reaktionszeiten aus.

In einer Nachfass-Erhebung mit 22 Teilnehmenden, diesmal nur unter Verwendung der Schriftarten *Times* und *Arial*, fanden Hill und Scharff zudem, dass sich die Kombination „Schwarz auf Mittelgrau“ (Grauwert etwa 35 %) als am besten lesbar herausstellte. Die schlechtesten Werte erzielten die Kombinationen „schwarze Schrift auf weißem Grund“ sowie „schwarze Schrift auf dunkelgrauem Grund“ (Grauwert etwa 60 %). Diese beiden Kombinationen erwiesen sich sogar als signifikant schlechter lesbar als die Kombination „Schwarz auf Mittelgrau“. Aus den Resultaten der beiden Erhebungen folgern die Forscher, dass Negativsatz nicht zu empfehlen sei am Bildschirm; jedoch sei darauf zu achten, dass die Hintergrundfarbe nicht zu hell, gar weiß ausfalle unter schwarzen Buchstaben.

3.11.4.3 Schwarze Schrift auf planem und texturiertem Farb-Hintergrund

Scharff et al. untersuchten 1999 die Auswirkungen von Hintergrundfarben und Hintergrundmustern auf die Lesbarkeit⁷⁵⁷. Sie konfrontierten 52 Probanden mit Texten, in denen sich Verbalisierungen verschiedener grafischer Formen fanden (zum Beispiel „Quadrat“), die bildlich auch neben dem jeweiligen Textblock platziert waren. Die Teilnehmenden wurden gebeten, die Texte zu überfliegen und stets, wenn sie eines der „Figur-Wörter“ gefunden hatten, die entsprechende bildliche Figur anzuklicken. Scharff et al. maßen dabei die Dauer von Lesebeginn bis zum jeweiligen Klick auf ein Symbol als Reaktionszeit.

⁷⁵⁶ Vgl. Hill/Scharff 1997

⁷⁵⁷ Vgl. Scharff et al. 1999

Die Texte waren im Experiment durchgehend gesetzt in der Farbe schwarz, die Hintergründe bildeten jeweils eine plane Farbfläche, ein fein ziselirtes „Stoffmuster“, ein grobes „Wollmuster“ sowie ein sehr ausdifferenziertes und unregelmäßiges Linienmuster aus weißen und mittel- bis dunkelgrauen Linien; diese Hintergrundflächen waren in verschiedenen Testkonstellation überfärbt in grau, blau oder gelb. Es zeigte sich, dass die besten Werte für die planen Hintergründe gefunden wurden; alle texturierten Hintergründe, in welcher Farbe auch immer, behinderten demnach die Lesbarkeit, im Falle des Wollmusters sogar signifikant im Abgleich mit dem planen Hintergrund. Insgesamt erwiesen sich die grauen und gelbfarbenen Hintergründe als relativ besser als die in Blau gesetzten; den absolut besten Durchschnitts-Messwert fanden Scharff et al. für die hellgrau-plane Grundfläche, den schlechtesten für die blaue „Wollfläche“ (die Reaktionszeiten für diese fielen um satte 71 Prozent schlechter aus als für die plane Graufäche).

In Kontrollexperimenten (58 und 16 Teilnehmende) fanden Scharff und Alhumada 2002 und 2003 diese Ergebnisse im Wesentlichen bestätigt⁷⁵⁸. Es zeigte sich jedoch, dass die Identifikationsleistungen der (wiederum schwarz gesetzten) Texte zwar vor planem Hintergrund wiederum durchgehend die besten Messergebnisse induzierten – dass jedoch mit wachsendem Helligkeitskontrast zwischen Schrift und Farbgrund die Messzeiten für die Hintergründe „Wollmuster“ und „Linienmuster“ sich immer näher an die der planen Grundflächen annäherten. Bei einem Helligkeitskontrast von 45 Prozent zwischen Schriftfarbe und Farbgrund waren die Unterschiede in den Messwerten zwar nicht nivelliert, aber erheblich einander angenähert (die texturierten Flächen induzierten hier nur etwa um 5 Prozent schlechtere Messwerte als die plane Fläche). Bei 30 Prozent Helligkeitsdifferenz hatten die Leseleistungen für die texturierten Farbgründe dagegen noch um ein 2,5-Faches über denen der planen Grundfläche gelegen. In einer weiteren Nachfass-Erhebung stellte sich zudem heraus, dass sich die gemessenen Identifikationsleistungen der Probanden nicht verbesserten unter den Bedingungen von 85 oder 100 Prozent Text-Grund-Kontrast, bei Kontrasten unter 80 Prozent aber merklich verschlechterten.

3.11.4.4 Differenzen zwischen grauer Schriftfarbe und grauen Farbgründen

Alhumada und Scharff maßen in einem weiteren Experiment 2003 (13 Teilnehmende) die Identifizierbarkeit von Buchstaben, die sich in ihrer Graufärbungs-Intensität um 10, 20 oder 40 Prozent vom jeweiligen Grau des Schriftgrundes abhoben⁷⁵⁹. Sie fanden, dass sich die Identifikationsgenauigkeiten wie die Erkennungsgeschwindigkeiten mit wachsender Graukontrastierung sukzessive verbesserten – und zwar weitgehend unabhängig davon, ob Buchstabe oder Grund die dunklere „Materie“ bildete. Die Erkennungszeit sank mit wachsendem

⁷⁵⁸ Vgl. Scharff/Ahumada 2002; vgl. ferner Scharff/Ahumada 2003

⁷⁵⁹ Vgl. Ahumada/Scharff 2003, S. 67ff.

Kontrast um fast 40 Prozent zwischen 10- und 40-prozentiger Graufächen-Differenz; hingegen wurde bei 10-prozentiger Differenz der Grau-Intensitäten von Buchstaben und Untergründen nur jede fünfte Letter korrekt identifiziert, bei 40 Prozent Kontrast dagegen im Schnitt 80 Prozent.

3.11.5 Zwischenfazit: Schriftfarben

Die vorliegenden empirischen Forschungsergebnisse zum Thema „Schriftfarbe“ und „Farbgrund“ zeigen im Wesentlichen in eine ähnliche Richtung wie die Vorgaben der relevanten Literatur. Danach scheint Positivsatz (also dunkle Schrift auf hellem Grund) am Bildschirm insgesamt besser lesbar als Negativsatz. Die Konstellation „schwarze Schrift auf weißem Grund“ jedoch als kontrastreichste Ausprägung des Positivsatzes scheint nicht die beste Wahl zu sein; eher spricht einiges für einen leicht verringerten Helligkeitskontrast zwischen Schriftfarbe und Farbgrund, der sich allem Anschein nach im Bereich von 80 bis 90 Prozent Helligkeitsdifferenz bewegen sollte. Gut lesbar am Bildschirm wäre in diesem Sinne also beispielsweise eine schwarze Textschrift auf einem Grauraster von 15 bis 20 Prozent als Untergrund. Ebenfalls gut lesbar in diesem Verständnis ist, folgt man den Erkenntnissen von Hill und Scharff, grüne Schrift auf gelbem Grund.

4 Das Experiment

4.1 Notwendigkeit einer Studie

Die vorangegangenen Kapitel haben hinreichend deutlich werden lassen: Die Probleme und Fallstricke der Browser-Typografie sind, was die Partei der Webdesigner als Kommunikatoren betrifft, gut und durchaus ausreichend tief dokumentiert. Dringender Forschungsbedarf existiert daher in diesem Teilfeld nicht. Denkbar wäre zwar beispielsweise eine Befragung von Kommunikatoren zu den von ihnen angewandten Tricks und Wegen, den Unzulänglichkeiten von HTML und CSS kreativ zu Leibe zu rücken und eine möglichst verlässliche Typografie über unterschiedliche Browser, Bildschirmarten und Betriebssysteme hinweg zu gewährleisten. Möglich wäre durchaus auch eine systematische und erschöpfendere Betrachtung populärer (journalistischer) Websites auf ihren Umgang mit Typografie hin in Form einer dann eher technisch orientierten Inhaltsanalyse.

Weitaus dringlicher jedoch erscheint eine detailliertere Befassung mit der „Gegenseite“: der Partei der Rezipienten nämlich und deren berechtigten Anforderungen nach „gut lesbarer“ Typografie. So bekannt nämlich die „Fußangeln“ sind, die Gestaltern von Webseiten in Sachen Typografie lauern, so uneinheitlich und umstritten sind die Konsequenzen und Praxisratschläge, die Autoren aus der Erkenntnis dieser Unzulänglichkeiten ziehen. Ein bemerkenswertes, wenn auch erklärliches Manko: Wo die Herstellerseite ihre weitestgehend technisch basierten Probleme sehr genau identifiziert hat und bereits sehr weit in der Entwicklung und Publizierung von Strategien zur Behebung dieser Schwierigkeiten gediehen ist, sind die Sorgen und Nöte der Rezipientenseite der überwiegenden Mehrheit der Autoren ein zwar wohlwollendes, aber doch oft vages oder hoch spekulatives Augenmerk wert. Dass technische Probleme leichter zu analysieren sind als Lesekomforts, erklärt diese augenfällige Diskrepanz zwar, macht sie aber längst nicht erträglicher.

Nur ein Beispiel mag den Missstand illustrieren. So herrscht, wie gezeigt, Einigkeit in der Feststellung, dass aufgrund der geringen Auflösungsqualität zeitgenössischer Computer-Bildschirme die Serifen und feinen Strichstärken-Differenzen in Antiqua-Schriftarten am Monitor entweder „verschluckt“, meist jedoch in ihren Proportionen extrem verzerrt werden. Die Schlüsse, die eine Mehrheit der zitierten Autoren aus diesem Faktum zieht, wirken dagegen fast durchgehend argumentativ sehr gewagt unterfüttert. Die meisten leiten aus der Monitor-Verzerrung der Zeichendarstellung eine deutlich verschlechterte Lesbarkeit der Schrift insgesamt ab – eine zumindest gewagte Unterstellung von Kausalität.

Diese Feststellung gewinnt weiter an Brisanz, zieht man in Erwägung, dass durchaus einige, wenn auch, wie gezeigt, nur relativ wenige empirische Untersuchungen vorliegen zur Lesbarkeit typografischer Anordnungen in Browsern – darunter einige, die die im vorigen Absatz wiedergegebene These mehr oder minder eindeutig widerlegen. Die Mehrheit der zitier-

ten Autoren jedoch rekuriert so gut wie gar nicht auf die Forschungs-Erkenntnisse, was durchaus fahrlässig genannt werden muss.

Doch auch der Stand der empirischen Forschung zur Typografie am Bildschirm im Allgemeinen und zu Typografie in Webbrowsern im Besonderen ist durchaus nicht völlig befriedigend. Dies ist – neben der allgemein immer noch eher dürftigen Menge vorliegender Studien überhaupt – zum einen festzumachen an der teilweise sehr geringen Teilnehmerzahl insbesondere vieler US-amerikanischer Studien, die deren Ergebnissen nur eingeschränkte Aussagekraft zu verleihen vermag (Experimente mit teilweise weniger als 20, ausschließlich aus studentischem Milieu rekrutierten Probanden sind zumindest hinterfragenswert in Sachen Verallgemeinerbarkeit der Resultate). Insbesondere aber ist festzuhalten, dass die Auswirkungen der Manipulation einiger durchaus wichtiger typografischer Faktoren am Bildschirm – beispielsweise des Zeilenabstands – nach meinen Recherchen bislang nur Gegenstand einiger weniger „Pilot“-Untersuchungen war, deren Resultate mithin einer Überprüfung und Verfestigung bedürfen. Andere Studienergebnisse wiederum – beispielsweise jene, die den typografischen Faktor „Zeilenbreite“ unter die Lupe nahmen – ergaben ausgesprochen widersprüchliche Ergebnisse, so dass an dieser Stelle der Forschungsbedarf nicht minder evident ist.

Hinzu tritt ein in meinen Augen zentrales Versäumnis fast aller vorliegenden Untersuchungen empirischer Natur: Von wenigen Ausnahmen abgesehen wurden die beschriebenen Einzelausprägungen und -faktoren von Schriftanordnungen in den vorliegenden Studien weitestgehend isoliert betrachtet. Es wurde insbesondere bislang kaum untersucht, inwiefern nicht nur Modifikationen einzelner typografischer Faktoren innerhalb von Schriftanordnungen (wie beispielsweise der Schriftart) den Leseerfolg beeinflussen, sondern auch mögliche Interaktionen dieser Faktoren. Auch Bayer fand 2003 keine Vorhaben dieses Inhalts: „Für die Bildschirmtypografie existieren noch keine Studien, die sich mit der Interaktion der Variablen Schriftart, -größe, Zeilenlänge und Durchschuss beschäftigen.“⁷⁶⁰

Es fehlt also auf quantitativer wie qualitativer Ebene nach wie vor an belastbaren und empirisch begründeten Erkenntnissen zur Wirksamkeit verschiedener typografischer Anordnungen in Webbrowsern. Die Folge: Webdesign vollzieht sich heute immer noch weitestgehend auf Basis spekulativer „Standards“. Was gute, lesefreundliche Typografie in Webbrowsern ist, wird nach wie vor hauptsächlich vermutet. Die Deutungshoheit besitzen dabei nach wie vor ausgewiesene oder vermeintliche Praktiker des Fachs.

Ich habe den Versuch unternommen, mit der im Folgenden beschriebenen Studie diesem Misstand ein wenig abzuhelpfen. Schließlich gilt unverändert: „Wenn es in der Wissenschaft

⁷⁶⁰ Bayer 2003, S. 100

zu einer Auseinandersetzung zwischen aussagekräftigen Daten und der Meinung von Experten kommt, gewinnen immer die Daten“⁷⁶¹.

Die Fragestellung war zunächst so simpel wie anspruchsvoll: Welche typografischen Anordnungen sind in gängigen Webbrowsern gut, welche eher schlechter lesbar?

4.2 Sinn und Machbarkeit einer Browser-Lesbarkeitsstudie

Eine Grundsatzfrage freilich ist dringend zu klären, bevor die Diskussion meines empirischen Vorhabens weitergeführt werden kann. Und diese Frage berührt nicht weniger als den Wesensgehalt des Projekts insgesamt.

Denn: Wenn die Betrachter von Webseiten nach wie vor so viel Einfluss nehmen können auf die Darstellung von Webseiten in „ihrem“ Browser auf „ihrem“ PC, wie ich dies im vorherigen Kapitel beschrieben habe – ergibt dann eine Studie wie die von mir projektierte überhaupt einen Sinn? Kann es Nutzen entfalten, die Lesbarkeit von Schriftkonstellationen zu überprüfen, die jeder Web-User theoretisch jederzeit – zumindest in Teilbereichen – in seinem Sinne manipulieren und damit ihres vom Webdesigner konzipierten Ur- oder besser: erwünschten Idealzustands entheben kann? Wenn insbesondere – dies ist das Hauptproblem – ein Großteil von Betrachtern im Web die Schriftgröße auf einer Seite nach individuellem Gusto herauf- oder herabsetzen kann, lohnt dann überhaupt eine Messung der Lesbarkeit verschiedener Größen, die im Ergebnis die Definition einer vermeintlich „optimalen“, realiter aber bestenfalls einer wünschenswerten Schriftgröße zeitigen könnte?

Die Antwort ist in hohem Maße abhängig davon, für wie „mündig“ und – insbesondere – für wie „experimentierfreudig“ man den wie auch immer definierten „durchschnittlichen Web-Nutzer“ hält. Die Frage ist also,

- wie viele Webnutzer wissen überhaupt, dass sie durchaus Einfluss nehmen können auf die typografische Darstellung von Webseiten? Und wenn sie um die Möglichkeit wissen, kennen sie dann auch das Prozedere?
- wie viele Webnutzer, die das Prozedere kennen, manipulieren auch tatsächlich aktiv die Schriftdarstellung in ihrem Browser? Und wenn ja, wie oft?

Leider ist nach meinen Recherchen bislang weder erhoben noch verlässlich bestimmbar, wie groß die beiden Populationen sind. Ich denke jedoch, dass Lazar nicht falsch liegt, wenn er ausführt: „Den wenigsten Nutzer ist bekannt, dass sie die Textgestaltung des Entwicklers manipulieren können, und daher wird dies auch selten getan“⁷⁶². Das World Wide Web ist längst ein Massenmedium, dessen Nutzer zu überwältigendem Anteil ausschließlich Konsum-

⁷⁶¹ Zimbardo/Gerrig 2004, S. 28

⁷⁶² Vgl. Lazar 2005, S. 148f. (Übersetzung M.L.)

menten, zu geringstem Anteil jedoch Kommunikatoren sind. Das heißt: Es handelt sich beim breiten Web-Publikum nicht mehr, wie in der Anfangszeit der frühen 1990er Jahre, um eine eher kleine Gemeinde von Computerbegeisterten, bei denen eine gewisse Versiertheit im Umgang mit Programmen zu erwarten wäre, also auch in der Handhabung detaillierterer Voreinstellungs-Konstellationen von Browsern. Eher ist davon auszugehen, dass ein überwältigender Anteil von Nutzern ihren Browser gleichsam als „Web-Fernseher“ betrachtet, den sie ein- und wieder ausschalten. Und sonst gar nichts: Als Schriftart ist *Times New Roman* eingestellt, und die Schriftgröße steht auf „mittel“. Und wenn der Online-Ausflug beendet ist, wird der Browser geschlossen. Unmanipuliert, in typografischer Hinsicht.

4.3 Lesbarkeit und ihre Operationalisierung

Die typografische, die gestalterische, auch die allgemeine Literatur über Webdesign geizt nicht mit dem Begriff: Lesbarkeit. Gute Lesbarkeit ist damit gemeint in aller Regel – jenes vorrangige, viele meinen auch: das Ziel schlechthin jeder typografischen Arbeit. Implizit, davon ist auszugehen, hatten alle der im vorigen Kapitel zitierten Autoren „gut lesbare“ typografische Anordnungen im Sinn, wenn sie mal jenen Zeilenabstand als empfehlenswert, mal eine Schriftart als „eher ungeeignet“ klassifizierten. Doch was ist „gute Lesbarkeit“? Ist „Lesbarkeit“ in ihrer Güte empirisch operationalisierbar, mithin also auch quantifizierbar? – das wäre immerhin die Voraussetzung, die eine Studie wie die von mir projektierte überhaupt erst möglich machte⁷⁶³. Und wenn „Lesbarkeit“ operationalisierbar ist – in welcher Einheit ist sie dann bestimmbar?

Wohlgedacht: Die Rede ist hier und wird weiterhin sein von „typografischer Lesbarkeit“, dem Gewerk also, Texte in visuelle Form zu gießen, die sie „gut rezipierbar“ geraten lässt – nicht gemeint ist die „sachliche Lesbarkeit“, die inhaltliche Textgestaltung also, jenes spannende Feld, das auch in Journalistik und Sprachwissenschaft immer stärkere Aufmerksamkeit erfährt⁷⁶⁴. In diesem Sinne werde ich auf den Folgenden Seiten auch nicht jene frappierenden Befunde von Spool et al. diskutieren, die feststellten, dass die „Usability“ von Web-Angeboten merklich wachse, wenn die Schwierigkeit der Texte auf den Seiten zunehme, dass genau diese Seiten zudem mehrheitlich als „klar“, „seriös“, „vollständig“ und „nützlich“ empfunden wurden von Teilnehmern an Usability-Tests⁷⁶⁵. So interessant und einer Hinterfragung würdig diese Befunde auch sein mögen: Sie berühren das Thema dieser Arbeit nur sehr marginal. Wenn daher im Folgenden die „typografische Lesbarkeit“ als Phänomen untersucht wird, so wird behelfsweise allgemein unterstellt, dass die zu untersuchenden Schriftzeichen Texte visualisieren, in deren Aufbau, Wortwahl und grammatikalischer Richtigkeit das

⁷⁶³ Vgl. Zimbardo/Gerrig 2004, S. 31; vgl. ferner Kromrey 2006, S. 189f.

⁷⁶⁴ Vgl. beispielhaft Ballstaedt et al. 1981

⁷⁶⁵ Vgl. Spool et al. 1999, S. 70f.

Bemühen eines Autors um „Verständlichkeit“ festzumachen ist; dass es sich, kurz gesagt, um gemeinhin „lesbare“ Texte im stilistischen wie inhaltlichen Sinne handelt.

Was also ist „Lesbarkeit“ im typografischen Sinne? Für McLean überraschenderweise ein „gefährliches – und interessantes – Wort. Es ist gefährlich, weil es oft so benutzt wird, als hätte es eine definitive oder absolute Bedeutung. Die hat es nicht. Es ist ein persönliches Wort, weder wissenschaftlich noch präzise. Wenn jemand sagt: ‚Das ist lesbar‘, meint er nur, dass er selbst es lesen kann. Er weiß aber nicht, ob ich es kann“⁷⁶⁶. In anderem Zusammenhang als dieser Arbeit wäre dieser Gedanke wert, weiter verfolgt zu werden; für die hier angestrebte Operationalisierung des Begriffs allerdings hilft McLeans Einlassung wenig weiter.

Stattdessen steht zu vermuten, dass Kommer und Mersin eine herrschende Meinung wiedergeben, wenn sie feststellen: „Die Lesbarkeit – ein komplexer Begriff – bemisst sich an der Zeit, die erforderlich ist, einen Text ohne Ermüdungserscheinungen zu erfassen“⁷⁶⁷. Bergner argumentiert inhaltsgleich⁷⁶⁸. Auch die im vorhergehenden Kapitel referierten empirischen Studien zu typografischen Phänomenen am Bildschirm folgerten zu überwiegendem Anteil Tauglichkeiten bestimmter typografischer Anordnungen aus den unterschiedlichen Lesegeschwindigkeiten, in denen diese Anordnungen rezipiert wurden.

Ist also ein Text, der typografisch so angelegt ist, dass Lesende ihn schnell rezipieren können, damit automatisch auch gut lesbar? Es existieren zumindest Ansätze, die andere Kriterien als die bloße Rezeptionsgeschwindigkeit zum Maßstab erheben. Diese werden im Folgenden diskutiert.

4.3.1 Legibility und Readability: Sichtbarkeit und Lesbarkeit

Zunächst ist jedoch noch auf eine bedeutsame Zweigeleisigkeit in den bisherigen empirischen Forschungsansätzen rund um die Qualität typografischer Darbietungen hinzuweisen, der weit über eine rein terminologische Unterscheidung hinaus reicht. Während nämlich ein Großteil vor allem früherer und frühester Untersuchungen (bis in die Anfänge des 20. Jahrhunderts hinein) sich vor allem auf die Qualität der bloßen Erkennbarkeit von Buchstaben, Wörtern und Sätzen fokussierte⁷⁶⁹, steht in moderneren Studien der eigentliche, praktizierte Leseprozess als ganzheitliche Verrichtung unter Beobachtung. Es wird daher fein und sinnvollerweise unterschieden zwischen „Legibility“ (im weiteren Sinne also der „Sichtbarkeit“ oder „Erfassbarkeit“ von Buchstaben, Wörtern und sonstigen typografischen Konstellationen weitgehend unabhängig vom transportierten Inhalt) und der „Readability“ (der „Lesbarkeit“ als meist linearer Prozess der inhaltserfassenden Informationsaufnahme von Schriftzeichen).

⁷⁶⁶ McLean 1992, S. 42 (Übersetzung M.L.)

⁷⁶⁷ Kommer/Mersin 2002, S. 116

⁷⁶⁸ Vgl. Bergner 1990, S. 20

⁷⁶⁹ Vgl. Rehe 1981, S. 19ff.

Dass „Readability“ durchaus einen Grad an „Legibility“ der publizierten Schriftzeichen voraussetzt, sei unbestritten; dass dies umgekehrt nicht notwendigerweise der Fall sein muss, dass „Readability“ also den eindeutig umfassenderen Ansatz bildet, wird allerdings im Folgenden unterstellt.

4.3.2 Die Legibility: Bekannte Forschungsdesigns

Die ersten bekannten empirischen Forschungsansätze zur typografischen Qualität dürfen aus heutiger Sicht samt und sonders als „Legibility“-Erhebungen gelten. Wendt kommt das Veienst zu, die gebräuchlichsten bislang eingesetzten Erhebungsweisen zusammengefasst zu haben. Er unterscheidet folgende Methodiken⁷⁷⁰:

- Das Tachistoskop

Das Tachistoskop ist eine Apparatur, die es ermöglicht, Buchstaben, Wörter oder Sätze in zeitlich exakt definierten, kurzen „Aufblitzern“ zu präsentieren – vergleichbar dem Verschluss eines Fotoapparates. Relevanter Messwert ist dabei die jeweils kürzeste Darbietungsdauer, in deren Anschluss der Betrachter den angezeigten Reiz zu benennen vermag.

- Der Entfernungsschwellenwert

In diesem Verfahren werden dem Betrachter Buchstaben, Wörter oder Sätze aus großer Entfernung dargeboten – beginnend mit einem räumlichen Abstand, der eine Identifikation des Reizes nach menschlichem Ermessen unmöglich macht, und diesen Abstand sukzessive verringern, bis der Betrachter den Reiz erkennt. Die Distanz, an dem die Erkennbarkeit einsetzt (also der Betrachter den Reiz aktiv benennen kann), ist dabei die abhängige Variable.

- Helligkeitsschwellenwert

In diesem Verfahren bleibt der Abstand zwischen Betrachter und Reiz konstant. Variiert wird stattdessen die Stärke der Beleuchtung des Reizes. Messwert ist in diesem Verfahren der Helligkeitsgrad, an dem der Betrachter den Reiz identifiziert.

- Das Luckiesh & Moss Visibility Meter

Auch dieses Verfahren sieht konstante Distanzen zwischen Reiz und Betrachter vor. Ein komplexes Linsensystem sorgt stattdessen dafür, dass der jeweilige Reiz sukzessive schärfer dargeboten wird; abhängige Variable ist in diesem Falle der Schärfeegrad, an dessen Erreichen der Betrachter den jeweiligen Reiz zu erkennen und benennen imstande ist.

Den vorgenannten Forschungsdesigns ist gemein, dass sie im Grunde nur eine Frage beantworten: An welchem Schwellenwert ist beispielsweise ein „A“ als Buchstabe, die Zeichenfolge „Apfel“ als Wort identifizierbar? Gemessen wird eben die „Erkennbarkeit“, die „Legibility“ von Zeichen in variablen typografischen Darbietungen. Dass aber beispielsweise ein „A“ in

⁷⁷⁰ Vgl. Wendt 1992, S. 279ff.

einer bestimmten Schriftart rascher als solches erkannt wird als in einer anderen Schriftart – sagt diese Erkenntnis etwas aus über die Qualität eines Alphabets an sich? Sind Rückschlüsse erlaubt, dass in der „überlegenen“ Schriftart gesetzte Texte auch inhaltlich besser „haften“ bleiben beim Betrachter – was aus journalistischer Sicht ganz bestimmt eine höherrangige Qualität besitzt als die bloße Letter-Identifikation in möglichst kurzer Spanne?

Als erstes Indiz für die Tauglichkeiten von Zeichen und Alphabeten dürfen „Legibility“-Tests zweifellos gelten. Ihnen die Qualität echter Gradmesser von Leserlichkeit typografischer Anordnungen zuzubilligen, erscheint jedoch allzu hoch gegriffen.

4.3.3 Ein Vorrang für die Readability

Ich habe mich aus den vorgenannten Gründen entschlossen, eine Untersuchung auf Grundlage der Zielformulierungen der „Readability“ zu konzipieren. Das heißt: In der projektierten Studie sollten typografische Anordnung weniger als bloß visuelle Phänomene begriffen werden denn vielmehr als funktionale Träger von (journalistischer) Information. Es sollte also Ziel der Studie sein, die Begünstigung oder auch Behinderung inhaltserfassender Leseprozesse durch ihre individuelle typografische Ausformulierung zu klären.

4.3.4 Mögliches Kriterium guter Lesbarkeit: Wissens-Zuwachs

Texte werden im Allgemeinen verfasst und publiziert, um Inhalte zu transportieren – Inhalte, die Verfasser bei potenziellen Lesern zu verankern suchen. Diese Feststellung mag zunächst banal erscheinen, insbesondere in Bezug auf journalistische Texte. In der Tat: Der rein spielerische Umgang mit Typografie – die „inszenierende Typografie“⁷⁷¹ –, um beispielhaft eine Ausnahme von der einleitenden Annahme anzuführen, bildet eher die Ausnahme in der modernen Medienlandschaft, in journalistisch orientierten Angeboten ohnehin.

Doch die vermeintliche Banalität der einleitenden Feststellung verblasst erheblich weiter, hält man sich vor Augen, dass, wie gezeigt, die „Lesbarkeit“ und damit die Güte typografischer Konstellationen scheinbar überwiegend gleichgesetzt wird mit der Geschwindigkeit, in der sie rezipierbar sind. Ob und in welcher Anordnung die Typografie eines Textes dagegen die Vermittlung von dessen Inhalten befördert oder behindert, lässt das Operationalisierungs-Kriterium „Lesedauer“ außer Acht.

Aus dieser Überlegung wäre also durchaus ein alternatives Kriterium zur Bestimmung und Quantifizierung der „Lesbarkeit“ eines Textes entwickelbar (und wurde als solches ja auch in einigen der im vorigen Kapitel beschriebenen empirischen Studien begriffen und operationalisiert): der Wissenszuwachs. So wäre durchaus denkbar, die Qualität einer Textanordnung

⁷⁷¹ Vgl. Willberg/Forssmann 2005, S. 14

daran zu messen, welche und wie viele der im Text enthaltenen Informationen vom Leser nicht nur „überflogen“, sondern nachhaltig aufgenommen, verstanden und gespeichert wurden. Ergeben sich unterschiedliche Wissenszuwachs-Erfolge zwischen Menschen, die zwar den selben Text, diesen Text jedoch in unterschiedlichen typografischen Anordnungen vorgelegt bekamen, so kann dies durchaus ein Indiz dafür sein, dass die Informationsaufnahme und -speicherung durch bestimmte typografische Anordnungen begünstigt, durch andere eher gestört wird. In diesem Sinne wäre die Qualität verschiedener typografischer Anordnungen operationalisierbar, indem die Lernerfolge verschiedener Probanden miteinander verglichen werden.

4.3.4.1 Operationalisierbarkeit des Wissens-Zuwachses

Die empirische Messung von inhaltlichem Wissen – und vor allem: des Wissenszuwachses – wäre prinzipiell möglich durch das Instrument der Befragung. So könnte der inhaltliche „Leseerfolg“ im Anschluss an einen experimentellen Rezeptionsprozess gemessen werden durch die gezielte Befragungen von Probanden – mündlich oder schriftlich, in offenen oder geschlossenen „Multiple-Choice“-Fragen. Die Fragen müssten sich dabei auf die im Text transportierten Informationen beziehen – die Güte einer typografischen Anordnung würde in diesem Modell quantifizierbar über die Anzahl, eventuell auch die Tiefe inhaltlich korrekter Antworten der Probanden im Anschluss an den Leseprozess.

Zumindest für eine quantitative Messung des Wissens-Zuwachses schied in meinem Verständnis sehr früh der Ansatz von O’Hara und Kenton aus⁷⁷²: Die Autorinnen forderten ihre Probanden, wie gezeigt, im Anschluss an einen experimentellen Leseprozess auf, eine Inhaltszusammenfassung des rezipierten Textes zu verfassen. Abgesehen davon, dass eine Auswertung dieser „Summaries“ zweifellos sehr zeitintensiv ausfallen würde, müsste diese Auswertung wiederum recht komplex operationalisiert werden im Sinne einer überprüfenden Inhaltsanalyse.

4.3.4.2 Grenzen der Operationalisierbarkeit des Wissens-Zuwachses

Der Ansatz, die Güte einer typografischen Anordnung über die Menge gewonnener Informationen zu ermitteln, entbehrt nicht eines Charmes. Immerhin würde in diesem Verfahren Qualität (Informationsgewinn) über Quantität (Lesegeschwindigkeit) gestellt – ein im journalistischen Sinne sehr reizvoller Ansatz. Dennoch ergeben sich in diesem Modell auch folgende Probleme:

⁷⁷² Vgl. O’Hara/Sellen 1997

- Dass Nutzer bestimmte inhaltliche Fragen nicht beantworten können im Anschluss an den Leseprozess, muss nicht zwingend damit zusammenhängen, dass sie den Text nicht oder nicht aufmerksam genug gelesen haben. Vielmehr ist nicht auszuschließen, dass bestimmte Teilnehmer eine Information zwar flüchtig aufnehmen, aber nicht speichern, weil sie dem rezipierten Faktum schlicht keine Relevanz zumessen im Verlauf des Rezeptionsprozesses. In diesem Sinne könnte hier also eine verhängnisvolle, wenn auch nachvollziehbare Selektion der aufgenommenen und dann gespeicherten Inhalte durch Nutzer stattfinden, die aber mithin nicht unbedingt auf einen nicht erzielten, sondern eventuell nur qualitativ anderen „Wissenszuwachs“ deutet.

- Andererseits muss die Tatsache, dass Probanden inhaltliche Fragen nach Beendigung des Leseprozesses korrekt beantworten können, nicht oder zumindest nicht unmittelbar mit der Tatsache zusammenhängen, dass der Text aufmerksam und inhaltserfassend gelesen wurde. Eventuell war dem richtig respondierenden Probanden die korrekte Antwort auf die Frage bereits bekannt, bevor er den relevanten Text überhaupt zum Lesen vorgelegt bekam.

Zumindest die zweitgenannte Problematik ließe sich prinzipiell auffangen und minimieren. So wäre möglich:

- die Abfrage des inhaltlichen Vorwissens der Versuchsteilnehmer. Denn um einen Wissenszuwachs nach dem Lesen eines Textes zu konstatieren, muss eines zwingend bekannt sein: das Vorwissen, das gespeicherte Wissen des Probanden vor Erscheinen des Textes. Nur bei Bekanntsein dieses Datums kann ein Zuwachs von Wissen überhaupt konstatiert werden. Dieses Vorwissen wäre erhebbar in Form einer dem Leseprozess vorgeschalteten Befragung der Probanden.

Das Abfragen bestimmten Vorwissens freilich könnte wiederum potenzielle Teilnehmer eines Experiments bereits auf eindeutige Fährten lenken: Die Probanden könnten aufgrund der Vorbefragung begründet mutmaßen, welche Informationen im ihnen noch nicht bekannten, auf die Vorbefragung folgenden Text von Bedeutung sein könnten. Der anschließende Leseprozess würde damit aber bereits im Vorhinein determiniert: Viele Probanden würden potenziell nur Informationen aufnehmen und speichern, die absehbar für die anschließende Befragung relevant sein könnten.

Dieses Problem der absehbaren wie ungewollten Präparierung der Probanden für den anschließenden Lesetest wäre wiederum teilweise umgehbar durch eine Variierung des Experimentalkonzepts insgesamt. Dies könnte wie folgt aussehen:

- Die Probanden werden bereits vor Erscheinen des Textes, den durcharbeiten sie gebeten werden, über die Fragen informiert, die sie im Anschluss an den Leseprozess zu beantworten haben. Technisch realisierbar wäre sogar eine noch weiter gehende Variante: Die Beantwortung der inhaltlichen Fragen erfolgte dabei nicht im zeitlichen Anschluss, sondern bereits während des Leseprozesses. So könnte – zeitlich vor der oder sogar auch während der Prä-

sens des eigentlichen Textes – die im Text implizit beantwortete Frage gestellt werden: „Wie heißt der portugiesische Ministerpräsident?“, und die Antwort wäre beispielsweise zu geben in Form eines Mausklicks auf besagten Namen im Text.

Auch diese Variante entbehrt zweifellos nicht eines gewissen Reizes. Schließlich ist, folgt man unter anderem Wirth⁷⁷³, davon auszugehen, dass zumindest routiniertere User das Internet vor allem problemorientiert bemühen, also, um gezielte Antworten auf bestimmte, selbst- oder fremdgestellte Fragen zu erlangen. Virulent wäre aber in diesem Modell wiederum das Problem, dass nicht feststellbar wäre, ob eine korrekte Beantwortung von Testfragen Ausfluss einer sorgfältigen Textrezeption ist – oder der Proband lediglich vorhandenes Sachwissen abgerufen hat, im Zweifel sogar, ohne eine einzige Textzeile gelesen zu haben.

Einen weiteren, wirklich interessanten Ansatz verfolgten beispielsweise Dyson und Kipping in ihren Experimenten 1997 und 1998⁷⁷⁴: Sie stellten ihren Probanden die intelligente Frage, ob ein bestimmtes inhaltliches Faktum dem soeben gelesenen Text entnehmbar gewesen sei – „ja“ oder „nein“. Mit dieser Messmethodik schalteten sie zumindest potenziell das Problem des nicht erhobenen „Vorwissens“ aus, kontrollierten aber dennoch recht effizient die Inhaltstrennung der Probanden. Ein Problem der Methodik ist meines Erachtens jedoch, dass die Fragestellung etwas „indirekt“ konzipiert ist und sich eventuell nicht allen Probanden unmittelbar erschließen würde; eventuell erfordert sie daher eingehendere Erläuterungen, die vor allem in einem Online-Experiment schwerlich zu geben sind. Zudem bietet eine „Multiple-Choice“-Frage wie die von Dyson und Kipping vorgesehene Probanden immer noch eine „Vermutungs-Chance“ (bei „Ja-Nein“-Fragen sogar eine mit einer Erfolgswahrscheinlichkeit von 50 % selbst bei völliger Verweigerung aktiven Lesens), die wiederum die erhobenen Daten nur teilweise zuverlässig als Indizien eines sinnerfassenden Leseprozesses erscheinen lassen.

4.3.4.3 Zwischenfazit: Wissens-Zuwachs als Lesbarkeits-Kriterium

So interessant und aus insbesondere journalistischer Sicht nahe liegend der Ansatz ist, die Qualität typografischer Anordnungen über das Ausmaß inhaltlichen Wissenserwerbs zu bestimmen, so zahlreich sind die Fallstricke und Unwägbarkeiten der Operationalisierung. Wissens-Erwerb ist eine schwer quantifizierbare Variable, die Anzahl der Störeinflüsse in diesem Modell unabsehbar. Einzig das beschriebene Modell von Dyson und Kipping hielt ich eine Zeitlang für sehr erwägenswert, wenn auch, wie gezeigt, für eventuell etwas zu weit „um die Ecke herum gedacht“ und in seinem „Multiple-Choice“-Ansatz zu unzuverlässig. In der Gesamtschau erschien mir für das projektierte Vorhaben der Operationalisierungsansatz via „Inhaltsabfrage“ insgesamt als zu unsicher; ich habe ihn letztlich verworfen.

⁷⁷³ Vgl. Wirth 2004, S. 207ff.

⁷⁷⁴ Vgl. Dyson/Kipping 1997, S. 705f.; vgl. ferner Dyson/Kipping 1998, S. 155

4.3.5 Mögliches Kriterium guter Lesbarkeit: körperliche Beanspruchung

Eine nicht abwegige These unterstellt, dass Rezipienten, die Texte in einer wie auch immer geartet „schlechten“ typografischen Anordnung lesen, im Verlaufe des Konsumtionsprozesses mehr körperliche Mühe aufzuwenden haben als Menschen, die denselben Text in einer vermeintlich tauglicheren Konstellation zu verarbeiten haben. Der Grad der physischen Beanspruchung durch einen Leseprozess wäre damit ein weiterer potenziell geeigneter Indikator für die Qualität einer Schriftanordnung.

4.3.5.1 Operationalisierbarkeit des Grades körperlicher Beanspruchung

Wie aber misst man zuverlässig die körperliche „Beanspruchung“ während eines Leseprozesses? Rehe erwähnt hier das Prinzip der „Zwinker-Technik“⁷⁷⁵, das ist die Messung der „Lidschlag-Frequenz“, der „Blinkrate“: Je öfter danach ein Proband im Verlaufe des Leseprozesses die Augen erholungshalber schließt, desto schwerer scheint ihm das Lesen zu fallen – desto geringer also ist in diesem Forschungsansatz die Qualität der Typografie des dargebotenen Textes. Wendt schränkt allerdings ein: „Gleichzeitig ist die Häufigkeit des Augenzwinkerns aber auch ein Anzeichen für allgemeine Ermüdung oder Wachheit, die nichts mit den typographischen Eigenschaften eines gelesenen Textes zu tun haben muss“⁷⁷⁶. Analog zum Operationalisierungsansatz des „Wissens-Erwerbs“ wäre hier also eine Vorab-Messung des körperlichen „Urzustands“ vonnöten, um körperliche Degenerationserscheinungen, die während und durch einen Leseprozess ausgelöst werden, verlässlich zu quantifizieren.

Ebenfalls auf beobachtbare körperliche Symptome hebt der Ansatz ab, die Anzahl der Augen- beziehungsweise Pupillenbewegungen und Blickverweilungen im Prozess des Lesens zu erfassen und zur Grundlage zu nehmen für die Bewertung der typografischen Qualität einer Textanordnung. Die Augensprünge, also die bereits beschriebenen Blick-„Sakkaden“ und die Pausen zwischen je zwei Sakkaden, werden erfasst und für die Bewertung der jeweiligen typografischen Anordnung zugrunde gelegt: Je mehr und damit kürzere Augensprünge, so die Unterstellung dieses Operationalisierungsmodells, desto schlechter ist der Text lesbar.

Augenbewegungen über einem Textgesamt zu erfassen und vor allem: einzelnen Textpassagen zuzuordnen, erfordert allerdings aufwändige Messinstrumentarien, zumindest geschulte und aufmerksame Beobachter. Zwar werden so genannte „Blickbewegungskameras“ (auch: „Blickaufzeichnungsgeräte“) zunehmend erschwinglich und liefern im Zusammenspiel mit geeigneten Computer-Programmen auch interessante Ergebnisse⁷⁷⁷ – es sind jedoch Zweifel angebracht, ob Probanden mit einer „Labor-Brille“ auf der Nase tatsächlich noch Ergebnisse liefern, die Rückschlüsse zulassen auf den ungezwungenen Leseprozess „am heimischen

⁷⁷⁵ Vgl. Rehe 1981, S. 20

⁷⁷⁶ Wendt 1992, S. 281

⁷⁷⁷ Stoessel 2002, S. 80

Rechner“ (später wird noch zu sprechen sein über diesen „Laboreffekt“). Für eine Online-Studie, die zumindest als mögliche Erhebungsform bereits in der Konzeptionsphase meiner Untersuchung eine Rolle spielte, schied das Verfahren ohnehin aus; schließlich war es rein logistisch ausgeschlossen, jedem potenziellen Teilnehmer ein Blickaufzeichnungsgerät zur Verfügung zu stellen, damit er oder sie es am Ort der Wahl – Heim-PC, Bürorechner, Internet-Café oder wo sonst auch immer – aufsetzen und dann in Eigenregie die Studie würde durchführen können. Ähnliches gilt im Grunde für alle weiteren denkbaren Ansätze, das Ausmaß physischer Beanspruchung zu messen: Sei es die Herzfrequenz, der Atemrhythmus oder eben die Lidschlag-Häufigkeit – die Messung all dieser körperlichen Phänomene erforderte zumindest die Präsenz eines empirischen Beobachters beim Probanden oder das Vorhandensein technischer Messgeräte vor Ort.

Dillon et al. verzichteten in ihren Studien 2004 und 2006⁷⁷⁸ von vorneherein auf technisches Gerät; sie erhoben die „Ermüdung“ ihrer Testteilnehmer schlicht via Befragung. Die Probanden sollten ihre Zustimmung signalisieren zu Aussagen wie „mir tut der Nacken weh“ oder „ich fühle mich erschöpft“ im Anschluss an die Leseprozesse. Mit einer verlässlichen Messung der körperlichen Beanspruchung hat dies freilich nur sehr am Rande etwas zu tun; eine individuelle Ansicht ist eben etwas definitiv anderes als eine medizinische Erhebung der körperlichen Beanspruchung.

4.3.5.2 Zwischenfazit: körperliche Beanspruchung als Lesbarkeits-Kriterium

Auch das Kriterium „physische Beanspruchung im Leseprozess“ erscheint wenig praktikabel als Instrument zur Messung der Qualität typografischer Anordnungen – finanziell und logistisch ohnehin, im Rahmen eines Online-Experiments in besonderem Maße. Eine Erhebung der körperlichen Beanspruchung via Befragung schied aus dem genannten Grunde ebenfalls aus. Ich verwarf daher auch diese Operationalisierungs-Option für mein empirisches Vorhaben.

4.3.6 Mögliches Kriterium guter Lesbarkeit: Urteil der Probanden

Eine dritte alternative Messmethodik der Lesbarkeit typografischer Anordnungen soll an dieser Stelle nur der Form und Vollständigkeit halber zur Sprache kommen: Die Lesbarkeit würde in diesem Modell operationalisiert über das einschätzende Urteil der Probanden. Die Teilnehmenden würden also per Befragung um eine Einschätzung der Lesbarkeit einer vorgelegten typografischen Anordnung gebeten.

⁷⁷⁸ Vgl. Dillon et al. 2004; vgl. ferner Dillon et al. 2006

4.3.6.1 Das Urteil der Probanden – kein sinnvoller Ansatz

Die Nachteile dieses Ansatzes liegen freilich auf der Hand. Von einer empirischen Objektivierung des Phänomens „Lesbarkeit“ kann definitiv nicht mehr die Rede sein, würde diese anhand notwendigerweise subjektiver Meinungen operationalisiert. Was in Sonntagsfragen erwünscht ist, kann dem Ziel meines Vorhabens daher nicht gerecht werden: individuelle Urteile. Ich verwarf daher auch dieses Modell sehr rasch.

4.3.6.2 Das Urteil der Probanden – jenseits der Lesbarkeit durchaus interessant

In einer anderen Hinsicht allerdings entfaltet das Modell der Meinungsabfrage dann durchaus doch noch Potenzial. Wie bereits gezeigt, bergen insbesondere Schriftarten in ihrer Formgestaltung so etwas wie eine unterschwellige Tonalität, eine Anmutung, eine Ausstrahlung, die von vielen Autoren durchaus als bedeutsam für den Rezeptionsprozess insgesamt erachtet wird (vergleiche Kapitel 3.4.1.1). Ich kam auf den Gedanken, dass dieser schwer greifbare, schimäre Faktor der Typografie eben über eine Befragung durchaus Eingang in mein Forschungsvorhaben finden könne – wenn nicht zur Quantifizierung der „Lesbarkeit“, so eben ergänzend zur Eruierung ästhetischer typografischer Präferenzen der Versuchsteilnehmer.

4.3.7 Mögliches Kriterium guter Lesbarkeit: Lesegeschwindigkeit

Nachdem ich die beschriebenen potenziellen Operationalisierungsmöglichkeiten jenes schimären, zumindest aber vielschichtigen Phänomens „Lesbarkeit“ – nämlich Wissenserwerb, körperliche Beanspruchung, Urteil der Probanden – aus den argumentierten Gründen heraus verworfen hatte, blieb er dann doch als einziger und wohl tauglichster übrig: der Faktor „Lesezeit“ als Operationalisierungskriterium. Es sei zugestanden, dass wohl auch die anderen drei, oben diskutierten Operationalisierungsmodelle in der Realisierung nicht gänzlich ohne die Einbeziehung von Zeitmessungen oder zumindest -beschränkungen ausgekommen wären; diese temporären Bedingungen wären jedoch lediglich Vehikel zur Sicherstellung der Erfassung von Daten anderer Einheit gewesen. Das ist etwas anderes, als die Lesezeit an sich zum Kernkriterium zu erheben!

Und genau dies entschied ich zu tun: Ich würde das Phänomen „Lesbarkeit“ operationalisieren über die Messung von Lesegeschwindigkeiten.

4.3.7.1 Operationalisierbarkeit der Lesegeschwindigkeit

Wie aber misst man zuverlässig das Lesetempo? Zunächst musste ich in jedem Falle von der Annahme ausgehen, dass die Rezeption von Texten im Internet dem hergebrachten Prinzip

folgt: dass Lesende nämlich im Allgemeinen mit der Informationsaufnahme am Textanfang beginnen und sie, linear dem Wort- und Zeilenablauf folgend, mit dem letzten Wort des Textes beenden. Dass viele Nutzer im Web Texte auch gerne und häufiger auf andere Art und Weise konsumieren – scannend, skimmend beispielsweise – wurde bereits beschrieben und bleibt unbestritten. Eine Messung der eigentlichen Lesegeschwindigkeit freilich war definitiv nur unter dieser Prämisse der Unterstellung linearen Lesens überhaupt realisierbar – unter der Annahme des „springenden“ Lesers dagegen praktisch ausgeschlossen. Es galt also – und damit war bereits die erste Vorgabe erkannt, der ich mich zu stellen hatte –, im Experiment Texte vorzulegen, die in inhaltlichem und stilistischem Aufbau so weitgehend wie möglich einen linearen Rezeptionsprozess erzwingen.

Sehr bald erkannte ich auch: Die Messung der Lesegeschwindigkeit beziehungsweise der Lesedauer – das ist die zeitliche Differenz zwischen Lesebeginn und Leseschluss unter der Voraussetzung, dass der Text tatsächlich bis zum Ende rezipiert wird – ist prinzipiell klar definierbar, dafür umso schwieriger zuverlässig operationalisierbar.

Das beginnt bereits beim Akt des Rezeptionsbeginns – salopp formuliert: Wann war die Stoppuhr zu starten? Ein denkbare Modell bestand darin, den Zeitpunkt des Erscheinens eines Textes auf dem Bildschirm ins Ermessen der Probanden zu stellen – beispielsweise durch einen sensitiven Button, der auf Mausclick den Text auf dem Bildschirm erscheinen lässt. Das Erscheinen des Textes auf dem Monitor dann mit dem Beginn des Leseprozesses gleichzusetzen – diese Unterstellung ist dem Ansatz implizit –, ist sicherlich ein etwas heikler Rückschluss. Eine echte Alternative fand ich freilich nicht. Ich beschloss, die Probanden zumindest zu Beginn des Experimentaldurchgangs zu bitten, den Leseprozess so rasch wie möglich nach Vorlage eines Textes zu beginnen.

Noch schwieriger gestaltet sich die Messung des Zeitpunktes, an dem der Leseprozess abgeschlossen ist. Grundsätzlich könnte auch die Bestimmung dieses Datums in die Gewalt des Probanden gelegt werden, wiederum durch einen sensitiven Button, auf dessen Klick der User die Beendigung des Leseprozesses signalisieren und damit messbar gestalten könnte.

In verschiedenen der im vorigen Kapitel beschriebenen Experimentalansätze wurde dieser Weg beschritten⁷⁷⁹. Dies setzt freilich eine „Ehrlichkeit“ der Versuchsteilnehmer voraus, von der nicht ohne weiteres auszugehen war und ist: So könnten Probanden, die sich zwar auf die Teilnahme am Experiment eingelassen, am Erfolg des Experiments aber von vorneherein oder auch nach Betrachten nicht aller Versuchs-Texte desinteressiert sind, den Button jederzeit betätigen, wenn ihnen schlicht die Lust am Durchlesen des angebotenen Textes abgeht. Dies kann dazu führen, dass „Lesezeiten“ registriert werden, die realiter keine sind und die das Ergebnis erheblich verfälschen – in der Regel nach unten, vielleicht auch nach oben.

⁷⁷⁹ Vgl. etwa Weisenmiller 1999; vgl. ferner Morrison/Noyes 2003; vgl. ferner Tullis et al. 1995; vgl. ferner Andreyev/Martynov 2000; vgl. ferner Dyson/Kipping 1998

Und dies, wohlgemerkt, ohne, dass in irgend einer Form abgesichert wäre, dass der Text auch nur annähernd sinnentnehmend rezipiert wurde.

Eine Alternative erwog ich nur kurz und verwarf sie: Denkbar wäre immerhin prinzipiell gewesen, die Teilnehmenden die Texte auf dem Bildschirm schlicht vorlesen zu lassen. Beginn und Abschluss des Leseprozesses wären in diesem Modell einigermaßen zuverlässig registrierbar, also messbar gewesen. Demgegenüber ergaben sich allerdings auch mannigfaltige Schwierigkeiten. Abgesehen davon, dass in einer möglichen Online-Studie dieses Modell wiederum die Präsenz von empirischen Beobachtern notwendig gemacht hätte – und damit logistische Aufwände unabsehbaren Umfangs –, ist lautes Vorlesen von Texten zweifellos eine Verrichtung, der erwachsene Menschen, zumal bei der Rezeption von Webseiten, nur sehr selten nachgehen. Die Lesegeschwindigkeit würde hier also mittels einer Lese-Methodik ermittelt, die wenig alltagsähnlich ist. Dass Vorlesen wohl langsamer vonstatten geht als stummes Lesen, muss grundsätzlich kein Nachteil sein; schließlich sollten im projektierten Experiment ja Unterschiede typografischer Konstellationen ermittelt werden, weniger aber absolute Geschwindigkeitsdaten. Die Tatsache jedoch, dass das korrekte sprechende Verbalisieren einer verschriftlichten Mitteilung noch längst nicht bedeutet, dass der Text auch inhaltlich erfasst wurde, ließ mich rasch auch diese Variante verwerfen.

Mir wurde klar, dass eine reine Lesezeitmessung – in welcher Detailmethodik auch immer – kaum Sinn ergeben würde, wenn nicht gleichzeitig in irgend einer Form sicher gestellt wäre, dass die User sich auch tatsächlich mit den Texten inhaltlich befassten. Es galt also, eine taugliche Mischmethode zu entwickeln, die zwar weiterhin die Zeitmessung als zentrales Element begriff und berücksichtigte, aber komplementär auch einen inhaltserfassenden Rezeptionsprozess der Teilnehmenden gewährleistete.

4.3.7.2 Der erste Pretest

Bereits die vorangegangenen Zeilen zeigen: Es waren noch einige instrumentelle Ungewissheiten aus dem Weg zu räumen, bevor das eigentliche Experiment starten konnte. Um die Tauglichkeit verschiedener Ansätze nicht nur theoretisch abwägen zu können, sondern zumindest in kleinerem Rahmen auch einem ersten Praxistest unterwerfen zu können, programmierte ich Ende August 2005 ein sehr simples, PHP-basiertes und auf einer MySQL-Datenbank aufsetzendes Skript, das bei Aufruf im Webbrowser einen auf HTML-Formulartechnik zurückgreifenden Pretest aktivierte. Dieser Pretest ist nach wie vor abrufbar unter <http://www.j-pr.de/lesen/texte.php>.

Es sei betont, dass dieser Pretest höchstens sehr am Rande bereits aus einem Erkenntnisinteresse bezüglich des Phänomens „Lesbarkeit“ an sich heraus konzipiert war. Vorrangig diente er ausdrücklich der Evaluation der Tauglichkeit bestimmter Erhebungsmethoden und -instru-

mentarien. Daher soll im Folgenden eine Grobbeschreibung dieser Vorstudie ausreichen; auch die eigentlichen Messergebnisse werde ich nur streifen.

Ich band vier Testtexte von jeweils durchschnittlich 104 Wörtern in den Pretest ein; einer befasste sich mit den Waldbränden in Portugal, der zweite mit dem Rechtsstreit der CDU mit den „Rolling Stones“ wegen der Verwendung des Songs „Angie“ im Unions-Wahlkampf, der dritte behandelte wissenschaftliche Erkenntnisse zur Lesegeschwindigkeit, der letzte die Entlassung von Trainer Mario Basler beim Fußball-Regionalligisten FC Regensburg. Die Nutzer wurden aufgefordert, die Texte jeweils mittels Klick auf den Button „Fertig – Weiter“ auf ihren Bildschirm zu projizieren, sie daraufhin sofort, sorgfältig, dabei aber auch möglichst zügig zu lesen und am Ende ihres individuellen Rezeptionsprozesses wiederum den Button „Fertig – Weiter“ anzuklicken. Als „Lesezeit“ wurde die Dauer zwischen Anwahl der beiden Buttons gewertet. Im Anschluss an jeden Text folgten drei Multiple-Choice-Fragen, mittels derer ich den Wissenserwerb der Probanden im vorherigen Text zu quantifizieren und einen sinnerfassenden Rezeptionsprozess sicherzustellen hoffte.

Einzig unabhängige Variable in diesem Pretest war der typografische Faktor „Zeilenbreite“. Per Zufall bekamen die Teilnehmer die Texte in Zeilenbreiten von durchschnittlich 25, 50 oder 75 Anschlägen vorgelegt. Als Schriftart war durchgehend die *Verdana* vorgegeben, der Zeilenabstand nicht explizit definiert, so dass hier der vom Browser voreingestellte Wert (in aller Regel 120 Prozent der Schriftgröße) zum Tragen kam. Auch andere typografische Faktoren wurden nicht explizit vorgegeben.

Ich bat einige Freunde und Bekannte sowie Mitglieder verschiedener Webforen, online – also von zu Hause aus – an diesem Pretest teilzunehmen, und mir vor allem Rückmeldung zu erstatten, inwiefern sie den Test als selbsterklärend, schlüssig und durchführbar erachteten.

Erstaunliche 198 Menschen führten den Test zwischen dem 28.8. und dem 14.9.2005 durch. Dabei erwies sich sehr frühzeitig insbesondere, dass es offenkundig wenig Sinn ergibt, die User in die Lage zu versetzen, den Abschluss ihres Leseprozesses eigenmächtig durch Betätigung von sensitiven Buttons zu signalisieren. So bewegten sich die registrierten „Lesezeiten“ zu über 15 Prozent unterhalb von 5 Sekunden – wohlgerne für Texte, die im Schnitt 104 Wörter aufwiesen. Hier hatten offenkundig viele Teilnehmer schlicht keine Lust, sich der Lektüre zu widmen – und klickten lange vor Ende des individuellen Rezeptionsprozesses, den sie eventuell gar nicht erst begonnen hatten, auf den Button.

Ein weiteres Problem, das im Pretest zutage trat, wurde bereits weiter oben beschrieben: Die Abfrage von Wissens-Erwerb durch den rezipierten Text erbringt höchstens mit Vorsicht zu genießende Daten. Das Vorwissen wurde nicht abgefragt; mehrere Teilnehmer wiesen mich in ihren Rückmeldungen darauf hin, dass sie einige Fragen durchaus auch ohne vorherige Lektüre der Testtexte hätten beantworten können.

Auch räumten einige Teilnehmer ein, dass sie sich durchaus einer virtuellen „Hintertür“ bedient hatten, um die Fragen korrekt zu beantworten: Sie betätigten, nachdem sie die Fragen gesehen hatten, schlicht den „Zurück“-Button ihres Browsers und lasen die Texte nochmals gezielt auf die abgefragten Fakten hin. Diese Umgehungsmethode hatte ich zwar technisch dergestalt aufgefangen, dass sie registriert und der Leseprozess des Teilnehmers daraufhin als „ungültig“ in die Datenbank einging – was freilich auch bedeutete, dass etwa 11 Prozent aller erhobenen Daten als unverwertbar deklariert werden mussten. Eine gefährlich hohe Aussonderungsquote, die ohne die Multiple-Choice-Fragen im Pretest erheblich geringer ausgefallen wäre.

Ein weiteres Problem trat hinzu. Einige Teilnehmer berichteten, sie hätten die korrekte Beantwortung der Fragen als derart zentral betrachtet für das Mess-Ergebnis insgesamt, dass sie die Texte „lieber zwei Mal“ und ausgesprochen sorgfältig lasen. Dass sie auf diese Weise gerade die Messung des zentralen Faktors „Lesegeschwindigkeit“ torpedierten, muss angesichts des offenkundigen guten Willens dieser Teilnehmer als geradezu tragikomisch bezeichnet werden.

Schließlich bemerkten einige Teilnehmer in ihren Rückmeldungen richtigerweise, dass im Test nicht sichergestellt war, dass die Teilnehmer die Schriftgröße „mittel“ in ihrem Browsermenü aktiviert hatten – auf dieser Grundannahme aber basierte die gesamte Programmierung des Tests. Wer nämlich in seinem Browser eine größere Basis-Schriftgröße definiert hatte, bekam deutlich kürzere Zeilen (heißt: Zeilen mit weniger Anschlägen) zu lesen als die anderen Teilnehmer. Ob und wie viele Teilnehmende tatsächlich die Schriftgröße – und damit die Zeilenbreiten – auf diese Weise manipulierten, ist leider bis heute mit keinem Programmskript eruierbar.

Angesichts all dieser Unwägbarkeiten nimmt es wirklich Wunder, dass sich in der groben Endauswertung des Pretests die Zeilenbreite „50“ als im Schnitt um 1,5 Sekunden und damit gut 6 % schneller lesbar erwies als die beiden anderen gemessenen Breiten. Die Unterschiede erwiesen sich in einem t-Test mit den unbereinigten Daten sogar als statistisch leicht signifikant (wobei der t-Test, wie ich noch herausstellen werde, in diesem Zusammenhang eher eine ungeeignete inferenzstatistische Methode darstellte). Wohlgedenkt: Ich nahm dieses Ergebnis zwar nicht als auf wissenschaftlich gesicherter Basis zustande gekommen an – eine gute Portion Mut und Optimismus für die Fortentwicklung des Projekts aber bescherte mir dieser experimentelle Achtungserfolg definitiv. Es konnte weitergehen.

4.3.8 Nachbesserungsbedarf

Die Ergebnisse des Pretests, insbesondere jedoch die Rückmeldungen von Teilnehmenden, brachte es an den Tag: Es gab in Konzeption und programmiertechnischer Hinsicht noch einiges zu tun. Folgende Kernprobleme und Imperative waren erkennbar:

- Die Entscheidung, wann ein Leseprozess im Experiment als beendet zu betrachten ist, musste dem individuellen Ermessen der Nutzer entzogen werden.
- Die Abfrage von durch die Lektüre der Testtexte erworbenem Wissen erfordert zumindest eine Kombination von Vorab- und Nachbefragung, um den Wissenszuwachs wirklich quantifizieren zu können. Ferner muss die Manipulation der Befragung durch die Nutzer (Stichwort: Zurück-Button) ausgeschlossen werden, ohne Daten dabei ungültig werden zu lassen. Schließlich darf das bloße Antizipieren der Fragen die Teilnehmenden nicht dazu bewegen, die Texte übertrieben sorgfältig und damit vermutlich alltagsunähnlich zu lesen.
- Die browserspezifischen typografischen Manipulationsmöglichkeiten der Teilnehmenden mussten auf ein Minimum reduziert werden, um im Experiment vergleichbare und valide Messwerte ermitteln zu können.

Das vorrangige Problem war dabei sicherlich das erstgenannte: Wie war in einem Online-Experiment zuverlässig insbesondere das Ende eines Leseprozesses zu erfassen – ohne diese Bestimmung den Probanden selbst zu überlassen?

Wie im Folgenden zu zeigen sein wird, existieren durchaus brauchbare Instrumentarien zur Lösung dieses Problems. Mehr noch: Es wird deutlich werden, dass sogar Methoden vorliegen, die die Erfassung des Rezeptionsendes mit der Sicherstellung eines sinnerfassenden Leseprozesses zu kombinieren vermögen.

4.3.9 Eingeführte Methoden zur Messung der Lesegeschwindigkeit

Eine genauere Analyse der in der empirischen Bildschirm-Lesbarkeitsforschung verwandten Methoden zeigt, dass die Messung zu überwiegendem Anteil nach recht ähnlichen, wenn auch nicht immer völlig identischen Mustern vorgenommen wird. Der Grundansatz besteht dabei darin, dass die den Probanden vorgelegten Testtexte zielgerichtet an einer oder mehreren Stellen manipuliert werden; hat der Proband diese Manipulationen identifiziert, wird einerseits unterstellt, dass ein sinnentnehmender Rezeptionsprozess stattgefunden hat, und andererseits, dass der Rezeptionsprozess als Ganzer abgeschlossen ist.

Eine ebenso einfache wie offenbar taugliche Methode ist dies offenkundig – eine Methode, die in ihrem Wesenskern zurückgeht auf zwei „Nestoren“ der Lesbarkeitsforschung: die US-amerikanischen Psychologen Donald Paterson und Miles Tinker, deren Lesbarkeits-Studien aus den späten zwanziger und dreißiger Jahren des vergangenen Jahrhunderts bis heute als richtungweisend und Hauptreferenzen für die Beurteilung typografischer Anordnungen gelten. Paterson und Tinker haben sich über Jahrzehnte hinweg intensiv mit der Erforschung von Leseprozessen befasst, weshalb ihre Basisthesen im Folgenden etwas detaillierter vorgestellt werden.

4.3.9.1 Ein „guter“ Lesbarkeitstest: Die Kriterien nach Paterson und Tinker

Paterson und Tinker stellten 1929 in einer ihrer ersten, jedoch wegweisenden Arbeiten zentrale Kriterien auf, die sie inhaltlich und methodisch für eine effiziente Messung von Lesegeschwindigkeiten als unabdingbar bezeichneten⁷⁸⁰. Es sind die folgenden:

- Ausreichende Anzahl von Versuchsteilnehmern

Paterson und Tinker betonen, dass Lesegeschwindigkeiten a priori stark differieren; die Streuung der individuellen Lesetempi verschiedener Menschen sei groß, größer in jedem Falle als die Durchschnittsdifferenzen, durch die sich die Lesbarkeiten verschiedener typografischer Anordnungen a priori voneinander abgrenzten. Um signifikante Ergebnisse zu erhalten, müsse also eine hinreichend große Anzahl Probanden akquiriert werden.

- Schwierigkeitsgrad der vorgelegten Texte

Der Schwierigkeitsgrad der vorgelegten Texte muss nach Paterson und Tinker konstant bleiben, um eine Vergleichbarkeit auch unter Vergleichsgruppen mit verschiedenen Probanden als Angehörigen herstellen zu können.

- Gleichwertigkeit verschiedener vorgelegter Texte

Werden innerhalb einer Studie mehrere Texte nacheinander vorgelegt, so müssen diese Texte in ihrer Inhaltsgestaltung „Trainingseffekten“ Rechnung tragen. Diese Trainingseffekte werden vermutet, wenn Probanden eine Aufgabenform verinnerlicht und somit im Verlaufe des Experimentalverlaufs prinzipiell gleichwertige Texte immer schneller zu bearbeiten in der Lage sind. Diese Trainingseffekte sollten laut Paterson und Tinker durch eine sukzessive Steigerung der Schwierigkeitsgrade der Texte im Studienverlauf aufgefangen werden.

- Sicherstellung sorgfältigen Lesens

Paterson und Tinker betonen, dass Lesegeschwindigkeiten nur untereinander vergleichbar seien, wenn überprüfbar sei, dass alle Texte mit vergleichbarer Gründlichkeit und Sorgfalt gelesen worden seien.

- Sicherstellung eines „natürlichen“ Leseprozesses

Die Gründlichkeit des Leseprozesses muss, wie angesprochen, ein wichtiges Kriterium sein. Das Bemühen der Probanden um diese Gründlichkeit darf aber nicht zu einer Verfälschung des Ergebnisses führen. Sie müssen immer noch so alltagsähnlich wie möglich lesen.

- Länge der Texte

Die vorgelegten Texte müssen laut Paterson und Tinker eine ausreichende Länge aufweisen, um die Lesezeiten zuverlässig vergleichen zu können (was hierbei unter „ausreichend“ zu verstehen ist, wird noch zu darzustellen sein).

⁷⁸⁰ Vgl. Paterson/Tinker 1929, S. 359ff.

4.3.9.2 Die Messmethodik nach Tinker und Paterson

Paterson und Tinker beschrieben 1936 ihr standardisiertes Verfahren zur Messung der Lesegeschwindigkeiten von Probanden, das ihren oben referierten Kriterien Rechnung trug und das sie in den Vorjahren erfolgreich in verschiedensten Studien angewandt hatten⁷⁸¹.

Der Tinker-/Paterson-Test basiert auf 450 Absätzen beziehungsweise Einzelaufgaben; jeder dieser Absätze besteht aus 30 relativ einfachen, allgemein als bekannt vorausgesetzten Wörtern, die einen bis zwei gut verständliche Sätze formen. Erzählt werden in diesen Texten eher belanglose „Alltagsgeschichten“, die bewusst emotionale Reize nicht zu wecken versuchen, um einen möglichst „neutralen“ Rezeptionsprozess sicherzustellen.

Um zudem einen konzentrierten Lesevorgang ihrer Versuchsteilnehmenden zu induzieren, installierten Paterson und Tinker im jeweils zweiten Teil eines jeden Absatzes ein Wort, das syntaktisch wie semantisch korrekt platziert ist, inhaltlich aber im Sinn-Gegensatz zum ersten Teil des Absatzes steht. Die Probanden werden nach der Methodik auf diese Tatsache hingewiesen und angewiesen, das jeweils sinnfremde Wort zu identifizieren, indem sie es durchstreichen. Die Zeit, die Probanden für die Bewältigung dieser Aufgabe benötigen, wird gemessen und als abhängige Variable in die Auswertung einbezogen.

Alle Teilnehmenden erhalten im Tinker-/Paterson-Modell zu Beginn eines Gesamt-Experimentaldurchgangs identische Textabsätze in identischer typografischer Konstellation vorgelegt. Im zweiten Teil wird dann ein typografischer Faktor variiert, und die Teilnehmenden setzen den Durchgang mit inhaltlich andersartigen Texten fort. Aus den individuellen Geschwindigkeitsunterschieden, die sie bei den einzelnen Teilnehmenden konstatierten und in Durchschnitts-Abweichungen ausdrückten, folgerten Paterson und Tinker in ihren Experimenten typografische Einflüsse auf die Lesegeschwindigkeiten.

Wendt übertrug das Versuchsdesign von Paterson und Tinker in den 1960er Jahren ins Deutsche, wobei er verschiedene Ursprungstexte von Tinker und Paterson leichten Änderungen und zeitgemäßen Anpassungen unterwarf. Ein Text-Beispiel aus seinem Design:

„Ilse putzte am Samstagmorgen das ganze Haus, weil sie ihre Mutter damit überraschen wollte. Als Frau Piper vom Einkaufen nach Hause kam, fand sie das Auto sauber und glänzend“⁷⁸² (hier war also das Wort „Auto“ zu identifizieren). Mit seiner deutschen Variante führte Wendt auch mehrere Print-Experimente durch, wobei er allerdings – im Gegensatz zu Tinker und Paterson – mehrere typografische Abwandlungen testete und nicht nur eine. Ich werde später noch detaillierter auf seine modifizierte Gruppeneinteilung eingehen.

⁷⁸¹ Tinker/Paterson 1936, S. 132ff.

⁷⁸² Vgl. Wendt 1992, S. 271

4.3.9.3 Modifikationen der Paterson-/Tinker-Methode

Die Messmethode von Paterson und Tinker wurde seit den 1930er Jahren von vielen Forschern im Grundsatz übernommen. Auch Lesbarkeitsstudien am Bildschirm wurden bereits nach wesentlichen Prinzipien der US-Amerikaner konzipiert. Allerdings sind auf Grundlage der Ansätze von Tinker und Paterson durchaus auch sinnvolle Modifikationen der Methodik denkbar. Ich stelle im Folgenden zwei interessante, dem Paterson/Tinker-Modell ähnliche, im Detail aber andere Wege beschreitende Modelle von Lesbarkeitstests vor.

- „Grammatikalische Methode“

Bernard und Mills setzten in einer Bildschirm-Studie 2000 auf das Prinzip, orthografische Fehler in den Testtexten unterzubringen. „Um sicherzustellen, dass die Teilnehmer die Textabschnitte lasen, wurden zufällig zehn Substitutionswörter in jedem Abschnitt platziert. Die Teilnehmer wurden angewiesen, diese Wörter durch lautes Aussprechen zu identifizieren. Die Substitutionswörter waren so konzipiert, dass sie bei sorgfältigem Lesen als unpassend zum Kontext des Abschnitts erkannt werden konnten. Die Wörter variierten grammatisch von den Originalwörtern – so wurde beispielsweise das Nomen ‚sun‘ durch das Adjektiv ‚fun‘ ersetzt“⁷⁸³.

Dieser Ansatz von Bernard und Mills ist nicht uninteressant, da er in gewisser Hinsicht noch kritischer die inhaltsentnehmende Rezeption von Texten durch Probanden quantitativ wie qualitativ kontrolliert als der Uransatz Tinkers und Patersons. Der Ansatz unterscheidet sich vom Ursprungsmodell insbesondere dadurch, dass er die Lesegenauigkeit der Probanden nicht nur inhaltlich-kontextuell misst, sondern zusätzlich die Komponente der grammatischen Texterfassung addiert.

Zusätzlich interessant ist an Bernards und Mills' Ansatz, dass die Forscher – im Gegensatz zum Modell Tinkers und Patersons – nicht nur ein einziges zu identifizierendes Wort je Textabschnitt vorsehen, sondern deren zehn pro Absatz. Dieser Ansatz birgt die Chance, die Sorgfalt eines Rezeptionsprozesses deutlich differenzierter zu bewerten: Wo bei Tinker und Paterson nur die Kategorien „gefunden – nicht gefunden“ existieren, wäre in Bernards und Mills' Ansatz auch denkbar, beispielsweise bereits das Auffinden von acht der zehn Substitutionswörter als Indiz für einen erfolgreich abgeschlossenen, inhaltsentnehmenden Leseprozess zu interpretieren. Eine kurze Phase der Unachtsamkeit des Probanden kann also im Urmodell von Tinker und Paterson bereits die empirische Klassifizierung des gesamten Testprozesses als „gescheitert“ bewirken, im „Orthografie-Modell“ dagegen wäre beispielsweise das Überlesen eines oder zweier Wörter immer noch guten Gewissens als lässliche Ausfallerscheinung bewertbar, die nicht gegen einen gelungenen Rezeptionsvorgang insgesamt spricht.

⁷⁸³ Vgl. Bernard/Mills 2000 (Übersetzung M.L.)

- Stolperwörter

Begriff und Prinzip des „Stolperworts“ gehen auf den Berliner Pädagogen Wilfried Metze zurück. Metze hat 2003 einen sehr einfachen Test entwickelt, der Grundschul-Lehrern der Klassen 1 bis 4 Aufschlüsse über den Lese-Lernstand ihrer Klassen verschaffen soll⁷⁸⁴ – Metztes Ansatz ist mithin kein originär typografischer, sondern ein pädagogischer. Dies enthebt ihn aber meines Erachtens keineswegs der Betrachtungswürdigkeit.

Wilfried Metze, dies sei kurz erläuternd eingeschoben, verfolgte und verfolgt bei der Konzeption und Publikation des Tests augenscheinlich keine kommerziellen Interessen: Die Testbögen sind als PDF-Dateien von seiner Homepage herunterladbar, als Gegenleistung verlangt Metze lediglich, dass Lehrer die Ergebnisse, die sie in ihren Klassen mit dem „Stolperwort-Test“ erzielen, zur Verfügung stellen und damit der Einarbeitung in eine umfangreiche statistische Auswertung zustimmen – eine Bedingung freilich, deren Einhaltung er augenscheinlich nicht dezidiert überwacht. Publikationen zum Stolperwort-Prinzip liegen im Übrigen, jenseits der Website, zumindest von Metze selbst nicht vor (dies bestätigte er mir am 27.10.2005 per E-Mail ausdrücklich).

Was unterscheidet Metztes Prinzip von den vorgenannten? Vor allem dies: Es basiert zwar im Grundsatz – darin gleicht es den obigen Modellen – ebenfalls auf gezielt platzierten Wörtern als zu identifizierenden „Fremdkörpern“ in Texten. Metztes „Stolperwörter“ jedoch korrumpieren weitaus eindeutiger und frappierender den erzählerischen Fluss der textlichen Information, als dies bei Tinker und Paterson der Fall ist, da sie die Textlogik nicht durch textimmanenten inhaltlichen Widerspruch brechen, sondern durch bewusste syntaktische Verstöße unter Beibehaltung einer – wenn auch randständigen – Inhaltsstringenz. Diese rudimentäre Stringenz ist es auch, die Metztes „Stolperwörter“ von Bernards und Mills’ „Substitutionswörtern“ abgrenzt: Diese nämlich büßen durch die gezielte grammatikalische Manipulation jeden Inhaltszusammenhang ein. Metztes Ansatz umgeht dieses Problem, indem seine „Stolperwörter“ nicht substitutiv in die Ursprungstexte eingearbeitet werden, sondern komplementär; sie stehen damit dem logischen Textfluss auch physisch schlicht „im Wege“, lassen damit den einigermaßen aufmerksamen Leser „stolpern“. Metze selbst nennt folgende Beispiele zur Veranschaulichung seines Konzepts:

- „1 Mir gefällt dein schön Bild gut.
- 2 Im Winter eisig ist es oft kalt.
- 3 Mein Heft ist schreiben voll.
- 4 Mein Bruder ist gestern acht Jahre alt.
- 5 Der Hund bellt Wut.“

Bereits das erste dieser Beispiele illustriert Metztes Ansatz, wie ich meine, sehr treffend. Das Stolperwort „schön“ nämlich kann bei flüchtigem Lesen durchaus in die logische Stringenz

⁷⁸⁴ Vgl. Metze O.J.

des Satzes eingeordnet werden; dass es dagegen syntaktisch fehlplatziert ist, erschließt sich nur bei sorgfältigerem Lesen. In diesem Sinne halte ich Metzses Ansatz für eine durchaus gelungene Symbiose der beiden vorgenannten Ansätze.

4.3.9.4 Ein weiterer Ansatz zur Messung der Lesegeschwindigkeit

Ein weiterer methodischer Ansatz zur Messung von Rezeptionsgeschwindigkeiten fand bereits weiter oben kurze Erwähnung; Weisenmiller verwandte ihn in seiner bereits erwähnten Untersuchung: den „Nelson-Denny-Test“⁷⁸⁵.

Weisenmiller forderte seine Versuchsteilnehmenden zu einem fixen Zeitpunkt auf, mit dem Lesen der Versuchstexte zu beginnen; nach einer Minute wies er sie dann mündlich an, das letzte Wort zu markieren, das sie bis zum verbalen Stopp-Signal gelesen hatten. Aus dem Quotienten von gelesener Wortanzahl und Lesezeit schloss Weisenmiller dann auf die individuell erreichte Geschwindigkeit.

Problematisch an dieser Herangehensweise ist, dass Weisenmiller zumindest in Ansätzen praktizierte, was ich im ersten Pretest als eher untaugliches Prinzip erkannt hatte: Er legte den Messvorgang zu weiten Teilen in die Hände der Probanden. Damit aber nahm er notwendigerweise die Gefahr von Manipulationen durch die Teilnehmenden in Kauf, die ja im Prinzip frei waren, jedes beliebige Wort im Text als Letztgelesenes zu markieren – Weisenmiller spekulierte also auf einen hohen Mitwirkungswillen seiner Probanden. Ob Manipulationen auftraten – seien es Manipulationen aus destruktivem Antrieb heraus, seien es Manipulationen aus dem Wunsch nach Lieferung von Ergebnissen im Sinne der „Erwünschtheit“ – eruierte Weisenmiller nicht aktiv. Zwar kontrollierte er im Anschluss an den Leseprozess auch erinnerte Textinhalte in Form einer Befragung; doch auch dieses Instrument hatte ich, wie gezeigt, nur bedingt als tauglich erkannt im Rahmen meines Pretests.

4.3.10 Entscheidung für eine „Stolperwort“-Studie

Nach Abwägung der verschiedenen vorgestellten methodischen Modelle zur Messung von Lesegeschwindigkeiten entschied ich mich relativ rasch dafür, die von mir beabsichtigte Untersuchung nach dem Muster von Metzses Konzept der „Stolperwörter“ durchzuführen. Die Gründe seien im Folgenden erläutert.

Der Reiz des Stolperwörter-Prinzips liegt meines Erachtens insbesondere darin begründet,

- dass es Geschwindigkeitserfassung und Sicherstellung eines inhaltsorientierten Rezeptionsprozesses methodisch koppelt. Damit erübrigte sich auch und insbesondere die vor-

⁷⁸⁵ Weisenmiller 1999, S. 54ff.

und/oder nachgeschaltete Erhebung eventueller Lernzuwächse durch die vorgelegten Texte, die, wie gezeigt, in der Umsetzung schwierig und in methodischer Hinsicht fragwürdig als zuverlässige Datenquelle ist.

- dass es im Vergleich zu den verwandten Modellen keine Höchst-, sondern eine eher durchschnittliche physisch-psychische Konzentration im Rezeptionsprozess von potenziell Teilnehmenden verlangt; ich unterstelle aber, dass dies insbesondere im Modell von Bernard und Mills nicht unbedingt der Fall ist. Deren sehr distinkte Manipulationen von Wörtern – „sun“ zu „fun“ – ist meines Erachtens als Prinzip nicht unbedenklich, fordert es doch einen Rezeptionsprozess in einem Maße von Fokussiertheit ein, der dem Alltags-Lesen eher unähnlich ist. Der von mir unterstellte, eher mittlere Grad an eingeforderter Leseintensität in Metzkes „Stolperwort-Modell“ dagegen scheint mir dem verbreiteten, gleichsam „web-alltäglichen“ Leseverhalten eher verwandt.
- dass die Stolperwörter in Schwierigkeitsgrad und Anzahl leicht variierbar sind; Tinkers und Patersons Modell des einen „Schlüsselworts“ im jeweils zweiten Teilabsatz zeigt sich demgegenüber quantitativ und qualitativ deutlich unflexibler.
- dass die Stolperwörter in den Testtexten inhaltlich mehr oder weniger eng an den Textinhalt angelehnt werden können. Auf diese Weise erzwingen sie beim Teilnehmenden nicht nur einen eindimensional auf Fehlersuche orientierten Suchprozess, der auch durch grobes „Überfliegen“ des Textes realisierbar wäre und wie ihn etwa Bernards und Mills' Experimentaldesign ansatzweise induzieren könnte.

Bestehend ist das Prinzip der Stolperwörter jedoch vor allem und nicht zuletzt auf Grund seiner Simplität und leichten Begreifbarkeit. Ein bis zwei Beispiele wie die oben angeführten würden, so meine Erwartung, in aller Regel zu Testbeginn ausreichen, um den Teilnehmerinnen und Teilnehmern einer Untersuchung das Prinzip zu verdeutlichen; im weiteren Testverlauf würde dann als Arbeitsanweisung ein schlichtes „Finden Sie die Stolperwörter!“ genügen. Für ein Experiment wie das von mir geplante, das spätestens als Online-Experiment so einfach wie irgend möglich, am besten „selbst erklärend“ gehalten werden müsste, ist das ein höchst gewichtiges Argument.

Metzke selbst zeigte sich im Übrigen beeindruckt von meinem Vorhaben (er war Teilnehmer an einem der Pretests im Oktober 2005), bezeichnete es als „sehr gute Adaption“ seines Basis-Ansatzes und erteilte mir ausdrücklich den Segen dazu, sein Stolperwort-Prinzip in seinen Wesenselementen für mein Experiment zu übernehmen (dies ebenfalls per Mail vom 27.10.2005).

4.4 Methode der Wahl: experimentelle Beobachtung

Die Abwägung und Auswahl der empirischen Erhebungsmethode war für mein projektiertes Untersuchungsvorhaben recht einfach und daher rasch vollzogen – spätestens nachdem die

Entscheidung für einen Stolperwort-Test nach Grundzügen der Paterson/Tinker-Messmethodik gefallen war. Denn für die Messung des Faktors „Lesbarkeit“ schien damit nur eine einzige der klassischen empirischen Erhebungsmethoden wirklich tauglich: die experimentelle Beobachtung. Ich sah und sehe bis heute die Entscheidung für dieses Erhebungsverfahren als einzig zielführende an – nicht zuletzt im Abgleich der gewählten Methode in ihrer vermuteten Tauglichkeit mit der der übrigen eingeführten Methoden der empirischen Sozialforschung:

- Die Erhebungsmethode „Befragung“ vermöchte höchstens die Meinung oder Vermutung von Probanden bezüglich der Lesbarkeit einer typografischen Anordnung zu erfassen, jedoch nicht die Lesbarkeit als solche. Dass eine Befragung als dann ergänzendes Erhebungsinstrument im projektierten Experiment durchaus Sinn ergeben würde (und auch zur Anwendung kam), ändert nichts an der Grunderkenntnis: Lesbarkeit sollte im hier diskutierten Vorhaben forschend objektiviert werden – und eine Befragung war dafür kein geeignetes Mittel der Wahl.
- Ähnliches gilt für die Methode der „Inhaltsanalyse“. Zwar ließe sich durchaus spekulieren, dass beispielsweise die Zugriffsraten auf bestimmte Webseiten in ihrer Quantität durchaus beeinflusst sind von der Qualität der Typografie auf der jeweiligen Seite – einen unmittelbaren und vor allem monokausalen Zusammenhang zu unterstellen zwischen der Typografie einer Webseite und dem Grad des Publikumszuspruchs erscheint jedoch, gelinde gesagt, gewagt. Nicht nur beeinflussen auf den allermeisten Webseiten auch bildliche und farbflächige Elemente die Wahrnehmung – an dieser Stelle lohnt der Hinweis, dass eine vermutet überwältigende Mehrheit von Online-Nutzern Seiten wegen ihrer Inhalte aufsucht und nicht wegen ihrer gestalterisch-typografischen Eigenqualität.
- eine nicht-experimentelle Beobachtung schließlich hätte insbesondere unabsehbare Probleme bezüglich der verlässlichen Datenerfassung mit sich gebracht. Zwar wäre es durchaus denkbar gewesen, Probanden bei der „alltäglichen“ Rezeption real existierender texthaltiger Webseiten zu beobachten – sei es beispielsweise durch Installierung von „Beobachtern“, sei es durch filmischen Mitschnitt⁷⁸⁶, mithilfe digitaler Erfassung der Mausbewegungen oder über Messung von Verweildauern auf bestimmten Seiten. Die (wie auch immer erfassten) Daten wären aber in diesen Modellen nur sehr schwer von ablenkenden Elementen auf der betreffenden Seite wie insbesondere Bildern, Navigationselementen oder Werbebannern zu isolieren und damit in unmittelbaren Bezug setzbar gewesen mit der eigentlichen Lese-Performance – dies widerspricht der Forderung, dass „die Versuchsanordnung so angelegt sein (muss), dass eine einzige verursachende Variable isoliert werden kann. Diese isolierte Variable wird als unabhängige Variable bezeichnet und stellt die Ursache dar, auf die eine beobachtbare Veränderung zurückgeführt werden kann“⁷⁸⁷. Dass die Installation von Beobachtern zusätzlich erhebliche logistische (und damit wohl auch finanzielle) Auswirkungen auf die

⁷⁸⁶ Vgl. Kromrey 2006, S. 348

⁷⁸⁷ Klammer 2005, S. 274

Durchführung und damit vor allem auf die potenzielle Anzahl akquirierbarer Teilnehmer genommen hätte, sei nur der Vollständigkeit halber erwähnt.

4.4.1 Das Experiment als empirisches Forschungsinstrument

Es gilt zunächst herauszustellen, dass ein „Experiment“, genau betrachtet, keine eigenständige wissenschaftlich-empirische Erhebungsmethode ist, die also im Range gleichwertig neben den Methoden Befragung, Inhaltsanalyse oder Beobachtung stünde⁷⁸⁸. Eher ist das Experiment zu charakterisieren als eine spezifische Versuchsanordnung, innerhalb derer die Daten allerdings wiederum durch Beobachtung oder Befragung erhoben werden⁷⁸⁹.

Laut Klammer kann ein (sozial-)wissenschaftliches Experiment „allgemein definiert werden als eine wiederholbare Beobachtung, die unter kontrollierten Bedingungen stattfindet, wobei einzelne Variablen der Versuchsanordnung so manipuliert werden, dass ein vermuteter Kausalzusammenhang überprüft werden kann“⁷⁹⁰ – mithin als „Herstellung bestimmter Bedingungen in der Welt mit dem Ziel, bestimmte vermutete Sachverhalte beobachtbar zu machen“⁷⁹¹. Probanden werden im Experiment eindeutig festgesetzten Stimuli (sogenannten „unabhängigen Variablen“) ausgesetzt, ihre Reaktion auf diese Stimuli nach einem definierten Prinzip gemessen (der Messwert ist dann die „abhängige Variable“)⁷⁹². Im Idealfall lässt sich aus einem Experiment ein eindeutiges Ursache-Wirkungs-Schema deduzieren⁷⁹³.

Im projektierten Experiment ergab sich die Art der Stimuli im Prinzip von selbst: Es sollten Probanden mit verschiedenen typografischen Anordnungen in einem Browser konfrontiert werden. Die Bestimmung der unabhängigen Variablen im Detail dagegen galt es noch zu entwickeln. Ich werde später eingehender auf dieses Problem zu sprechen kommen.

Es sei an dieser Stelle erwähnt und abschließend festgestellt, dass ich im Folgenden vorrangig konzeptionelle Aspekte des Experimentaldesigns behandeln werde – insbesondere ethische Aspekte der Experimentkonzipierung, wie sie auch in jüngerer Zeit bezüglich medizinischen Versuche diskutiert wurden, finden ausdrücklich keine Berücksichtigung, da ich sie für das projektierte Experiment als nicht oder höchstens marginal relevant erachtete⁷⁹⁴.

⁷⁸⁸ Vgl. Brosius/Koschel 2005, S.176f.

⁷⁸⁹ Vgl. Eichhorn 1993, S. 163

⁷⁹⁰ Klammer 2005, S. 273f.

⁷⁹¹ Hofstätter/Wendt 1974, S. 3

⁷⁹² Vgl. Zimbardo/Gerrig 2004, S. 31; vgl. ferner Brosius/Koschel 2005, S.177

⁷⁹³ Vgl. Kromrey 2006, S. 97

⁷⁹⁴ Vgl. Brosius/Koschel 2005, S.184; vgl. ferner Klammer 2005, S. 280; vgl. ferner Zimbardo/Gerrig 2004, S. 45f.

4.4.2 Deskription, Exploration oder Hypothesenprüfung?

Experimentelle Beobachtungen können, wie alle anderen empirischen Erhebungsmethoden auch, verschiedene Ziele verfolgen.

- Deskriptive Forschung dient dabei dazu, „Informationen über bestimmte Realitätsbereiche zu erfassen“⁷⁹⁵. Dieser „Realitätsbereich“ kann beispielsweise ein Meinungsbild sein, etwa die Mehrheitsmeinung der Bürger der Bundesrepublik Deutschland zu der Frage, welches politische Problem derzeit das dringendste im Lande sei. So wäre womöglich ermittelbar, dass die Bundesbürger die Arbeitslosigkeit für eben das vorrangige Problem ihrer Zeit halten. Warum dies freilich so ist, ist im Rahmen eines deskriptiven Forschungsansatzes in aller Regel nicht eruierbar.
- An diesem Punkt setzen explorative Forschungsvorhaben an. Hier wird untersucht mit dem Ziel, herauszufinden, „wo mögliche Ursachen für ein Phänomen liegen könnten, ohne dass wir über theoretische Vorannahmen verfügen“⁷⁹⁶. Die Daten werden also mehr oder minder „auf gut Glück“ erhoben, in der Hoffnung, aus den erfassten Daten Erkläransätze für ein sozialwissenschaftliches Phänomen induktiv extrahieren zu können.
- Hypothesenprüfende Forschung schließlich versucht festzustellen, „ob sich Vorhersagen, die im Rahmen einer Theorie gemacht werden, empirisch nachweisen lassen“⁷⁹⁷.

Grundsätzlich gilt für natur- wie sozialwissenschaftliche empirische Forschung, dass diese, wenn möglich, theoriegeleitet sein sollte. Ein hypothesenprüfender Forschungsansatz gilt mithin allgemein als wünschenswert⁷⁹⁸. Für das von mir projektierte Vorhaben ergaben sich in dieser Hinsicht jedoch einige Schwierigkeiten.

Zwar existieren für einige typografische Phänomene in Webbrowsern durchaus Annahmen, die zu überprüfaren Hypothesen ausarbeitbar und mithin in einer empirischen Studie überprüfbar wären. Dies gilt beispielsweise für die erwähnte Vermutung, dass die starke „Verpixe- lung“ von Buchstaben auf dem Bildschirm zu einer schlechteren Lesbarkeit von serifentragenden und in den Strichstärken ausdifferenzierteren Schriftarten im Vergleich zu „Grotesken“ führt. Ziel des projektierten Experiments war jedoch, wie erwähnt, nicht zuletzt auch die Erforschung wechselseitiger Interaktion typografischer Faktoren in verschiedensten Ausprägungen; die Literatur-Empfehlungen wie die wenigen vorliegenden empirischen Erkenntnisse gerade in diesem Problemfeld erschienen jedoch qualitativ wie quantitativ zu dünn, meist theoretisch kaum fundiert und vor allem allzu oft als derart krass widersprüchlich, dass sie als Grundlagen begründeter Hypothesen ausschieden. Meines Erachtens wären also fixe Hypothesen für das beabsichtigte Experiment notwendigerweise so hoch spekulativ ausgefal-

⁷⁹⁵ Vgl. Eichhorn 1993, S. 163

⁷⁹⁶ Eichhorn 1993, S. 163

⁷⁹⁷ Eichhorn 1993, S. 164

⁷⁹⁸ Vgl. Kromrey 2006, S. 52f.

len, dass sie letztlich nur der Verbrämung eines de facto explorativen als vermeintlich hypothesenprüfenden Forschungsansatz gedient hätten.

Ich entschied mich daher für die meines Erachtens „ehrlichere“ Variante: das projektierte Experiment würde einem explorativen Forschungsansatz folgen. Und damit auch jener hoch-simplen „Nullhypothese“⁷⁹⁹, dass typografische Manipulationen von Texten keinerlei Einfluss auf die Lesegeschwindigkeit ausüben – so erwünscht die Widerlegung dieser These war, bot sie insbesondere die Chance, Daten zunächst einmal sehr unvoreingenommen zu erfassen und sie dann einer eingehenderen Analyse zu unterwerfen.

4.4.3 Labor- versus Feldbeobachtung

Eine weitere Frage, die im Vorfeld des Experiments zu klären war, war die nach dem zu konstruierenden Versuchsumfeld. Probanden würden mit typografisch experimentell gestalteten Texten konfrontiert werden, so weit war das Vorhaben längst beschrieben – aber an welchem Ort dies geschehen sollte, war ungeklärt. In einem Versuchsraum, in den dann alle potenziell Teilnehmenden zu bitten wären, um dort das Experiment durchzuführen? Oder am heimischen, am Büro- oder PC-Pool-Rechner der Hochschule, mithin in einem Umfeld, in dem die Probanden auch im Alltag Webseiten konsumieren? Kurzum: Würde eine „Laborstudie“ eher zum Ziel führen oder eine „Feldstudie“?

Beide Varianten haben ihre Vor- und Nachteile. Im Kern machen sich die Unterschiede an den sogenannten „Kontrollbedingungen“ fest. Schließlich ist grundsätzlich Absicht und Ziel einer experimentellen Beobachtung, vermutete Einflussfaktoren auf menschliches Verhalten – in meinem Vorhaben: typografische Faktoren – gezielt zu isolieren und durch Variierung ihrer Ausprägungen ihre alleinursächliche Wirkung auf Verhaltensmuster zu identifizieren⁸⁰⁰. Im Idealfall versucht der empirische Forscher also, „alle Variablen und Bedingungen konstant zu halten, bis auf diejenigen, die in direktem Zusammenhang mit der zu testenden Hypothese stehen“⁸⁰¹. Wer also – wie ich – den Einfluss typografischer Variationen auf die Lesbarkeit von Webtexten experimentell untersuchen möchte, muss in diesem Verständnis beispielsweise dafür sorgen, dass alle Teilnehmenden das Experiment durchführen

- an identischen Bildschirmen
- an Rechnern identischer Leistung
- in identischen Browsern
- auf identischen Betriebssystemen
- im selben Raum
- bei identischer Raumtemperatur

⁷⁹⁹ Vgl. Hofstätter/Wendt 1974, S. 6f.

⁸⁰⁰ Vgl. Kromrey 2006, S. 97

⁸⁰¹ Zimbardo/Gerrig 2004, S. 34

- und identischen Lichtbedingungen
- zur identischen Tageszeit
- bei identischer Ausgeruhtheit und Motivation der Teilnehmenden.

Die obige Liste ist zweifellos unvollständig (was sie keineswegs weniger respekt einflößend geraten lässt). Im strengen Verständnis qualifiziert dabei schon jede kleinste Variation dieser nicht unmittelbar messrelevanten Variablen als „Störvariablen“⁸⁰². Eichhorn nennt daher auch als zentrales Apodiktum, der experimentelle Forscher habe „eine Situation (zu) schaffen, in der nur die Kräfte, deren Effekte er messen will, Wirkung auf das zu untersuchende Phänomen ausüben können, und andere Faktoren, die ebenfalls Einfluss haben könnten, ausgeschaltet sind“⁸⁰³.

Es liegt auf der Hand, dass der Ausschluss von Störvariablen – beziehungsweise deren Anerkennung und folgerichtigen Konstanthaltung im Untersuchungsverlauf – stets nur näherungsweise realisierbar ist. Störvariablen sind in der Sozialwissenschaft stets präsent und widerwillig akzeptierte Begleiter: „Menschliches Handeln, das ja Gegenstand experimenteller Untersuchungen ist, erfolgt nicht immer rational und lässt sich selten auf einfache Ursache-Wirkung-Zusammenhänge reduzieren“⁸⁰⁴. Wird dennoch versucht, die Umgebungsbedingungen einer Untersuchung konstant zu halten, so ist dies prinzipiell nur möglich, wenn ein kontrollierter Rahmen geschaffen wird; es wird eben ein „Laborexperiment“ konstruiert. Insbesondere heißt dies in aller Regel, dass alle Probanden das Experiment am selben Ort durchführen.

Nun „liegt auf der Hand, dass die künstliche Atmosphäre eines Laborexperiments (...) fast allen realen, sozialen Situationen (...) widerspricht“⁸⁰⁵, deren Einflusszusammenhänge ja gerade interessieren im Rahmen von experimentellen Beobachtungen: „Die Kontrolle sämtlicher möglicher Einflussfaktoren und die Standardisierung des experimentellen Ablaufs können dahin führen, dass das Experiment unter völlig unrealistischen Bedingungen abläuft und damit nicht mehr sicher gestellt ist, dass man die Ergebnisse überhaupt auf die Realität übertragen kann“⁸⁰⁶. Die Probanden werden in eine eben nicht alltägliche Umgebung platziert, um Alltagsverhalten zu erforschen⁸⁰⁷. Es spricht aus diesem Grunde viel für die Gattung der „Feldbeobachtung“, die Beobachtung der Probanden also in einem vermeintlich „natürlichen“, alltäglichen räumlichen und sozialen Umfeld. Die Kehrseite freilich ist evident: „In der Künstlichkeit des Labors lassen sich Ursache-Wirkung-Zusammenhänge gut herausarbeiten und nachweisen, eine Generalisierung dieser Ergebnisse ist wegen dieser künstlichen Rahmenbedingungen jedoch nur schwer möglich. Genau umgekehrt ist es bei Feldexperimenten: Sie

⁸⁰² Vgl. Brosius/Koschel 2005, S.180f.; vgl. ferner Klammer 2005, S. 276

⁸⁰³ Eichhorn 1993, S. 167

⁸⁰⁴ Klammer 2005, S. 276

⁸⁰⁵ Klammer 2005, S. 279

⁸⁰⁶ Brosius/Koschel 2005, S.182

⁸⁰⁷ Vgl. Kromrey 2006, S. 100

zeichnen sich durch eine hohe Authentizität der Untersuchungssituation aus, verbunden leider mit all den dazugehörigen intervenierenden Variablen, die die Komplexität sozialer Wirklichkeit nun einmal enthält und die einen ungestörten Blick auf Ursache und Wirkung unterbinden“⁸⁰⁸.

Zimbardo und Gerrig halten Beobachtungen unter natürlichen Bedingungen besonders in den Anfangsstadien einer Untersuchung für nützlich. Sie hülften, „den Verbreitungsgrad eines Phänomens zu erfassen oder einen Eindruck davon zu bekommen, welche Variablen und Zusammenhänge wichtig sein könnten“⁸⁰⁹. Nun kann die Lesbarkeit von Web-Typografie sicherlich nicht als gänzlich unerforscht gelten; mein Vorhaben freilich, auch wechselseitige Interaktionen typografischer Faktoren im Web experimentell zu untersuchen, darf nach meinen Recherchen durchaus den Rang eines Pionierprojekts beanspruchen.

Bereits weiter oben habe ich erwähnt, dass das Konzept einer Online-Studie bereits in einem frühen Stadium der Konzeptionsphase meines Untersuchungsvorhabens eine wichtige Rolle spielte als methodische Alternative. Im Sinne der zitierten Literatur ist „die“ Online-Studie wohl grundsätzlich eindeutig als Variante des Konzepts des „Feldexperiments“ zu verstehen. Da ich der Literatur allerdings auch keine durchschlagenden Argumente gegen die Methode des Feldexperiments entnehmen konnte, entschloss ich mich, Konzept und Erkenntnisaussichten einer Online-Feldstudie detaillierter unter die Lupe zu nehmen.

4.4.4 Das Online-Experiment als spezielle Form der Feldbeobachtung

Das von mir konzipierte Experiment zielte auf die Erforschung der Lesbarkeit verschiedener typografischer Anordnungen in Webbrowsern – was also lag näher, als das Experiment dann auch im Rahmen eben jenes Mediums durchzuführen, dessen Weiterentwicklung es absehbar dienen sollte? Dass ich mich letztlich tatsächlich für diese Plattform als Instrumentarium zur Projekt-Durchführung entschied, sei an dieser Stelle als schlichtes Faktum vorweggenommen. Warum ich es allen Unwägbarkeiten und instrumentarischen Schwächen zum Trotz tat, sei im Folgenden geschildert.

4.4.4.1 Chancen und Potenziale von Online-Erhebungen

Das Aufkommen und die Popularisierung des World Wide Web haben der empirischen Sozialforschung in vieler Hinsicht neue Dimensionen erschlossen⁸¹⁰. Couper und Coutts rekurrieren zwar vornehmlich auf das Instrument der Befragung als empirischer Methode, skizzieren aber meines Erachtens die Potenziale des Web für Untersuchungen auch anderer Basis-Anla-

⁸⁰⁸ Klammer 2005, S. 282

⁸⁰⁹ Zimbardo/Gerrig 2004, S. 45

⁸¹⁰ Vgl. Bandilla 1999, S. 14f.; vgl. ferner Janetzko 1999, S. 5f.; vgl. ferner Welker et al. 2005, S. 14ff.

gen sehr treffend, wenn sie feststellen: „Web-Befragungen bieten der Forschung enorme Möglichkeiten für preiswerte selbstadministrierte Befragungen unter Verwendung einer großen Auswahl an Stimulus-Materialien (Ton, Abbildungen, Video etc.), die bislang nicht verfügbar oder für den breiten Einsatz in Befragungen mit Interviewern zu teuer waren“⁸¹¹. Insbesondere die Möglichkeit, die Person des Interviewers in Web-Befragungen durch ein unpersonales, mithin emotionsfreies und „neutrales“ Formularelement ersetzen zu können, entfaltet in diesem Zusammenhang besonderen Reiz. Und dies eben nicht nur aus Kostenerwägungen: Antworten im Sinne „sozialer Erwünschtheit“ treten in Online-Untersuchungen als Einflusselement deutlich in den Hintergrund⁸¹². Oft ist hier auch vom sogenannten „Erwartungseffekt“ die Rede; dieser tritt im klassischen Verständnis auf, „wenn ein Forscher oder Betrachter dem Probanden auf subtile Weise mitteilt, welches Ergebnis er erwartet – und so erst die gewünschte Reaktion hervorruft“⁸¹³. Die „Unpersönlichkeit“ des Web – also vor allem die physische Absenz einer befragenden oder beobachtenden Person – darf daher durchaus als Chance begriffen werden für empirische Online-Untersuchungen⁸¹⁴.

Freilich sind Online-Erhebungen durchaus auch umstritten als Instrument. Im Folgenden werden die wichtigsten Grenzen und Risiken von Online-Erhebungen benannt und diskutiert.

4.4.4.2 Grenzen und Risiken von Online-Erhebungen

Couper und Coutts benennen drei der wichtigsten Fehler, die sie für Stichprobenziehungen im Internet identifiziert haben⁸¹⁵. Es sei allerdings darauf hingewiesen, dass diese beschriebenen potenziellen Hauptfehlerquellen meines Erachtens nur zum Teil als web-spezifische zu betrachten sind. Vielmehr beschreiben die Autoren hier allgemeine Probleme der empirischen Datenerhebung, die im Web wohl durchaus gravierender durchschlagen können als in der „klassischen“ Empirie, aber nicht unbedingt und ausschließlich auf die Spezifika des Mediums selbst zurückzuführen sind.

- Abdeckungsbezogene Fehler (coverage errors) resultieren dabei aus einer unvollständigen Berücksichtigung der Zielpopulation durch die Auswahlgesamtheit. Als Beispiel: Da nur rund 51 % der Bürger der Bundesrepublik Deutschland Zugang zu einem webfähigen Computer haben beziehungsweise diese Grundgesamtheit mindestens einmal monatlich online ist⁸¹⁶, sind Internet-Befragungen mit der Zielpopulation „alle wahlberechtigten Bürger“ nicht empfehlenswert, da Zielpopulation und Auswahlgesamtheit zu stark voneinander differieren: „So stellen abdeckungsbezogene Fehler derzeit wohl die Hauptursache für die mangelnde Verall-

⁸¹¹ Couper/Coutts 2006, S. 228

⁸¹² Vgl. Couper/Coutts 2006, S. 225

⁸¹³ Zimbardo/Gerrig 2004, S. 33

⁸¹⁴ Vgl. Brosius/Koschel 2005, S.192ff.

⁸¹⁵ Vgl. Couper/Coutts 2006, S. 219ff.

⁸¹⁶ Vgl. van Eimeren/Frees 2005, S. 364

gemeinerbarkeit von Ergebnissen aus Online-Befragungen dar“⁸¹⁷. „Nach wie vor werden sehr viele Befragungen im Internet als catch-all-Befragungen durchgeführt (votes), d.h. jeder der mitmachen will, kann an der Befragung teilnehmen. (...) Solche Umfragen haben mehr den Charakter eines Gesellschaftsspiels, ähnlich wie Straßenumfragen oder TED's (orthografischer Fehler im Original, M.L.) im Fernsehen“⁸¹⁸.

- Stichprobenfehler (sampling errors) resultieren daraus, dass nicht alle Mitglieder der Auswahlgesamtheit in eine jeweilige Stichprobe aufgenommen werden können – und zwar schlichtweg deshalb, weil schon die Auswahlgesamtheit praktisch unbestimmbar ist⁸¹⁹. Während beispielsweise bei Telefonbefragungen – zumindest theoretisch – eine echte Zufalls-Stichprobe durch einfaches randomisiertes Wählen von Telefonnummern gezogen werden kann, ist eine Vollerhebung der Auswahlgesamtheit von beispielsweise allen E-Mail-Adressen deutschsprachiger Internet-Nutzer so gut wie unmöglich. Es gibt eben kein „E-Mail-Telefonbuch“.
- Ein „Nonresponse“-Fehler liegt vor, wenn nicht alle Befragten in der Stichprobe gewillt oder in der Lage sind, die Fragen zu beantworten beziehungsweise an der Befragung teilzunehmen. Dies kann in der Tat Ergebnisse von Untersuchungen stark beeinflussen; ist eine Online-Befragung beispielsweise in den Augen von potenziellen Teilnehmenden mit höheren Bildungsabschlüssen zu „primitiv“ gestaltet, nehmen unter Umständen gerade viele Angesprochene aus dieser Gruppe nicht teil. Zu hohe intellektuelle Anforderungen dagegen – oder auch nur quantitativ übermäßig viele Aufgaben – bewirken eventuell einen gegenteiligen Effekt: Nicht nur, aber auch Menschen mit geringerem Bildungsabschluss könnten das Experiment schlicht aus dem Gefühl der Überforderung heraus abbrechen.

Weitere Gründe für Abbrüche oder grundsätzliche Nichtteilnahme von potenziellen Teilnehmenden können sein:

- Die Konstruktion des empirischen Vorhabens auf Grundlage einer Software, die viele Nutzer auf ihrem Rechner nicht vorhalten;
- Kostenerwägungen; zwar setzt sich das schnelle Übertragungssystem DSL zunehmend durch, doch sind nach wie vor auch User mit langsameren Modems und oft auch ohne „Flatrate“ im Netz unterwegs (nach der ARD-/ZDF-Online-Studie 2006 waren in diesem Jahr 2006 immer noch 25 % der Online-Nutzer „analog“ im Netz, also ohne ISDN- oder DSL-Technologie, 48 Prozent verfügten allerdings auch schon über DSL-Technologie⁸²⁰). Wie auch immer sich die Zahlen entwickelt haben: Die Teilnahme an einer Befragung kostet alle potenziellen Teilnehmenden Zeit (Übertragungszeit nämlich, aber ganz banal auch Freizeit), und – ohne das Vorhandensein einer „Flatrate“ – Geld;
- Unerfahrenheit der Nutzer im Umgang zum Beispiel mit Web-Formularen, die ein häufiges

⁸¹⁷ Couper/Coutts 2006, S. 219

⁸¹⁸ Vgl. Forschungsgruppe Wahlen O.J.

⁸¹⁹ Vgl. Couper/Coutts 2006, S. 220f.

⁸²⁰ Vgl. Fisch/Gscheidle 2006, S. 431

Element und im Grunde das einzige Mittel zur Datenerfassung in einer Browser-Umgebung darstellen;

- Sicherheitserwägungen; viele Webnutzer sind durch wiederholte Berichte über „Datenklau“ im Internet verunsichert. Sie könnten zögern, ein Formular auszufüllen, in dem beispielsweise ihr Haushaltseinkommen abgefragt wird, da sie den vertraulichen Umgang mit diesen Daten durch die Befragenden bezweifeln. Auch könnten Menschen zögern, an einer Webumfrage vom Computer ihres Arbeitsplatzes aus teilzunehmen, weil sie fürchten, dass die Geschäftsleitung dahinter kommt und „arbeitsfremdes“ Surfen moniert;
- Initiative der Teilnehmenden erforderlich – potenziell Teilnehmende an Online-Studien müssen den ersten Schritt tun, indem sie die Webseite mit der Befragung aktiv aufsuchen. Eine motivierende Begleitung durch Versuchsleiter entfällt;
- Das „Massen-Missverständnis“. „Vielen Web-Surveys liegt die falsche Annahme zugrunde, dass umfangreiche Stichproben automatisch zu zuverlässigen Ergebnissen führen (...). Inferenzstatistische Aussagen sind jedoch nur bei Zufallsstichproben möglich“⁸²¹.

4.5 Experimentaldesign

4.5.1 Projekt: Online-Studie

Die vorangegangenen Ausführungen haben, so meine ich, deutlich werden lassen: Feldexperimente im Allgemeinen, Online-Studien im Besonderen sind als sozialwissenschaftliche Erhebungsmethoden mit mannigfaltigen Unsicherheiten und potenziellen Störfaktoren belastet⁸²². Letztlich entschloss ich mich dennoch, das von mir projektierte Experiment in Form einer Online-Studie zu konzipieren und durchzuführen – allerdings mit Einschränkungen, die laut Kromrey mein Vorhaben letztlich als „Quasi-Experiment“ qualifizierten⁸²³. Im Folgenden soll kurz umrissen werden, welche Überlegungen mich zu diesem Beschluss führten, und welche Modifikationen klassischer Experimentaldesigns ich im Detail vornahm (und warum).

Zunächst ist Folgendes festzuhalten: Die mediumspezifische Ausgrenzung potenzieller Probanden vom projektierten Experiment (nämlich der implizite Verzicht auf die Mitwirkung von Menschen, die selten oder praktisch nie online gehen) war zwar unvermeidlich mit der Entscheidung für ein online-basiertes Vorhaben – aber ebenso verschmerzbar. Mehr noch: Da die Studie ja gerade und gezielt der Erforschung von Leseverhalten in Online-Medien galt, lag es nahe, insbesondere solche Personen als Teilnehmende einzubeziehen, die sich in eben diesen Medien alltäglich bewegen – der Ausschluss von „Offlinern“ war also im Vorhaben nicht Manko oder Störeffekt, sondern sogar in gewissem Sinne erwünschte „Vor-Selektion“. Grundgesamtheit würden eben nicht „alle“ des Deutschen und des Lesens mächtigen

⁸²¹ Couper/Coutts 2006, S. 223

⁸²² Vgl. auch Reips 1999

⁸²³ Vgl. Kromrey 2006, S. 102

Menschen sein, sondern jene, die sich regelmäßig online aufhalten. Ein, wie beschrieben, gängiger Haupteinwand gegen Online-Experimente – der der mangelnden Abbildung der Grundgesamtheit in der Stichprobe – war damit zumindest zu einem gewichtigen Anteil aus dem Forschungsinteresse selbst heraus entkräftet.

Stichproben- und Nonresponse-Fehler freilich waren als potenzielle Störfaktoren weiterhin virulent. Insbesondere angesichts knapper finanzieller, personeller und auch räumlicher Ressourcen war jedoch bereits in der Konzeptionsphase des Experiments absehbar, dass eine qualitativ wie quantitativ befriedigende und gleichzeitig demografisch ausgemittelte Stichprobe im Rahmen des Vorhabens auch in einer Laborstudie kaum besser realisierbar sein würde als in einer Online-Studie. Nicht allein, dass für ein laborgestütztes Experiment Probanden in ausreichender Anzahl, in gewichteter Altersstruktur und eventuell auch unter den Kriterien anderer demografischer Eigenarten hätten akquiriert werden müssen; diese Probanden hätten auch bereit sein müssen, sich aktiv an den Ort der Messung zu begeben – was sicherlich nicht alle Angesprochenen tatsächlich getan hätten. Eine ausgemittelte Stichprobe an Versuchsteilnehmenden zu werben, wäre also im Rahmen einer Labor-Untersuchung womöglich schwerer gefallen als in der dann tatsächlich durchgeführten Online-Studie. Das dritte von Couper und Coutts angesprochene Hauptproblem von Online-Studien, das der „Non-Response“, ist schließlich in meinen Augen kein web-spezifisches: Auch wer sich weigert, an einer Labor-Studie teilzunehmen, betreibt im übertragenen Sinne „Non-Response“⁸²⁴.

Insgesamt war bereits in der Entstehungsphase des Vorhabens absehbar, dass ich nur auf sehr eingeschränkte räumliche Ressourcen zurückgreifen könnte im Rahmen einer eventuellen Labor-Studie; schließlich war ein mit funktionierenden Computern ausgestatteter, mir quasi unbeschränkt zur Nutzung bereitstehender Versuchsraum Voraussetzung für eine erfolgreiche Labor-Studie. Dieser uneingeschränkte Zugriff war mir nicht gestattet; zwar ergab sich für mich potenziell die Möglichkeit, gelegentlich zuzugreifen auf einen isolierten, mit sechs Rechnern ausgestatteten Raum des Instituts für Journalismus und Public Relations an der Fachhochschule Gelsenkirchen als „Labor“ – doch diente und dient dieser Raum vorrangig und nicht zuletzt qua Widmung als „Lehrredaktion Rundfunk“ und war aus diesem Grunde meiner uneingeschränkten und zuverlässigen Nutzung entzogen. Dass die Fachhochschule Gelsenkirchen zudem eher schlecht erreichbar ist mit öffentlichen Verkehrsmitteln und auch sehr am Rande des Gelsenkirchener Ortsteils Buer ihren Standort hat, hätte die Akquirierung von „Labor-Probanden“ jenseits des studentischen oder akademischen oder nicht-wissenschaftlichen Bediensteten-Milieus an der FH zusätzlich erschwert.

Dass sich zudem die Nachteile von Labor-Untersuchungen (deren alltagsfremde Erhebungsumgebung insbesondere) in Waage halten verglichen mit Feldstudien (weniger Möglichkei-

⁸²⁴ Vgl. Klammer 2005, S. 280

ten, Störeinflüsse auszuschalten), wurde bereits erwähnt – und von mir ebenfalls dahingehend interpretiert, dass eine Online-Studie für mein Vorhaben durchaus eine gangbare Methodik darstellen würde. Unbestritten: Insbesondere würde ich nicht kontrollieren können, ob die Probanden das Experiment alleine durchführten und ob die Beleuchtungsverhältnisse, unter denen die Teilnehmenden das Experiment durchführten, vergleichbar waren. Ich werde allerdings noch beschreiben, dass ich zumindest eine eingeschränkte Kontrolle der Versuchsumgebung – in Bezug beispielsweise auf verwendete Betriebssysteme der Teilnehmenden oder deren gelegentliche Versuche, das Experiment mehrfach durchzuführen – auch im Rahmen des Online-Experiments durchaus herzustellen in der Lage war. Und ich werde beschreiben, dass ich dies auch tat, mithilfe programmiertechnischer Prüfmechanismen, die mir bereits in der Konzeptionsphase des Experiments bekannt waren und mir zusätzlich Mut und Zuversicht verschafften, dass eine Online-Studie mit verlässlicher Datensammlung möglich sei.

Letztlich, und dies sei nicht verschwiegen, reizte mich jedoch vor allem die Aussicht, dass ich im Rahmen einer Online-Studie eine weit höhere Anzahl an Teilnehmenden würde akquirieren können, als dies bei einem lokal gebundenen Labor-Experiment der Fall sein würde. Obgleich mir natürlich bewusst war, dass „Masse“ in diesem Sinne nicht gleich „Klasse“, also unmittelbar Qualität der Daten erzeugen würde, schien mir dieser offenkundige Nachteil einer Online-Erhebung verschmerzbar – eben weil ich abwägend davon ausging, dass die Chance, eine quantitativ befriedigende Abbildung einer repräsentativen „Online-Nutzerschaft“ im Rahmen eines Labor-Experiments unter den Bedingungen der vorhandenen räumlichen, jedoch auch personellen und finanziellen Ressourcen herzustellen, nicht viel größer sein würde als in einem Online-Experiment.

Fazit: Mit der Entscheidung für ein Online-Experiment verzichtete ich zunächst vor allem aktiv auf ein gezieltes „Sampling“ der Teilnehmenden im Sinne einer demografisch repräsentativen Abbildung der vermuteten Grundgesamtheit der „regelmäßigen Online-User“. Dies tat ich vor allem zugunsten einer vermutlich deutlich höheren Teilnehmer-Anzahl, vermittels derer ich, so meine Hoffnung und Annahme, eventuelle demografische Verzerrungen zumindest teilweise auszugleichen in der Lage sein würde.

4.5.2 Wie misst man Dutzende von typografischen Konstellationen?

Es war, wie dargelegt, eines der wesentlichen Ausgangsziele der hier beschriebenen Untersuchung, nicht nur einen oder zwei isolierte typografische Faktoren auf ihre Lesbarkeit hin zu prüfen, sondern auch vermutete wechselseitige Interaktionen dieser typografischen Faktoren zu identifizieren. Obwohl in der frühen Konzeptionsphase damit noch keineswegs feststand, wie viele und welche Konstellationen im Detail letztlich im Experiment tatsächlich getestet werden würden, war doch absehbar, dass die Anzahl dieser Konstellationen einen Umfang

erreichen würde, der die Konstruktion eines verlässlichen Forschungsdesigns zumindest nicht einfach geraten lassen würde. Es würde auf ein echtes mehrfaktorielles Versuchsdesign⁸²⁵ herauslaufen.

Ein simples Rechenexempel verdeutlicht die Dimensionen, die das Projekt bereits in einer frühen Überlegungsphase gewann. So erzeugt beispielsweise allein die wechselseitige Kombination von zwei Schriftarten einerseits mit zwei Schriftgrößen andererseits bereits vier zu untersuchende Einzel-Konstellationen; zwei Schriftarten in jeweils drei Schriftgrößen untereinander zu „kreuzen“, bringt bereits sechs Designs hervor. Fügt man nun als Kriterien nur zwei divergierende Zeilenabstände hinzu, hätte man bereits 12 Experimental-Designs zu verwalten; die ergänzende Hinzunahme von drei variierenden Zeilenbreiten würde deren Anzahl auf stolze 36 anschwellen lassen. Es sei an dieser Stelle der Vorgriff erlaubt, dass ich im hier beschriebenen Projekt letztlich 90 typografische Konstellationen auf ihre Lesbarkeit untersucht habe. Auch wenn ich in der frühen Entwicklungsphase diese letztliche Größenordnung noch nicht vor Augen hatte: Mir war rasch klar, dass ich absehbar eine Fülle von typografischen Konstellationen empirisch zu beobachten beabsichtigte, die in Vollständigkeit bestenfalls ein hoch motivierter Proband zu bearbeiten bereit und imstande wäre. Da das Experiment zudem als Online-Studie konzipiert war, die praktisch keine Kontrolle der Versuchsumgebung meinerseits erlaubte und daher den potenziellen Probanden unter anderem den jederzeitigen, völlig sanktionsfreien Abbruch des Experiments gestattete, erschien mir wenig erfolgversprechend, auf eine „Vollabdeckung“ der Experimentalkonstellationen durch wirklich alle potenziell Teilnehmenden zu spekulieren.

Denn wie viele Versuchsteilnehmer – so die Überlegung – würden bereit sein, auch „nur“ 30 experimentelle typografische Anordnungen am Stück zu rezipieren, geschweige denn jene 90, die es am Ende im Experiment tatsächlich waren? Selbst bei einer unterstellten durchschnittlichen Lesedauer von einer halben Minute pro Text und Test-Anordnung hätte dies von den Probanden summa summarum ein praktisch pausenfreies Zeitopfer von gut 50 Minuten oder mehr eingefordert. Jedem Probanden also sämtliche zu testenden typografischen Konstellationen vorzulegen, erschien, zumal im Rahmen eines Online-Experiments, ein utopischer Ansatz.

Worin bestand die Alternative? Ein akzeptables Maß an Gültigkeit hätte zweifellos ein Experiment beanspruchen können, in dem einer hohen Anzahl von Probanden ein einziger, fixer Text, gesetzt in jeweils einer der 90 avisierten Schriftanordnungen, zur Bearbeitung vorgelegt worden wäre. Jeder Teilnehmer hätte in diesem Modell also exakt eine der 90 zu testenden Konstellationen zu lesen bekommen – und sonst keine. Durch zufällige Zuweisung der Test-Konstellationen an die Teilnehmenden wäre nach diesem Ansatz eine gewisse empiri-

⁸²⁵ Vgl. Brosius/Koschel 2005, S.199ff.

sche Gleichverteilung hergestellt worden, die die erhobenen Daten durchaus vergleichbar gestaltet hätte.

Doch auch dieses Denk-Modell barg Schwächen. Denn, dies war mir klar, es bedürfte einer ausgesprochen großen Anzahl Probanden, um jede der typografischen Anordnungen in einer Häufigkeit zu testen, die statistisch wirklich verwertbare Daten für jede einzelne Anordnung erbrächte. Um beispielsweise für jede der tatsächlich projektierten 90 Konstellationen lediglich 10 Testergebnisse zusammenzutragen, wäre in diesem Ansatz die Akquirierung von 900 Versuchspersonen vonnöten gewesen. Für aussagekräftigere 25 Testergebnisse pro getesteter typografischer Konstellation wären bereits 2250 Teilnehmende zu akquirieren gewesen – eine angesichts des damals unabsehbaren Akquirierungserfolgs zumindest sehr ehrgeizige Zielmarke. Eine Zielmarke, die zu erreichen ich seinerzeit für kaum möglich hielt (und die Ergebnisdarstellung wird zeigen, dass ich damit offenbar nicht ganz falsch lag).

Gelöst war das grundlegende Problem damit freilich noch lange nicht: Würden sich potenziell unbestimmt viele, aber wünschenswerterweise eventuell mehr als 30 typografische Konstellationen in einem einzigen Experiment überhaupt wirksam auf ihre Lesbarkeit hin messen lassen? Und selbst wenn sich dieses Vorhaben als prinzipiell umsetzbar herausstellen würde: Welche Anzahl an Teilnehmenden setzte dies voraus, um ein vertretbares Maß an Gültigkeit für die Studie beanspruchen zu können? Gefragt war mithin ein systematischerer Blick auf mögliche, taugliche methodische Alternativ-Ansätze.

4.5.2.1 Eigentlich unmöglich: Einrichtung einer Kontrollgruppe

Das klassische Experimentaldesign sieht grundsätzlich die Einrichtung je einer Kontroll- und mindestens einer Versuchsgruppe vor⁸²⁶, zwei Gruppen von Experimentarteilnehmenden also, die, zumindest im Optimalfalle, „in ihrer Zusammensetzung hinsichtlich soziodemografischer und anderer Merkmale vollkommen identisch sind“⁸²⁷. Dabei wird der Stimulus, dessen Wirkung getestet werden soll, der Kontrollgruppe vorenthalten, die Versuchsgruppe wird dagegen mit dem Stimulus konfrontiert⁸²⁸. Treten Mess-Unterschiede auf zwischen Experimental- und Kontroll-Gruppe, kann daraus – so zumindest die Grundannahme – Art und Ausmaß der Wirkung des Stimulus’ interpretiert und quantifiziert werden.

Da im vorliegenden Experiment die Wirkung typografischer Anordnungen (als Stimulus) auf die Lesegeschwindigkeit (als abhängiger Variable) getestet werden sollte, erwies sich jedoch bereits die Umsetzung dieser basalen Vorgabe als schwierig. Denn ein „typografie-freier“ Stimulus ist, in Bezug auf Lese-Aufgaben, schlechterdings nicht kreierbar: Es gibt keine Schriftanordnung, der keine typografische Entscheidung zugrunde liegen würde. Text ohne Typo-

⁸²⁶ Vgl. Kromrey 2006, S. 97f.

⁸²⁷ Klammer 2005, S. 274

⁸²⁸ Vgl. Klammer 2005, S. 275

grafie gibt es nicht. Selbst Menschen, die einen Text in *Word* verfassen, ohne eine einzige aktive typografische Wahl zu treffen, fällen eine typografische Entscheidung, eine passive: Nämlich die, die Standardeinstellungen des Programms zu übernehmen.

Dies gilt natürlich nicht für alle existenten typografischen Faktoren: Natürlich kann Kursivsatz prinzipiell abwesend sein in einer Schriftanordnung, in einer anderen dagegen vorhanden. Der typografische Stimulus „Kursivsatz“ wäre also prinzipiell durchaus isolierbar. Jedoch: Die Ausschaltung eines grundlegenden Faktors wie beispielsweise des Zeilenabstandes ist schlechterdings nicht möglich: Dieser Stimulus ist nicht absentierbar, sondern höchstens variierbar. Ähnliches gilt für die Faktoren „Schriftgröße“ und „Zeilenbreite“: Nur „kein Text“ enthält diese wichtigen Stimuli gar nicht.

4.5.2.2 Lösungsansätze bisheriger Studien

Wie versuchte die bisherige experimentelle Lesbarkeitsforschung dem Dilemma beizukommen, dass eine Ausschaltung typografischer Stimuli per se unmöglich ist – und damit im Grunde auch die Definition von Kontrollgruppen-Aufgaben, mithin von Kontrollgruppen schlechthin? Die Antwort ist ebenso schlicht wie im Detail erläuterungsbedürftig: Entweder durch willkürliche Setzung einer bestimmten typografischen Anordnung als Ausgangs-Stimulus – oder durch konsequenten Verzicht auf die Einrichtung von Kontrollanordnungen und damit -gruppen überhaupt.

Tinker und Paterson hielten es in den meisten ihrer Experimente mit einem zweistufigen Testsystem, in dem jeder Proband gleichsam als Kontroll- und Versuchsteilnehmer in Personalunion agierte. So wurden alle Teilnehmenden in der jeweils ersten Hälfte eines Experimentaldurchgangs mit der einen, in der zweiten Hälfte mit der anderen typografischen Anordnung konfrontiert, in der jeweils ein typografischer Faktor experimentell variiert wurde. Relevanter Messwert waren dabei nicht die Durchschnitts-Lesezeiten der beiden Testformen, sondern die Lesezeit-Differenzen zwischen den getesteten typografischen Anordnungen. Erfasst wurde also nicht im Kern absolute Lesezeiten, sondern individuell der Geschwindigkeitszuwachs beziehungsweise die Verlangsamung, die eine fix definierte Variation eines definierten typografischen Faktors induzierte. Auf diese Weise hofften Paterson und Tinker, den individuellen Lesegewohnheiten ihrer Probanden besser gerecht zu werden: Ein grundsätzlich „langsamer“ Leser würde die vermeintlich „besser lesbare“ Schriftanordnung absolut womöglich immer noch langsamer lesen als ein routiniert-schneller Leser die jeweils „schlechtere“ typografische Anordnung; die relative Beschleunigung jedoch war ein verlässlicher Wert. Tinker und Paterson verwendeten hier also ein sogenanntes „within-subjects-“ oder „Messwiederholungsdesign“, das im Prinzip „jeden Probanden zu seiner eigenen Kontrollgruppe (macht)“⁸²⁹.

⁸²⁹ Zimbardo/Gerrig 2004, S. 35

Wendt kritisiert als Nachteil dieses Verfahrens, „dass man immer nur so viele typografische Formen miteinander vergleichen kann, wie man Parallelförmigen des Tests zur Verfügung hat – das waren auch bei Tinker und Paterson nur zwei – und dass man bei Verwendung mehrerer Parallelförmigen nacheinander doch immer damit rechnen muss, dass die letzten durch Lernprozesse gegenüber dem ersten begünstigt sind“⁸³⁰. Wendt modifizierte das Prinzip in einem eigenen, deutschsprachigen Test deshalb dahingehend, dass allen Probanden zunächst ein einheitlicher Textabschnitt in identischer typografischer Anordnung vorgelegt wurde, in der zweiten Hälfte des Durchgangs jedoch Texte in insgesamt 16 von der Ursprungs-Schriftanordnung differierenden typografischen Konstellationen nach Zufallsprinzip und gleichmäßig verteilt über die Teilnehmenden ausgegeben wurden. Beibehalten hatte Wendt in der Auswertung freilich das Prinzip, nicht die absoluten Lesezeiten, sondern die jeweilige Veränderung der Geschwindigkeiten zum Maßstab zu erheben.

Redelius dagegen verzichtete in dem in seiner Dissertation beschriebenen Bildschirm-Lesbarkeits-Experiment gänzlich auf die Einrichtung einer Kontrollgruppe⁸³¹. Stattdessen legte er jedem seiner 50 Probanden am Bildschirm insgesamt 15 verschiedene Texte nach Zufallsprinzip in „zufälliger Gleichverteilung“, also je Proband in individuell unterschiedlicher Reihenfolge vor. Jeden Text hielt Redelius dabei in jeweils nur einer typografischen Gestaltung vor. Obgleich er die oben beschriebenen Normen von Wendt sowie Tinker und Paterson – die ja durchaus apodiktisch das Vorhandensein einer Kontrollgruppe einfordern – durchaus beschreibt und diskutiert, belässt es Redelius in der Begründung seines Verzichts auf eine Kontrollgruppe beim Hinweis, dass „die Reihenfolge der Bildschirmpräsentation und eventuell längere Lesezeiten am Beginn der Versuchsreihe über alle Testformen gleichverteilt sind und sich damit herausmitteln“⁸³². Nach einem ähnlichen Prinzip maßen Bernard und Mills im Jahre 2000 die Lesbarkeit von zwei Schriftarten in zwei Schriftgrößen und wiederum zwei Auflösungsverfahren am Bildschirm⁸³³: Den 35 Probanden wurden alle acht interessierenden typografisch-monitorspezifischen Varianten, die wiederum jeweils inhaltlich an einen einzigen Text gekoppelt waren, in jeweils zufälliger Reihenfolge vorgelegt. Aus den erhobenen Mittelwerten wurden dann Schlüsse gezogen.

Auf die Einrichtung einer Kontrollgruppe verzichtete auch Weisenmiller in der bereits erwähnten Untersuchung, die den Kern seiner Dissertation bildete⁸³⁴. Anders als Redelius sowie Bernard und Mills jedoch konfrontierte Weisenmiller alle akquirierten 264 Probanden mit ein und demselben Versuchstext, wobei er diesen einen Text in wechselseitig kombinierter Gleichverteilung in vier Schriftarten und drei medialen Varianten ausgab (nämlich in 1-Bit-Bildschirmdarstellung, 8-Bit-Bildschirmdarstellung und hochqualitativer Papierausdruck).

⁸³⁰ Wendt 1992, S. 292

⁸³¹ Vgl. Redelius 1998

⁸³² Redelius 1998, S. 87

⁸³³ Vgl. Bernard/Mills 2000

⁸³⁴ Vgl. Weisenmiller 1999

Hier erhielt also nicht jeder Proband mehrere Testtexte in wechselndem typografischen Design, sondern lediglich einen Testtext in einer der insgesamt zwölf gemessenen Ausprägungen medialer und typografischer Natur.

4.5.2.3 Systematisierung der bisherigen Ansätze

Betrachtet man die oben vorgestellten Vorgehensweisen, so lassen sich deren Ansätze meines Erachtens auf folgende Kernfragen veichten, die mithin am Beginn einer Experimental-konzeption zu beantworten sind:

- Welche typografischen Faktoren sollen auf ihre Lesbarkeit hin untersucht werden?
- In wie vielen Ausprägungen soll jeder einzelne dieser Faktoren auf seine Lesbarkeit hin untersucht werden?
- Werden alle zu untersuchenden Faktoren in allen vorgesehenen experimentellen Ausprägungen wechselseitig gekreuzt und damit auch auf mögliche Interaktionen hin untersucht – oder findet diesbezüglich eine Auswahl statt?
- Falls eine solche Auswahl stattfindet: Welche Interaktionen werden untersucht und welche nicht (und nach welchen Kriterien erfolgt der Ausschluss der Alternativen)?
- Können und sollen alle potenziellen Teilnehmer im Experiment jeweils mit allen zu untersuchenden typografischen Ausprägungen konfrontiert werden – oder nur mit einem Teil der Ausprägungen?
- Wie viele inhaltlich differierende Experimentaltexte sind vorzusehen, um jeden Teilnehmenden mit der gewünschten Anzahl Ausprägungen konfrontieren zu können, ohne dass textlich-inhaltliche Redundanzen das Messergebnis verfälschen können?
- Wird jeder einzelne Experimentaltext in lediglich einer der zu untersuchenden typografischen Anordnungen vorgehalten oder in mehreren, möglicherweise allen interessierenden typografischen Varianten?
- Werden allen potenziellen Probanden alle verfügbaren Experimentaltexte vorgelegt oder nur ein Teil derselben?
- Soll die Reihenfolge variieren, in der die Texte den Probanden – in welcher typografischen Anordnung auch immer – vorgelegt werden, oder ist diese Reihenfolge für jeden Einzelteilnehmer identisch zu gestalten?

Es schälten sich für mein projektiertes Experiment mithin im Wesentlichen folgende drei methodische Alternativen heraus:

Klassisches Experimentaldesign

Diese Variante verfolgt einen relativen Ansatz: Er sieht die Einrichtung einer Kontroll- sowie einer oder mehrerer Versuchsgruppen (Experimentalgruppen) vor, die jeweils mit einem oder mehreren unterschiedlichen, je individuellem Probanden jedoch gleichbleibenden typografischen Stimuli versorgt würden und deren jeweilige gruppenspezifische durchschnittli-

che Lesezeiten verglichen würden. Die Forschungshypothese lautet, verkürzt formuliert: Das typografische Design der Kontrollgruppe zeitigt keine besseren Ergebnisse als das der Experimentalgruppe. Im Ergebnis würden die Messresultate der Versuchsgruppen mit der der Experimentalgruppe konfrontiert und eine Kausalbeziehung konstruiert nach dem Muster: Experimentalgruppe las x % (eben um einen bestimmten, relativen Betrag) schneller als Kontrollgruppe. Im Grundansatz ist dieses Design unbestreitbar „sauber“, sofern auf in Bezug auf soziodemographische Struktur homogene Gruppenzuteilungen geachtet wird.

„Within-subjects“-Design nach Tinker und Paterson

Hierbei handelt es sich ebenfalls um einen relativen Ansatz: Er sieht die Messung individueller Beschleunigungen oder Verlangsamungen im Leseprozess vor, die ein und derselbe Proband beim aufeinander folgenden Lesen zweier oder mehrerer unterschiedlicher typografischer Anordnungen zeigt, und den anschließenden Gesamtabgleich dieser Beschleunigungen oder Verlangsamungen über alle Probanden hinweg. Im Ergebnis: Die (wiederum relative) Veränderung beispielsweise der Schriftgröße von 12 auf 14 Punkt bewirkt, unabhängig von der individuellen Lesegeschwindigkeit, einen bestimmten durchschnittlichen Betrag an Beschleunigung oder Verlangsamung bei den Probanden. Dieser Ansatz trägt der erwähnten Unterstellung Rechnung, dass „der Unterschied der individuellen Lesegeschwindigkeiten der einzelnen Testteilnehmer erfahrungsgemäß die größte Varianzquelle (ist), gegen die die durch typografische Variation zu erzielenden Unterschiede relativ gering sind“⁸³⁵. In der modifizierten Variante von Wendt erhalten alle Teilnehmenden einen Kontrolltext in ein und derselben Typografie, in einem zweiten Teil jedoch unterschiedliche Schriftkonstellationen zu lesen.

Randomisiertes Experimentaldesign ohne Kontrollgruppe

Diesen Ansatz möchte ich als einen „absoluten“ klassifizieren. Es existiert keine Kontrollgruppe, jede Experimentalgruppe erhält eine bestimmte Anzahl von Texten in unterschiedlichen typografischen Konstellationen vorgelegt. Innerhalb dieses Modells sind wiederum zwei Varianten vorstellbar, nämlich, dass

- alle Probanden den oder dieselben Texte in der jeweils selben, nämlich in allen interessierenden und zu messenden typografischen Konstellation vorgelegt bekommen, oder
- jeder Proband jeweils nur einen Teil der insgesamt zu messenden Konstellationen vorgelegt bekommt, nur wenige Teilnehmende also das exakt gleiche Forschungsdesign erleben.

Auch in diesem Ansatz wird nicht negiert, dass Menschen – unabhängig von Bildungsgrad, Alter und anderer individuell feststellbarer Spezifika – durchaus sehr unterschiedlich schnell lesen. Es wird jedoch, darüber hinaus, davon ausgegangen, dass sich eben diese Unterschiede bei ausreichend zahlreicher Teilnehmer-Menge sowie bestmöglicher Zufalls-Verteilung der Testaufgaben auf die Teilnehmenden zunehmend um einen bestimmbareren, „absoluten“

⁸³⁵ Wendt 1992, S. 291

Mess-Mittelwert sammeln und Ausreißerwerte einzelner Testpersonen sich in der Gesamtschau wechselseitig neutralisieren und ausgleichen – vulgo: „ausmenden“ – werden.

4.5.2.4 Bewertung der Ansätze

Ein entscheidender Nachteil im klassischen einerseits wie im Experimentaldesign nach Tinker und Paterson andererseits liegt meines Erachtens in der potenziellen Ressourcen-Verschwendung – zumindest in Fällen, in denen mehr als eine Experimentalgruppe vorgesehen ist, und dies, so war absehbar, würde im von mir konzipierten Experiment der Fall sein. In diesen Fällen geht eine – die „Kontrollgruppen“-Konstellation nämlich – numerisch stark überproportional im Vergleich zu den Experimentalgruppen in die Wertung ein, als willkürlich gesetzte relative Bezugsmarke, die damit aber Kapazitäten von Probanden bindet, die gut und gerne auch auf andere interessierende und untersuchbare Konstellationen verwendbar wären. Allerdings halten zumindest Brosius und Koschel die Einrichtung einer echten Kontrollgruppe in mehrfaktoriellen Designs ohnehin für überflüssig, „denn alle Gruppen werden miteinander verglichen und kontrollieren sich sozusagen gegenseitig“⁸³⁶.

Größter Nachteil des „randomisierten“ Ansatzes ohne Kontrollgruppe wie auch des klassischen Experimentaldesigns ist jedoch, dass individuelle Unterschiede in der Lesegeschwindigkeit verschiedener Testpersonen hier keine substantielle Berücksichtigung finden. Erzielt ein potenziell „langsamer“ Leser ein brillantes Teilergebnis in der Konfrontation mit einer „guten“ Schriftkonstellation, kann dieses Detail-Resultat immer noch deutlich unter jenem liegen, das ein vermeintlich „routiniert-schneller“ Leser in einer vermutet „schlechter lesbaren“ Anordnung erzielt hat. Dieses Manko fängt wiederum der „Within-subjects“-Ansatz von Tinker und Paterson vorbildlich auf. Indem hier nämlich jeder Einzel-Teilnehmer quasi zur Kontrollgruppe seiner selbst wird, ist der individuell erfahr- und messbare Fort- oder Rückschritt bei Wechsel einer typografischen Konstellation hervorragend herausarbeitbar. Dieser Vorzug muss aber, wie gezeigt, wiederum relativiert werden durch die diesem Modell immanente Inkaufnahme des Nachteils, dass die absolute Anzahl messbarer Konstellationen sehr klein ausfällt und die Kontrollgruppen-Resultate quantitativ sehr überproportional zu Buche schlagen. Auch Wendts Modifikation des Ursprungsansatzes ändert daran nichts.

Effizienter im Einsatz der „Ressource Proband“ ist diesbezüglich unbestreitbar der „randomisierte“ Ansatz. Der Verzicht auf eine Kontrollgruppe oder auch nur eine „Kontroll-Konstellation“ wird in diesem Modell durch eine hohe Teilnehmerzahl und eine dadurch erhoffte „Ausmittlung“ der individuellen Lesegeschwindigkeits-Unterschiede ausgeglichen, der Erwartung also, dass sich durch eine möglichst zufällige Gleichverteilung der Aufgaben auf die Teilnehmenden „schnelle“, „mittelmäßige“ und „langsame“ Leser im Sample eines jeden Untersuchungs-Elements die Waage halten und damit wechselseitig ausgleichen würden.

⁸³⁶ Brosius/Koschel 2005, S. 199

In der Gesamtschau erschien mir der „randomisierte“ Ansatz, bei allen beschriebenen Schwächen, der für mein Vorhaben tauglichste – insbesondere, weil er durch Verzicht auf Installation einer Kontrollgruppe absehbar die Ermittlung von Werten für eine höhere Anzahl von Testkonstellationen erlauben würde als die beiden anderen Modelle. Es handelt sich zudem, trotz der sichtbaren Abweichungen vom klassischen Experimentaldesign, um ein offenkundig akzeptiertes empirisches Verfahren, das jedoch vor allem eines voraussetzt: Eine ausreichend hohe Anzahl an Versuchsteilnehmern, die eine befriedigende Anzahl an Messergebnissen ins Ergebnis einfließen lassen und sich in ihrer Gesamtheit ausgleichen würden.

4.5.2.5 Ein heikler Ausweg: Ein Proband – mehrere Testtexte

Auch die Grundsatz-Entscheidung für das „randomisierte“ Verfahren hatte freilich ein weiteres, oben bereits angedeutetes Kernproblem noch nicht aus dem Weg geschafft: Es würden sich vermutlich weder ausreichend Probanden finden, die alle vorgesehenen 90 typografischen Testkonstellationen durcharbeiten würden in einem Experimentaldurchgang; selbst wenn in diesem Durchgang großzügige Pausen vorgesehen würden, eventuell sogar tagesweise Unterbrechungen – noch war wahrscheinlich, dass bei Konfrontation aller Teilnehmenden mit jeweils nur einem Text in einer Konstellation absolut ausreichend Messergebnisse zustande kommen würden.

Es führte also kein Weg an der Überlegung vorbei: Wie konnte die Anzahl erhobener Lesezeiten methodisch sauber auf ein empirisch verwertbares Maß gehoben werden, ohne andererseits die Teilnehmenden zeitlich, intellektuell und physisch zu überfordern? Abhilfe versprachen zwei Modelle, die im Prinzip auf das Prinzip der Messwiederholung⁸³⁷ referenzieren:

- Variante A war, jedem potenziell Teilnehmenden nicht nur einen, sondern nacheinander mehrere Texte in jedoch durchgehend einer der 90 zu testenden Schriftkonstellationen vorzulegen. Dass die Inhalte dieser Texte variieren mussten, lag dabei auf der Hand: Ein Lernbeziehungsweise Trainingseffekt hätte absehbar schon beim zweiten Erscheinen ein und desselben Textes so stark auf das Messergebnis durchgeschlagen, dass die Daten nicht mehr als verwertbar hätten gelten können.
- Variante B sah ebenfalls vor, jeden Teilnehmenden mit mehreren Texten, diese Texte jedoch in jeweils unterschiedlichen typografischen Konstellationen zu konfrontieren; dies würden dann aber eben nicht 90 Texte in den projektierten 90 Anordnungen sein können, sondern deutlich weniger. Jeder Proband würde also nur einen willkürlich gesetzten Teil der insgesamt zu messenden typografischen Konstellationen zu lesen bekommen.

Eine Konsequenz freilich war beiden Varianten gemein: Beide ließen es als unausweichlich erscheinen, die potenziellen Probanden des Experiments mit mehr als nur einem Text zu

⁸³⁷ Vgl. Brosius/Koschel 2005, S. 200f.

konfrontieren. Dies aber bedeutete, dass neben die zu messenden typografischen Faktoren im Experiment ein fünfter, ein eindeutiger Stör-Faktor „Text“ treten musste – dessen Einfluss aufs Messergebnis also genauso zu berücksichtigen (oder alternativ und bestenfalls: auszuschalten) war wie der der anderen, der typografischen. Dazu später mehr.

Fest stand ferner, spätestens an dieser Stelle der Überlegungen: Das Experiment würde definitiv nicht umsetzbar sein unter der Prämisse, dass alle Probanden den exakt gleichen Stimuli ausgesetzt sein würden. Nicht alle Probanden würden mit allen 90 zu testenden Schriftkonstellationen konfrontiert werden können, sondern mit deutlich weniger und also einer willkürlichen Teilmenge der zu testenden typografischen Anordnungen. Die Versuchsgruppen würden also nicht nur unvermeidlich soziodemografisch, sondern auch im Aufgabenprofil erheblich variieren. Dieses Problem freilich relativiert sich durch den Hinweis, dass auch in klassischen Experimentaldesigns mit Kontrollgruppen notwendigerweise nicht alle Teilnehmenden identische Stimuli erfahren – eine hohe Sorgfalt bei der Zuteilung von Probanden auf die unterschiedlichen Versuchsgruppen ist akzeptiert als Methode, die Störeinflüsse dieser inhomogenen Zusammensetzung von Vergleichsgruppen zu minimieren.

4.5.2.6 Zuteilungsverfahren im Experiment – eine Abwägung

Zuvorderst war nun freilich zu entscheiden, welcher der beiden oben beschriebenen Varianten der Vorzug zu geben war: Sollte jeder Proband mehrere Texte in gleichbleibender typografischer Anordnung zu lesen bekommen? Oder mehrere Texte in gleichzeitig typografisch variierenden Anordnungen?

Für beide Varianten sprachen Argumente. Variante A besticht insbesondere dadurch, dass sich in diesem Verfahren absehbare Unterschiede in den Testtext-Inhalten und damit den Basis-Rezeptionszeiten durch die für jeden Probanden individuell durchweg identischen typografischen Konstellation zumindest teilweise ausmitteln würden. Der Störfaktor „Textinhalt“ wäre in diesem Modell also methodisch wirksam dämpfbar in seinem Einfluss auf die Messergebnisse.

Nachteil dieses Verfahrens ist allerdings, dass es die Konstruktion einer erwünschten Gleichverteilung von „guten“ (oder auch: „routinierteren“ und damit „schnelleren“) Lesern über alle Messergebnisse hinweg erschwert. Würde in diesem Modell beispielsweise ein „guter“ Leser randomisiert an eine beliebige der 90 zu messenden Konstellationen geraten, würde dieser „gute“ Leser absehbar gleich mehrere, damit aber potenziell eben auch überproportional viele „gute“ Leseergebnisse für eben diese eine gemessene typografische Konstellation ins Gesamtergebnis einbringen. Würde dagegen der typografische Stimulus, wie in Variante B vorgesehen, über die aufeinander folgenden Texte hinweg modifiziert, würden die individuellen Lesegeschwindigkeiten „guter“ und „schlechter“ Leser vermutlich gleichmäßiger über alle Messvarianten hinweg verteilt.

Ein weiteres Argument gegen Variante A ergab sich aus den Erfahrungen und Rückmeldungen meines ersten Pretests heraus, in dem ich ein vergleichbares Prinzip angewandt hatte. Einige Teilnehmer hatten sich damals irritiert gezeigt, dass sie in einem Experiment, in dem doch ausdrücklich die Qualität unterschiedlicher typografischer Ausprägungen Gegenstand des Interesses war, mit durchgehend lediglich einer einzigen Konstellation über alle Experimentaltexte hinweg konfrontiert wurden. Auch aus motivatorischen und Transparenz-Erwägungen heraus lief also Vieles auf Variante B hinaus.

Auch durchführungspraktische Überlegungen mündeten in einer tendenziellen Präferenz für Variante B. Die Anzahl akquirierbarer Teilnehmer lag in der Konzeptionsphase des Experiments naturgemäß im Ungewissen; klar aber war als Ziel, dass sich die typografischen Faktoren zumindest nicht völlig ungleichgewichtig auf die Messergebnisse insgesamt verteilen sollten nach Abschluss der Studie. Was aber würde geschehen, wenn sich beispielsweise die typografischen Konstellationen „12“, „54“ und „67“ innerhalb aller Messergebnisse als unterrepräsentiert herausstellen sollten in einer Zwischenbilanz? In Variante A würde die Zusatz-Akquirierung von drei weiteren Probanden notwendig, die dann aber wiederum eventuell viel mehr Messergebnisse einbringen würden als zum wechselseitigen quantitativen Ausgleich der Konstellationen im Gesamtergebnis notwendig. In Variante B dagegen wäre nur ein weiterer Experimentaldurchgang eines einzigen weiteren Probanden erforderlich – weil dieser absehbar je ein Messergebnis für jede der unterrepräsentierten Konstellationen beisteuern würde. Die Ausbalancierung der typografischen Konstellationen über alle Messergebnisse hinweg würde also unter den Bedingungen von Variante B leichter fallen und weniger Teilnehmende erfordern.

Andererseits, um das Eingangs-Argument wieder aufzugreifen: Die Methodik in Variante B erforderte in weitaus stärkerem Maße als die von Variante A, die Versuchstexte als Einflussbeziehungsweise gar Störfaktoren im Prozess der Auswertung zu berücksichtigen. Das bedeutete wiederum, dass in Variante B de facto nicht mehr nur 90 Schriftkonstellationen zu vergleichen waren, sondern zusätzlich eine noch festzusetzende Anzahl Text-Inhalte in jeweils allen der 90 Schriftkonstellationen! Bei vier Versuchstexten pro Proband (dies ist nicht nur ein Beispielwert, sondern der letzte Wert aus der hier beschriebenen Online-Studie) ergaben sich mithin stattliche 360 Einzel-Konstellationen.

Fallen aber nicht zuletzt durch diese Multiplizierung die vorgebrachten Argumente für Variante B weitestgehend in sich zusammen? Nur bedingt, wie ich meine.

Zum einen ist ins Feld zu führen, dass es sich beim Faktor „Textinhalt“ um einen der meines Erachtens noch beherrschbarsten und in seinem Störeinflüsse am ehesten gezielt ausschaltbaren und nivellierbaren Faktor im Experiment handelte – vor allem das Risiko, „gute“ und „schlechte“ Leser ungleichverteilt über die Messergebnisse zu verteilen, schien mir in der Abwägung allemale größer und schwerer minimierbar. Zudem wollte ich ja eben nicht Diffe-

renzen inhaltlicher Textqualitäten messen, sondern typografische Manipulationen – deren durchführungspraktischer Kontrolle und Gleichverteilung war also in meinen Augen unbedingter Vorrang einzuräumen in der Konzeption, und dieser Vorrang drückte sich in Variante B in der Summe deutlich wirksamer aus als in Variante A.

Letztlich konstruierte ich mein Experiment daher nach den grundsätzlichen Leitideen von Variante B – was freilich nicht bedeutete, dass ich den evidenten Schwächen dieses Modells keine Beachtung mehr schenkte. Im Gegenteil: Ich fand, dass insbesondere der Störfaktor „Inhalt“ auch im Rahmen von Variante B durchaus beherrschbar gestaltet werden konnte, knüpfte man die randomisierte Zuteilung der typografischen Konstellationen an zusätzliche, erweiterte Vorgaben. Ich formulierte schließlich folgende Sachverhalte als erstrebenswert:

- Jede der 90 zu testenden Schriftkonstellationen sollte mit exakt der gleichen Anzahl erhobener Messwerte in die Endauswertung einfließen wie die jeweils 89 anderen.
- Auch jeder Testtext sollte über den Experimentalverlauf hinweg in gleicher Anzahl präsentiert werden und ins Messergebnis einfließen.
- Innerhalb aller Messergebnisse eines jeden einzelnen Testtextes sollten sich wiederum die 90 Konstellationen möglichst gleichverteilt wiederfinden.
- Diese dreifache Gleichverteilung sollte in einem möglichst hochgradig randomisierten Verfahren erzielt werden.

Dieses vielgestaltige Ziel, dies sei an dieser Stelle vorweggenommen, wurde erreicht. Wie es technisch und durch gezielte Datenauswahl realisiert wurde, wird weiter unten dargestellt.

4.5.2.7 Schwerwiegende Frage: Noch aussagekräftig?

Zumindest im Bereich der Lesbarkeitsforschung ist das oben beschriebene und von mir realisierte Experimentaldesign nach meinen Recherchen ohne Beispiel – nicht einmal ein ansatzweise vergleichbares Prinzip fand ich vor. Selbstverständlich war ich versucht, diese Tatsache alleinursächlich meiner eigenen konzeptionellen Kreativität zuzuschreiben. Dennoch gebot die Aufrichtigkeit, nochmals sorgfältig abzuwägen, ob der von mir gewählte Ansatz trotz seines Bemühens um quantitative Ausbalancierung erhobener Werte nicht doch an grundlegenden „Konstruktionsfehlern“ krankte, die seine Alleinstellung nicht nur erklären, sondern schlimmstenfalls empfindlich entwerten würden.

Insbesondere war eine Tatsache nicht von der Hand zu weisen: Es würden keine wirklich isolierbaren Versuchsgruppen mehr existieren in meinem Experiment. Beziehungsweise: Potenziell würde jeder Proband seine eigene, isolierte „Versuchsgruppe“ bilden.

So war beispielsweise die rein statistische Chance eines jeden Teilnehmers, Testtext 2 in der typografischen Konstellation „33“ präsentiert zu bekommen, eins zu neunzig – weil eben neunzig typografische Konstellationen je Text vorgesehen waren im Experiment. Dass jedoch

eben dieser Teilnehmer vorher Text 1 zum Beispiel in Konstellation Nummer „60“ gelesen hatte, war statistisch nur noch in einem von $(90 * 90 =) 8100$ denkbaren Fällen wahrscheinlich. Eventuell aber determinierte die in Text 1 rezipierte Typografie das Leseergebnis in Text 2 (und eben einer anderen Konstellation) mit?

Sicher: Sowohl Redelius als auch Bernard und Mills ließen „ihre“ Probanden Texte in einer Zufallsreihenfolge rezipieren; dennoch waren im Ergebnis stets alle Probanden allen experimentellen Stimuli ausgesetzt in diesen Studien.

Den Mut, den beschriebenen Weg weiter zu gehen, schöpfte ich insbesondere aus der „Entdeckung“ jenes statistischen Verfahrens, das – bei Einhaltung bestimmter Voraussetzungen im Experimentaldesign – offenbar sehr potent ist in der Identifikation von Auffälligkeiten auch in komplexeren und aus heterogeneren Stichproben gewonnenen Messergebnissen: der Varianzanalyse. Dazu später mehr.

4.6 Gemessene Faktoren und Attribute

Eine der zentralen Fragen, die in der Konzeptionsphase des Experiments zu klären war, war die nach Art und Anzahl der typografischen Faktoren, deren Einfluss auf die Lesbarkeit von Texten gemessen werden sollte. Dass hier Beschränkung not tat, lag auf der Hand; da schließlich eines der zentralen Erkenntnisziele der Untersuchung darin bestand, nicht nur einzelne typografische Faktoren in ihrem Einfluss auf die Lesegeschwindigkeit zu bestimmen, sondern auch und vor allem mögliche Interaktionen dieser Einzel-Faktoren, führte an einer wechselseitigen „Kreuzung“ dieser Faktoren im Experiment kein Weg vorbei. Dies wiederum bedeutete aber, dass jede Hinzunahme von Faktoren oder auch nur die Addition einer weiteren Ausprägung eines einzelnen typografischen Faktors die Anzahl zu testender Designs multiplizieren würde – und damit die Anzahl Testteilnehmer, die nötig würden, um aussagekräftige Resultate zu erzielen.

Welche Faktoren also waren zu messen, und in wie vielen Ausprägungen? An dieser Stelle war Spekulation unumgänglich, aber auch eine Art „Risikoabschätzung“. So stellten sich vor allem diese Kernfragen:

- Welche der diversen in Hauptkapitel 3 besprochenen typografischen Faktoren – zum Beispiel Schriftart, Schriftgröße, Zeilenabstand, Zeilenbreite, Schriftfarbe, Laufweite, Schriftausrichtung – bedingen die Lesbarkeit eines Textes zuvorderst? Welche spielen vermutlich eher eine untergeordnete Rolle?
- Wenn die wichtigsten Faktoren – vermeintlich – identifiziert sind: In wie vielen Ausprägungen sollten sie auf ihre Lesbarkeit hin gemessen werden?
- Wie viele Teilnehmende würden sich für die Mitwirkung am Experiment kalkulierbar akquirieren lassen? Diese Spekulation war unumgänglich, schließlich musste die Anzahl der zu

messenden Schriftkonstellationen in ein gesundes Verhältnis gesetzt werden zur Anzahl der Probanden, denen sie absehbar vorgelegt werden würden.

Welche typografischen Faktoren die Rezipierbarkeit einer Anordnung vorrangig beeinflussen, ist ein in der Literatur überraschend wenig diskutiertes Faktum. Die wenigen Auslassungen, die ich fand, bewegen sich jedoch in einer sehr eindeutigen und gleichgesinnten Richtung. So meinen Kommer und Mersin: „Einer der wichtigsten Faktoren für die Lesbarkeit eines Textes ist seine Größe. Fast ebenso einflussreich sind Zeilenlänge und Zeilenabstand“⁸³⁸. Bei Bergner liest es sich ähnlich: „Als wichtigster Faktor für die Lesbarkeit gilt die Größe einer Schrift. Zeilenlänge und Zeilenabstand, die mit der Schriftgröße korrelieren, haben auf die Lesbarkeit einen fast ebensolchen Einfluss“⁸³⁹. Für Brekle sind die entscheidenden Faktoren gleichfalls die Schriftart, die Schriftgröße, der Zeilenabstand, die Zeilenlänge – und schließlich noch das makrotypografische Momentum des Satzspiegels⁸⁴⁰.

Ich fand die obigen Aussagen auch aus meiner eigenen beruflichen Erfahrung heraus bestätigt. So entschloss ich mich, die vier Phänomene „Schriftgröße“, „Schriftart“, „Zeilenabstand“ und „Zeilenbreite“ im Experiment einzeln und im Wechselspiel genauer unter die Lupe zu nehmen.

4.6.1 Faktor Schriftart

4.6.1.1 Warum den Faktor Schriftart messen?

Ich bin in Kapitel 3.4.1.7 intensiv auf den traditionsreichen Meinungsstreit eingegangen, der um Lese-Tauglichkeit und Qualität der beiden meistverwendeten Schriftgruppen heutiger Tage gefochten wird: serifentragende und serifenlose Schriftarten. Ich habe gezeigt, dass das hartnäckig weitergetragene Apodiktum, dass serifentragende Schriften auf Papier besser lesbar seien, im Webdesign eine eindrucksvolle Umkehr erfahren hat. Am Bildschirm gelten nunmehr den meisten „Fachleuten“ serifenlose Schriften als Nonplusultra, und in der Schroffheit der Ablehnung des anderen, des Serifen-Lagers, reicht der Eifer bald an jenen heran, der auf dem „Print“-Feld in anderer Richtung zu konstatieren war und ist.

Nun ist zweifellos zu konstatieren, dass einige empirische Bildschirm-Studien bereits die Unhaltbarkeit dieser Theorie wenn nicht erwiesen, so doch als Schluss nahegelegt haben. Ich habe mich dennoch entschieden, den Faktor „Schriftart“ zumindest in kleinem Umfange in meine Messungen mit einzubeziehen, und zwar

- um die bislang kaum untersuchte Wechselwirkung dieses Faktors mit den Phänomenen „Zeilenbreite“ und „Zeilenabstand“ untersuchen zu können, und

⁸³⁸ Kommer/Mersin 2002, S. 116

⁸³⁹ Vgl. Bergner 1990, S. 21

⁸⁴⁰ Brekle 1994, S. 221ff.

- weil der Streit um die bestmögliche Web-Schriftart nach wie vor mit so unvergleichlicher Vehemenz betrieben wird, dass eine völlige Nichtbeachtung im Rahmen einer Studie insgesamt nur schwer vermittelbar gewesen wäre.

4.6.1.2 Anzahl und Ausgestaltung der gemessenen Attribute

Ich habe mich entschieden, im Experiment zwei Schriftarten miteinander zu vergleichen: *Verdana* und *Times* beziehungsweise deren *Microsoft*-Derivat *Times New Roman*. Die Entscheidung fiel aus folgenden Erwägungen heraus:

- die beiden Schriftarten gehören gänzlich unterschiedlichen Schriftgruppen an (*Verdana* als Groteske und *Times* als serifentragende Schrift), um deren Tauglichkeiten bezüglich der Lesbarkeit sich traditionell – und, wie gezeigt, im Web nicht weniger als auf dem Papier – heiße Dispute ranken.
- es sind die – neben der *Helvetica* respektive des *Microsoft*-Klons *Arial* – vermutlich aktuell meistgenutzten Schriftarten im World Wide Web. Für die Praxisrelevanz des Experiments war dieser Aspekt ein miteinscheidender.
- die *Times* ist nach wie vor voreingestellte „Standardschrift“ in den meisten gängigen Browsern. Das heißt, dass eine große Anzahl von Websites ohne CSS-Voreinstellungen nach wie vor in dieser Schriftart dargeboten werden dürfte.
- Die *Arial* wurde nicht ins Experiment einbezogen, da sie als Groteske einerseits derselben Schriftgruppe wie die *Verdana* angehört, andererseits offenkundig auch *Microsoft* von dieser seiner „Eigenschöpfung“ Abstand zu nehmen beginnt⁸⁴¹. Ebenfalls nicht berücksichtigt wurde die *Georgia*, die, wie meine Kleinerhebung in Kapitel 3.4.3.2 gezeigt hat, im Web nach wie vor eine recht geringe Akzeptanz genießt. Eine Praxisrelevanz konnte ich daher in der Messung der Lesbarkeitsqualitäten dieser Schriftart nicht erkennen.
- ein weiterer Reiz entstand aus dem Abgleich einer dezidiert für die Bildschirmanwendung entworfenen Schriftart (der *Verdana* nämlich) mit der am Monitor lediglich adaptierten, originär für die Verwendung in Print-Erzeugnissen geschaffenen Schriftart *Times*.

4.6.2 Faktor Schriftgröße

4.6.2.1 Warum den Faktor Schriftgröße messen?

Untersucht werden sollte weiterhin der Einfluss der Schriftgröße auf die Lesbarkeit von Browser-Texten. Dass dieser Faktor Eingang in die Untersuchung fand, hatte folgende Gründe:

- Die Empfehlungen in der Literatur zum Phänomen „Schriftgröße“ sind insgesamt recht unbefriedigend. Dass sie variieren, mag geschmacklichen Gründen oder variierenden Quellen

⁸⁴¹ Vgl. Klein 2005, S. 13; vgl. ferner Klein 2004, S. 14

geschuldet sein und ist zunächst nicht ungewöhnlich. Dass sie in ihrer überwiegenden Anzahl uneindeutig sind, liegt dagegen auf der Hand. Denn es ist zu bemängeln, dass die numerischen Punkt-Angaben, auf die sich die zitierten Autoren festlegen, fast durchgehend unter Missachtung des Faktors „Schriftart“ gesetzt werden. Dass aber beispielsweise die für die Größenwahrnehmung relevanten Mittelhöhen einer *Times New Roman* runde 15 % kleiner ausfallen als die einer *Verdana* in numerisch identischer Größe, bleibt bei den meisten Autoren unberücksichtigt. Damit aber verlieren die zitierten Größen-Empfehlungen erheblich an Wert.

- Derselbe Vorwurf ist an einen Großteil der vorliegenden empirischen Studien zum Thema „Schriftgröße“ zu richten. Auch hier wurden oft „Scheinvergleiche“ angestellt zwischen zwar numerisch identisch großen, aber in ihrer tatsächlichen Lettergröße teilweise stark differierenden Buchstaben verschiedener Schriftarten.

4.6.2.2 Anzahl und Ausgestaltung der gemessenen Attribute

Um einen aussagefähigen Vergleich von Schriftgrößen im Experiment auch unter den Bedingungen zweier verschiedener Schriftarten zu gewährleisten, wurden für *Verdana* und *Times* jeweils drei individuelle Größen festgelegt – nämlich jeweils eine „Variante klein“, eine „Variante mittel“ und eine „Variante groß“. Jede Variante einer Schriftart entsprach dabei in der x-Höhe bestmöglich der gleichnamigen Variante der jeweils anderen Schriftart.

Zunächst verfolgte ich, um den Handreichungs-Charakter der projektierten Studie für die Praxis möglichst hoch zu gestalten, das Ziel, in jedem Falle eine Schriftgröße zu testen, die in möglichst vielen etablierteren Web-Angeboten Verwendung findet. Diese war relativ rasch identifiziert. Mithilfe von Screenshots ermittelte ich, dass die x-Höhe sowohl der Standard-Fließtextschrift auf „Spiegel online“ (das ist die *Verdana*) im wahrscheinlich marktführenden Browser *Internet Explorer* unter der Standard-Schriftgrößen-Einstellung „mittel“ ebenso 7 Pixel betrug wie der „Brottext“ unter „Focus online“ (ebenfalls *Verdana*). Gleichfalls 7 Pixel hoch fällt unter der beschriebenen Standard-Einstellung die Normaltextschrift unter „kicker online“ aus (Schriftart *Tahoma*). Die „Netzeitung“ dagegen setzt ihre Standardschriftart *Times* mit einer Mittellänge von 8 Pixeln.

Somit drängte sich geradezu auf, im Experiment die Schriftarten *Verdana* und *Times* in jedem Falle in Schriftgrößen zu testen, die unter Browser-Normalbedingungen die x-Höhen beider Schriftarten 7 Pixel groß auf dem Bildschirm erscheinen lassen. Um eine größtmögliche Korrespondenz der Schriftgrößen herzustellen, fertigte ich eine Vergleichsmatrix an.

Abbildung 4.1 simuliert die Bildschirmdarstellungen der Schriften *Verdana* und *Times* in den jeweils angegebenen numerischen Größen (als Maßeinheit diente hier der Bildschirm-Pixel). Die einzelnen Größen-„Päckchen“ wurden dabei nach x-Höhen sortiert und vertikal einander gegenübergestellt.

Zunächst ist festzuhalten, dass sich die x-Höhen der Schriften keineswegs analog zur nominellen Schriftgröße verhalten – was freilich zu erwarten war, da die Relation zwischen Ober- und Mittelhöhe in beiden Schriftarten unterschiedlich ist. Dennoch finden sich einige interessante Konstellationen im direkten Vergleich. So ergibt sich für die *Verdana* eine x-Höhe von 6 Pixeln beispielsweise ausschließlich bei der definierten Schriftgröße „11 Pixel“, eine x-Höhe von 7 Pixeln resultiert für die *Times* dagegen aus nicht weniger als drei Größenangaben: „14 Pixel“, „15 Pixel“ und „16 Pixel“. Die Veränderungen in diesen drei Größen finden mithin nur im Bereich der Versalhöhen statt, der Buchstabenbreite und der Laufweite – die Mittelhöhe dagegen bleibt unverändert, obgleich zwischen 14 und 16 Punkt rechnerische 12,5 Prozent Größen-Differenz liegen.

Welche Schriftgrößen waren also zu definieren für jene ominöse Richtgröße „7 Pixel x-Höhe“? Immerhin kamen für die *Verdana* die nominalen Größen-Werte „12“ und „13 Pixel“ in Frage, um diese Mittelhöhe zu erzeugen. Ich entschied, als Referenzgröße 13 Pixel festzusetzen, und zwar aus folgenden Gründen:

- die Laufweite der 13-Pixel-Variante erscheint ausgeglichener, weniger „gequetscht“ als die kleinere Version;
- ein Blick in die Style-Sheet-Vorgaben unter anderem von „Focus online“ zeigte, dass die x-Höhe 7 Pixel aus der eindeutigen Festlegung einer Schriftgröße von 13 Pixeln resultiert (http://pg.squids.images.focus.de/fol/css/fol_home.css, Besuch 07.11.2006).

Die x-Höhen im Vergleich	
Verdana 10 px x-Höhe 5 px Hamburgefonts	Times 11 px x-Höhe 5 px Hamburgefonts
Verdana 11 px x-Höhe 6 px Hamburgefonts	Times 12 px x-Höhe 6 px Hamburgefonts Times 13 px x-Höhe 6 px Hamburgefonts
Verdana 12 px x-Höhe 7 px Hamburgefonts	Times 14 px x-Höhe 7 px Hamburgefont Times 15 px x-Höhe 7 px Hamburgefon
Verdana 13 px x-Höhe 7 px Hamburgefonts	Times 16 px x-Höhe 7 px Hamburgefo
Verdana 14 px x-Höhe 8 px Hamburgefonts	Times 17 px x-Höhe 8 px Hamburgefo
Verdana 15px x-Höhe 8px Hamburgefont	Times 18px x-Höhe 8px Hamburgef
Verd. 16px x-Höhe 9 px Hamburgefon	Tim. 19 px x-Hö. 9 px Hamburgef

Die *Times* und die *Verdana* im direkten Vergleich nach dem Kriterium der x-Höhe.

ABBILDUNG 4.1

Vergleichsweise einfach war dagegen die Bestimmung der numerischen Größe der *Times* mit 7 Pixel x-Höhe: Hier wählte ich unter den drei verfügbaren die „goldene Mitte“, also die Schriftgröße 15 Pixel.

Mit wie vielen weiteren – und vor allem: welchen – Schriftgrößen aber war diese „Standard-Größe“ nun im Experiment zu vergleichen? Ich entschloss mich, die „Spiegel online“-Größe als „Variante mittel“ zu definieren und diese abzugleichen mit jeweils einer größeren und einer kleineren Variante. Würde – so meine Überlegung – eine schrittweise Erhöhung von Schriftgrößen tatsächlich mit einer besseren Lesbarkeit korrelieren, so käme dieser Zusammenhang in einer Reihung von drei verglichenen Größen deutlicher zum Ausdruck, als wenn nur zwei Größen verglichen würden. Zudem bliebe die Frage der Schriftgröße unbefriedigend gelöst, würde beispielsweise am Ende des Experiments zwar die Erkenntnis stehen, dass eine Verkleinerung Einbußen bei der Lesbarkeit nach sich zöge – aber gleichzeitig die Frage unbeantwortet bliebe, ob sie bei einer Vergrößerung zunähme.

So definierte ich die „Variante klein“ (beziehungsweise, wie sie im Folgenden bezeichnet werden soll, als Größe „0“) als Schriftgröße mit einer resultierenden x-Höhe von 6, die „Variante groß“ (Größe „2“) als Größe mit resultierender x-Höhe 8 Pixel. Von der Messung einer möglichen Schriftgröße „3“ (x-Höhe 9 Pixel) sah ich ab – in diesen Satzgrößen entfalteten beide Schriftarten meines Erachtens bereits einen deutlichen Charakter als Überschriften-, als Akzidenz-Anordnungen, die ja ausdrücklich nicht den Gegenstand der hier beschriebenen Untersuchung bilden sollten. Zudem zeigt bereits ein nur explorativer Streifzug durch die Webseiten journalistisch orientierter Online-Anbieter, dass selbst Schriften in der von mir festgesetzten Experimental-Größe „2“ nur sehr selten Einsatz finden im Web – eine Messung noch größer angelegter Anordnungen hätte also potenziell Probanden-Bemühungen gebunden, die an anderer Stelle sinnvoller abrufbar erschienen.

Aus dieser Vorgabe freilich ergaben sich wiederum Probleme, wie ein neuerlicher Blick auf Abbildung 4.1 zeigt. Einzig die x-Höhe „6“ war – im Falle der *Verdana* – eindeutig der nominalen Schriftgröße „11 Pixel“ zuzuordnen; für die drei übrigen Konstellationen ergaben sich wiederum mehrere Optionen. Schließlich entschied ich mich für folgende Zuordnungen:

Tabelle 4.1: Einrichtung der Schriftgrößen für das Experiment			
	Größe 0 (x-Höhe 6)	Größe 1 (x-Höhe 7)	Größe 2 (x-Höhe 8)
<i>Verdana</i>	11 Pixel	13 Pixel	14 Pixel
<i>Times</i>	13 Pixel	15 Pixel	17 Pixel
<i>Quelle: eigenes Konzept</i>			

Auffällig ist an obiger Aufstellung sicherlich der fehlende „Rhythmus“ in der Größenabstufung der Schriftart *Verdana*. Diese Schriftart in der Größe „15 Pixel“ (denn dies wäre der eigentlich logische nächste Größen-Schritt gewesen) erschien mir jedoch im Abgleich mit der korrespondierenden Variante der *Times* allzu wuchtig (vergleiche dazu nochmals Abbildung 4.1).

4.6.3 Faktor Zeilenbreite

4.6.3.1 Warum den Faktor Zeilenbreite messen?

Gemessen wurden darüber hinaus fünf verschiedene Zeilenbreiten. Dass dieser typografische Faktor überhaupt Eingang in die Untersuchung fand, hat folgende Gründe:

- die Spanne der in der Literatur genannten, vermeintlich „optimalen“ Zeilenbreiten variiert – wie gezeigt – in abenteuerlichen Spannen. Wenn die empfohlenen Margen von „maximal 35 Zeichen, eher weniger“⁸⁴² bis hin zu 60 Zeichen als vermeintlicher Obergrenze⁸⁴³ reichen, diese 60 Anschläge aber andernorts glatt als Minimum definiert werden⁸⁴⁴, liegt der Forschungsbedarf auf der Hand: Hier müssen einzelne Autoren schlichtweg völlig daneben liegen mit ihren Empfehlungen! Wer das war, sollte untersucht werden.
- Kaum ein typografischer Faktor nimmt in einem vergleichbar großen Maße Einfluss auf das allgemeine Erscheinungsbild, die Makrotypografie also von Webseiten, wie die Zeilenbreite. Zumindest Seiten, die in merklichem Umfang Fließtexte bereitstellen (und die journalistisch orientierten Websites gehören definitiv zu dieser Gattung), müssen ihr Grund-Design vor allem am Raum ausrichten, den sie Lesetexten vorbehalten. Erkenntnisse zur anzustrebenden Zeilenbreite dürften also einen nicht unerheblichen Einfluss ausüben auf die Gestaltung von Websites im Allgemeinen.
- Die Zeilenbreite gilt als stark interagierender Faktor in der Typografie; vor allem der Zusammenhang von gewählter Zeilenbreite und definiertem Zeilenabstand wird häufig betont. Da das hier beschriebene Experiment nicht zuletzt auf das Zusammenwirken verschiedener typografischer Einzelfaktoren abhob, war die Messung der Effekte von divergierenden Zeilenbreiten geradezu zwingend.
- Die durchaus vorliegenden empirischen Studien zum Thema „Zeilenbreite“ haben bislang – analog zu den Meinungen der Literatur – ein sehr diffuses Bild ergeben. Teilweise schnitten extrem lange Zeilen mit gelegentlich über 100 Anschlägen am besten ab, teilweise wiederum sehr schmale Zeilenlängen. Eine eindeutige Empfehlung jedenfalls lässt sich aus dem empirischen Forschungsstand definitiv nicht herleiten.

4.6.3.2 Anzahl und Ausgestaltung der gemessenen Attribute

Der typografische Faktor „Zeilenbreite“ wurde im Experiment in fünf Ausprägungen gemessen: Anhand von Zeilen, die im Schnitt 28, 42, 56, 70 oder 84 Anschlägen Platz boten.

Warum fünf gemessene Abstände – eine Menge immerhin im Vergleich zu den Faktoren „Schriftart“ und „Schriftgröße“, die nur in drei Attributen gemessen werden sollten? Weil die zitierten Ratschläge aus der Literatur – und auch die vorliegenden empirischen Erkenntnisse

⁸⁴² Erben 1997, S. 92

⁸⁴³ Vgl. Kommer/Mersin 2002, S. 173

⁸⁴⁴ Vgl. Thissen 2000, S. 86

– in der Gesamtschau eine derart große Spannweite entfalten, dass kaum ein Weg daran vorbeiführte, Zeilenbreiten zwischen etwa 30 und – um auch am anderen Ende der Skala einen vermuteten „Extremwert“ zu platzieren – ungefähr 80 Anschlägen im Experiment zu überprüfen, eher noch mehr. Eine derart große Spanne freilich schien mit beispielsweise nur drei experimentell zu messenden Breiten nur ungenügend berücksichtigt.

Erste Überlegungen gingen daher in die Richtung, vier Breiten-Konstellationen zu testen im Experiment. In Betracht kam beispielsweise eine Messung von Zeilen, die Platz boten für 29, 46, 63 und 80 Zeichen pro Zeile. Freilich erschien die Lücke zwischen dem Wert „46“ und dem Wert „63“ in diesem Modell allzu ausgeprägt; ein Messwert in den „50ern“ schien mir unverzichtbar, nicht zuletzt, um den tradierten Setzer-Wert „abc mal 2“ in die Messung einzubeziehen – dieser Richtwert besagt, dass die optimale Zeilenbreite im Druck bei der Zeichenmenge des „doppelten Alphabets“ anzusiedeln sei⁸⁴⁵, also bei etwa 52 bis 58 Anschlägen (je nachdem, ob man Umlaute als eigenständige Buchstaben wertet oder nicht).

Möglich wäre zwar gewesen, diesem Missstand durch eine arhythmische Staffelung der Messattribute abzuhelfen (zum Beispiel 30, 45, 55, 75), jedoch hätte dies die Studie potenziell der Chance beraubt, eventuelle Linearbeziehungen der Zeilenbreiten eindeutig nachzuweisen. So entschloss ich mich für das bereits einleitend beschriebene „5er-Modell“.

4.6.4 Faktor Zeilenabstand

4.6.4.1 Warum den Faktor Zeilenabstand messen?

Schließlich wurde im Experiment der Faktor „Zeilenabstand“ auf seinen Einfluss auf die Lesbarkeit hin untersucht. Der Hauptgrund für die Berücksichtigung dieses Faktors lag auf der Hand: Es liegt meinen Recherchen nach lediglich eine Studie vor, die sich mit der Bedeutung des Faktors „Zeilenabstand“ in Bildschirmtexten auseinander gesetzt hat. Diese Studie zeitigte jedoch keine sehr erhellenden Erkenntnisse und verglich zudem auch nur den heute wenig verbreiteten Kompresssatz mit einem sehr großen Zeilenabstand. Hier war also unübersehbar eine Forschungslücke zu schließen. Dass der Zeilenabstand zudem als sehr stark mit anderen typografischen Faktoren interagierendes Phänomen gilt, bestärkte mich nur in meinem Vorhaben: Genau diese Interaktionen bildeten schließlich mit die zentralen Erkenntnisinteressen der hier beschriebenen Studie.

4.6.4.2 Anzahl und Ausgestaltung der gemessenen Attribute

Die Bestimmung der Messattribute für den Faktor „Zeilenabstand“ gestaltete sich besonders schwierig, da dieser unmittelbar mit dem Faktor „Schriftgröße“ korreliert. Wo beispielsweise

⁸⁴⁵ Vgl. Liebig 1999, S. 217

ein Zeilenabstand von 16 Pixel für eine 11 Pixel große Schrift durchaus als großzügig gelten darf, ist er für eine 18 Pixel große Schrift eindeutig zu klein bemessen.

Es führte also kein Weg daran vorbei, die gemessenen Attribute für den Zeilenabstand unmittelbar an die jeweils aktuell aufgerufene Schriftgröße zu koppeln. Hier waren also, im Gegensatz zu den bisher diskutierten Faktoren, nicht drei bis fünf fixe Werte zu bestimmen, sondern relative Marken. Da es sich bei der Bestimmung der Schriftgröße als sinnvoll erwiesen hatte, sich von den numerischen typografischen Größenangaben zu lösen und die Attributswerte unmittelbar aus den Schriftbildern der untersuchten Schriftarten heraus zu entwickeln, habe ich auch bei der Bestimmung der Zeilenabstände nach handhabbaren Kriterien gesucht, die vor allem den differierenden Schriftbildern der *Verdana* und der *Times* gerecht werden und vor allem aus diesen Schriftbildern unmittelbar generiert werden konnten.

Es erschien mir notwendig, bei der Festlegung der zu messenden Zeilenabstände auch die Versalhöhen der untersuchten Schriftarten zu berücksichtigen. Denn ist die vertikale Differenz zwischen Mittel- und Oberlängen einer Schrift sehr ausgeprägt, generiert dies bereits aus sich selbst heraus Weißraum, Luft also zwischen den einen Normaltext dominierenden Minuskeln und der Grundlinie der jeweils darüber liegenden Zeile. Hier erzeugt das Schriftbild aus sich selbst heraus bereits eine Art „Zeilenabstand“, weshalb herrschende Meinung in der Literatur ist, dass Schriften mit großer Versalhöhe weniger numerischen Zeilenabstand benötigen als solche mit geringeren Differenzen zwischen Versal- und Mittelhöhe.

Wiederum begann ich mit der Analyse verschiedener Screenshots. Besonderes Augenmerk widmete ich dabei wiederum den Brotschriften von „Spiegel online“ und „Focus online“. Ich stellte fest, dass sich bei diesen beiden Angeboten der Standard-Zeilenabstand mathematisch beschreiben ließ als „Versalhöhe plus Mittellänge“ mit „Pixel“ als Einheit. Die Formel griff ebenfalls für die Grundschrift bei „kicker online“ (Schriftart *Tahoma*) in der Version, die die Grundoptik dieses Angebots bis zum Relaunch am 1. Juni 2006 prägte. Bei der „Netzeitung“ hingegen, wie bereits erwähnt die einzige bedeutendere Publikation mit der *Times* als Grundschriftart, lautete die Formel „Versalhöhe plus Mittellänge minus 1“.

Es schien, als sei die beschriebene Formel insgesamt eine der Weiterverfolgung würdige. Als „Ausgangs-Zeilenabstand“ definierte ich daher für jede der sechs bereits festgesetzten Konstellationen „Schriftgröße-Schriftart“ zunächst die Summe aus Versal- und x-Höhe. Die Werte ermittelte ich mithilfe von Screenshots von gezielt eingerichteten HTML-/CSS-Dateien an einem PC im Browser *Internet Explorer* unter der Voreinstellung „Schriftgröße -> mittel“.

Damit freilich waren nur Ausgangswerte bestimmt. Doch welche – und nicht zuletzt: wie viele – Vergleichswerte für die isolierte Betrachtung des Faktors „Zeilenabstand“ boten sich an? Zunächst stellte sich die Frage, ob der ermittelte und stark an den Standards von „Spiegel online“ orientierte Grundwert wie im Falle der Schriftgröße als Mittelwert definiert werden sollte, dem vergleichshalber jeweils ein niedrigerer und ein höherer Vergleichswert zugeord-

net werden sollte – oder ob der dem „Spiegel-Wert“ entlehnte eher den untersten unter den gemessenen bilden sollte.

Insgesamt wirkten (und wirken bis heute) fast alle insbesondere der auf journalistisch orientierten Webseiten definierten Zeilenabstände auf mich bereits sehr eng bemessen. Auch in der Literatur finden sich, wie beschrieben, mehrheitlich Ratschläge, Zeilenabstände im Web eher großzügig als zu schmal anzulegen; Kompresssatz dagegen wird so gut wie nie empfohlen. Ich beschloss daher, dem beschriebenen „Ur-Wert“ im Experiment zwei größer bemessene Zeilenabstände vergleichsweise gegenüberzustellen; den „Spiegel“-Wert also als den kleinsten der zu vergleichenden zu definieren. Eher aus einer vagen Vermutung heraus entschied ich mich zudem, dem untersten Wert zwei Vergleichsabstände gegenüberzustellen – ob nicht auch eine einzige Vergleichsgröße reichen würde, war angesichts der mageren empirischen Quellenlage bedauerlicherweise kaum abschätzbar.

Ein letztes Problem ergab sich aus der Frage, in welchen Schritten die Vergleichs-Abstände zu vergrößern waren im wechselseitigen und im Verhältnis zu den Ausgangsschriftgrößen. Eine gleichmäßig prozentuale Vergrößerung der Messwerte beispielsweise in Relation zum jeweiligen Ausgangswert, das war absehbar, würde, wiewohl wünschenswert, kaum zu realisieren sein im Experiment („halbe“ Pixel gibt es im Web eben nicht). So vergrößerte ich letztlich die „Ur-Abstände“ mit jedem Schritt linear um 2 Pixel. Daraus ergaben sich schließlich die folgenden Messkonstellationen für mein Experiment:

Tabelle 4.2: Einrichtung der Zeilenabstände für das Experiment					
		Größe	Zeilenabstand 0	Zeilenabstand 1	Zeilenabstand 2
<i>Verdana</i>	Größe 0	11 px	14 px	16 px	18 px
	Größe 1	13 px	16 px	18 px	20 px
	Größe 2	14 px	18 px	20 px	22 px
<i>Times</i>	Größe 0	13 px	15 px	17 px	19 px
	Größe 1	15 px	17 px	19 px	21 px
	Größe 2	17 px	19 px	21 px	23 px

Quelle: eigenes Konzept

4.6.5 Zusammenfassung der Faktoren und Attribute

Hier also nochmals im Überblick: Ich beabsichtigte zu messen und wechselseitig erschöpfend untereinander zu kombinieren

- zwei Schriftarten (*Times* und *Verdana*)
- in drei Schriftgrößen (x-Höhen „6“, „7“ und „8“)
- und fünf Zeilenbreiten (28, 42, 56, 70 und 84 Anschläge pro Zeile im Schnitt)
- in drei Zeilenabständen („Versalhöhe + x-Höhe + 0“, „+2“ und „+4“).

Um also im Experiment jedes Attribut jedes Faktors auf sein Wechselspiel mit allen übrigen Attributen untersuchen zu können, ergab sich die Notwendigkeit, $(2 \times 3 \times 5 \times 3) = 90$ unterschiedliche Schriftkonstellationen auf ihre Lesefreundlichkeit zu testen.

4.6.6 Nicht gemessene Faktoren

Zahlreiche der in Hauptkapitel 3 erkannten und diskutierten typografischen Faktoren fanden keinen Eingang als Messgegenstände in die Konzeption meines Experiments. Es sei ausdrücklich betont, dass diese Nichtberücksichtigungen vor allem eine Entscheidung für die vier letztlich gemessenen Faktoren war, weniger eine gegen die unberücksichtigt gebliebenen. Hauptgrund war die rein praktische Erwägung, dass jeder weitere untersuchte Faktor die benötigte Anzahl an Versuchspersonen erheblich weiter nach oben gezwungen hätte.

Insbesondere der Faktor „Farbe“ spielte in meinen Vorüberlegungen eine Zeitlang durchaus eine erhebliche Rolle. Die auffallende Diskrepanz zwischen der herrschenden „Web-Realität“ (schwarze Schrift auf weißem Grund fast allenthalben) und den beschriebenen Empfehlungen der Literatur hätte Schriftfarbe und Farbgrund sicherlich zu einem weiteren interessanten Untersuchungsgegenstand geraten lassen. Letztlich aber fiel dieser Gegenstand genauso der bewussten Beschränkung auf die – vermeintlich – vier wichtigsten typografischen Einflussfaktoren zum Opfer wie Laufweite, Satzausrichtung und, nicht zuletzt, die diversen typografischen Auszeichnungsvarianten wie Fett- und Kursivsatz, Unterstreichung oder Kapitälcheneinrichtung. Folgestudien mit Fokus auf all diese in der hier beschriebenen Studie unberücksichtigt gebliebenen Faktoren sind jedenfalls wünschenswert.

4.7 Die Experimentaltexte

Die Rekapitulation sei erlaubt: Ich würde im projektierten Experiment nicht auf einen einzelnen Testtext als Messgrundlage rekurrieren können. Die potenziellen Probanden würden notwendigerweise mit unterschiedlichen Inhalten zu konfrontieren sein, die als potenzielle Störfaktoren den isolierten Blick auf die typografischen Anordnungen würden behindern können.

Es galt also in einem nächsten Schritt, Prämissen für die inhaltliche und stilistische Gestaltung der Experimentaltexte zu entwickeln – Leitlinien mit dem Ziel, den Einfluss der Textinhalte und ihrer Stilistik auf die Messergebnisse mindestens konstant zu gestalten, bestenfalls insgesamt zu minimieren; gänzlich ausschaltbar, dessen war ich mir bewusst, würde der Faktor „Text“ in seiner Beeinflussung der Messdaten ohnehin nicht sein. Meine Überlegungen mündeten schließlich in den folgenden basalen Vorgaben:

- Die Messtexte mussten in ihrem Schwierigkeitsgrad vergleichbar gestaltet werden; das intellektuelle und stilistische Anforderungsniveau zwischen den Texten war also möglichst egal zu gestalten, um den potenziellen Störfaktor „Textinhalt“ in seinem Einfluss auf die gemessenen Lesegeschwindigkeiten möglichst gering zu halten.
- Gleichzeitig war der vermutete „Trainings-“ oder „Lernfaktor“ zu berücksichtigen, der sich im Verlaufe des individuellen Experimental-Durchgangs eines jeden Probanden ergeben wür-

de. So war absehbar, dass Teilnehmende mit jedem Text mehr Routine in der experimentellen Bearbeitung der Texte entfalten würden. In diesem Sinne war also, um eine Vergleichbarkeit der Messergebnisse zu gewährleisten, dann doch eine kontinuierliche Steigerung der Textschwierigkeiten über den Zeitraum eines Untersuchungsdurchgangs herzustellen; in welchem Maße freilich die Texte „schwieriger“ würden werden müssen, musste zunächst spekulativ bleiben.

4.7.1.1 Die inhaltliche Gestaltung der Testtexte: Vorgaben

Grundsätzlich war ferner zu klären, welche Inhalte die Experimentaltexte überhaupt sinnvollerweise zu behandeln hätten: politische, wirtschaftliche, „bunte“? Zeitlose, aktuelle oder gänzlich fiktive? Meine Erwägungen resultierten in folgenden Apodikta: Die Texte:

- sollten Inhalte transportieren, die für einen absehbar großen Anteil der Teilnehmenden einen gewissen Neuigkeitswert bergen, damit einen gewissen „Nutzwert“ entfalten und die Motivation der Probanden aufrecht erhalten würden – und wenn dieser Nutzwert auch nur darin bestehen würde, dass in den Texten interessante Begebenheiten oder Tatsachen berichtet werden, die an Stamm- oder Kaffeetisch weitererzählbar sein würden. Bereits weithin bekannte Sachverhalte in den Texten dagegen würden, so meine Überlegung, Teilnehmende tendenziell zu nachlässigerem und weniger konzentriertem Lesen verführen.
- sollten nicht-fiktiv sein. Diese Aussage ist nicht nur der Tatsache geschuldet, dass das Basisinteresse dieser Arbeit der journalistischen Vermittlung im Web gilt; vielmehr wurde im Sinne der eben angesprochenen „Nutzwert“-Theorie unterstellt, dass Lesende zumindest in der Konfrontierung mit unbekanntem Sachverhalten realen Sachverhalten potenziell mehr Interesse entgegenbringen als fiktiven.
- sollten gleichzeitig Inhalte behandeln, die möglichst viele Probanden potenziell interessieren oder zumindest nicht langweilen würden. Dabei war zu berücksichtigen, dass die Teilnehmerschaft im Online-Experiment absehbar heterogen sein würde in Geschlecht, Alter, nicht zuletzt dem Bildungsstand. Dies schränkte die Themenwahl durchaus ein; galt es doch Text-Inhalte zu identifizieren, die absehbar Frauen wie Männer interessieren würden, Ältere und junge Menschen, Akademiker wie geringer Qualifizierte. Komplexere Themen aus der Welt der Wissenschaft (zum Beispiel Fragen der Gentechnik oder differenzierte Erkenntnisse der Kriminologie) fielen damit aus als Themen. Texte zum Thema „Unterhaltungs-Elektronik“ dagegen (zum Beispiel über neue Handy-Technologien) würden absehbar älteren Teilnehmenden mehr Probleme bereiten als jüngeren. Doch auch die Geschlechter galt es zu berücksichtigen. So stand zu vermuten, dass zumindest ein erheblicher Anteil männlicher Versuchsteilnehmer einen Text durchaus goutieren würde, der sich im weiteren Sinne mit Sport, im engeren mit Leistungs-Fußball auseinandersetzen würde – also beispielsweise mit einem bedeutsamen Turnier-Finalspiel und dessen Protagonisten. Bei vielen weiblichen Probanden dagegen wäre ein gänzlich gegensätzlicher Effekt zu erwarten gewesen.

- sollten zeitlos sein. Zeitpunkt und Dauer des Experiments standen in der Konzeptionsphase nur in groben Zügen fest. Texte mit allzu aktuellem zeitgeschichtlichen Bezug, so die Überlegung, würden in der ständigen Gefahr stehen, durch zeitgenössische Entwicklungen noch während der Durchführungsphase inhaltlich überholt zu werden. Abgesehen davon, dass dies viele Probanden absehbar irritieren und ihre Motivation einschränken würde, wäre damit auch eine spätere eventuelle Wiederholung des Versuchs mit dem Ziel der Reliabilitäts-Überprüfung erschwert, wenn nicht hinfällig als Option. Ein Text, der – beispielsweise – eine Geschichte über den Trainer eines Fußball-Bundesliga-Vereins erzählen würde, verlöre absehbar an Glaubwürdigkeit und Zuspruch, würde er vorgelegt zu einem Zeitpunkt, an dem dieser Übungsleiter bereits demissioniert wäre.
- sollten in jeder Hinsicht eine neutrale Position beziehen. Politisch-weltanschauliche Präjudizien sollten die Rezeption der Texte möglichst wenig beeinflussen beziehungsweise schlimmstenfalls: zum Lektüre-Abbruch verleiten.
- sollten stilistisch in informierendem, aber nicht übertrieben „nachrichtlich“-trockenem Duktus gehalten sein. Insbesondere sollte eine allzu boulevardeske Tonalität in der Formulierung vermieden werden, um die Studie insgesamt vor dem Vorwurf der Unseriosität, der gemeinhin als „BILD-Zeitungs-Stil“ denunzierten Komposition, zu bewahren. Allgemein strebte ich die Kreation von „Infotainment“-Texten mit durchaus erlaubtem feuilletonistisch, oft augenzwinkernd-analytischem Tenor an, wie sie in der seriöseren regionalen Tagespresse auf Seiten publiziert werden, die Rubrizierungsköpfe tragen wie „Aus aller Welt“, „Vermischtes“ oder „Welt im Spiegel“.

In Abwägung dieser Vorgaben gelangte ich nach längeren Recherchen zum Ansatz, kuriose, aber vermutlich weithin weniger bekannte historische Begebenheiten mit dennoch erkennbaren Gegenwartsbezügen zum Gegenstand der Testtexte zu machen. Dies garantierte

- einen gewissen (natürlich nur vermuteten) Neuigkeitswert der Informationen, der zudem durch zeitgeschichtliche Entwicklungen nicht unmittelbar bedroht erschien,
- einen dennoch existenten kontemporären Bezug zur Alltagswelt der Lesenden
- einen motivierenden Nachrichtenwert zumindest in dem Sinne, dass der Proband eine leicht weiterzuerzählende Anekdote aus dem Experiment mitnehmen würde.

Vor der eigentlichen Ausgestaltung der Experimentaltexte stand jedoch wiederum eine wichtige, durchführungspraktische Vorüberlegung: Wie lang sollten die Texte ausfallen?

4.7.1.2 Die Länge der Texte

Tinker und Paterson empfehlen eine „angemessene“ Textlänge für Lese-Experimente⁸⁴⁶. Aus den von den US-Forschern dokumentierten Erhebungen ist zu folgern, dass sie Lesezeiten

⁸⁴⁶ Vgl. Paterson/Tinker 1929, S. 364

von um die drei Minuten für angemessen hielten im Rahmen eines Forschungsdesigns. Im von mir projektierten Experiment freilich war ein Faktor zu berücksichtigen, den Tinker und Paterson nicht in ihre Überlegungen einbezogen hatten (weil sie es gar nicht mussten): die Bildschirm-Höhe.

Überlange Texte nämlich würden, so meine Überlegung, die Teilnehmenden zwingen, am Bildschirm vertikal zu scrollen oder – wenn ihnen die Möglichkeit eingeräumt würde – eine virtuelle neue Bildschirmseite „aufzuschlagen“. Beides jedoch würde absehbar ablenken vom eigentlichen, in seiner Dauer zu messenden Leseprozess, würde ihn unterbrechen in einem experimentell schwer bestimmbareren Umfang und damit einen unkalkulierbaren Zeitfaktor in die erhobenen Messwerte einfügen: Wie lange dauert das bloße Erkennen, dass zu scrollen ist, wie lange das „Mausfahren“ zum Scrollbalken, dessen Betätigung im Anschluss und schließlich die visuelle Akklimatisierung an das neu entstandene Monitor-Bild? All diese Faktoren würden absehbar das Messergebnis unkalkulierbar beeinflussen, wie auch verschiedene Studien von Dyson und Haselgrove gezeigt haben⁸⁴⁷.

Diesen Einfluss suchte ich einzuschränken. Ich entschied mich daher, dafür Sorge zu tragen, dass der vermutlich überwiegende Anteil potenzieller Teilnehmender mit sämtlichen Versuchstexten „auf einen Blick“, also auf ein und derselben Bildschirmseite vom ersten bis zum letzten Buchstaben konfrontiert sein würde. Scrolls oder Klicks, die nur dazu dienten, einen einzelnen Experimental-Text von vorne bis hinten zu rezipieren, sollte es im Experiment nicht geben.

Nun war diese Vorgabe zugegebenermaßen praktisch nicht umsetzbar – zumindest nicht hundertprozentig. Schließlich würde beim Online-Experiment die Bildschirmgröße der einzelnen Teilnehmenden nicht kontrollierbar sein – und selbst, wenn dies praktikabel wäre (in Maßen ist es das durchaus), wäre nie feststellbar, ob das vereinzelte Programm-„Fenster“, in dem das Experiment abgebildet wäre, die gesamte verfügbare Bildschirmhöhe ausfüllen würde oder nicht. Auch würde unmöglich feststellbar sein, ob Teilnehmende parallel zum Experiment weitere Programme geöffnet halten würden, deren Benutzeroberfläche eventuell die der Experiment-Oberfläche überlagern würden. Dennoch gab es durchaus Anhaltspunkte, an denen ich mich orientieren konnte.

Zunächst war davon auszugehen, dass die überwiegende Mehrheit der potenziellen Teilnehmenden das Experiment an einem Rechner durchführen würde, auf dem eine Version des Betriebssystems *Windows* der Firma *Microsoft* installiert ist. In diesem Betriebssystem aber füllen die meisten Programmdateien – also auch Webseiten als Dateien im Programmtyp „Webbrowser“ – nach ihrer Aktivierung standardmäßig die verfügbare Bildschirmfläche vollständig aus. Ich konnte also davon ausgehen, dass die Bildschirmfläche bei den meisten Teilnehmenden sehr weitgehend der sichtbaren „Experimental-Fläche“ entsprechen würde (ab-

⁸⁴⁷ Vgl. Dyson/Haselgrove 2000; vgl. ferner Dyson/Haselgrove 2001

züglich der jeweils programmeigenen Kopf- und Fußbereiche natürlich, die aber durchaus kalkulierbar sind).

Weiterhin war anzunehmen, dass die Bildschirme, an denen die potenziellen Probanden das Experiment durchführen würden, mindestens eine Bildschirmdiagonale von 15, eher von 17 Zoll aufweisen würden. Kleinere Monitore, im Maßstab von 14 Zoll beispielsweise, sind zumindest bei Nicht-Laptops so gut wie verschwunden vom Markt, und dies seit Jahren. Damit war aber auch davon auszugehen, dass ein relativ geringer Anteil Teilnehmender eine Bildschirm-Auflösung von weniger als 1024 x 768 Pixeln voreingestellt haben würde (dies darf angesehen werden als systemische Standardauflösung für Monitore ab 17 Zoll, aber auch möglich und oft verwendet für Monitore mit 15 Zoll Diagonale), ein absolut vernachlässigbarer Anteil eine Auflösung von weniger als 800 x 600 Pixeln. Es sei gestattet, an dieser Stelle vorweg zu nehmen, dass im Online-Experiment tatsächlich nur 7,5 Prozent aller Teilnehmenden das Experiment an einem Bildschirm durchführten, der eine Auflösung von weniger als 1024 x 768 Pixeln aufwies. In diesem Sinne darf auch im Nachhinein als richtige Entscheidung gelten, dass ich eine Bildschirmauflösung von 1024 x 768 Pixeln als Grundlage für die weiteren Überlegungen annahm.

Dabei war zu berücksichtigen, dass der bis heute meisteingesetzte Browser *Internet Explorer* der Firma *Microsoft* auf einem Monitor, der aufgelöst ist mit 1024 x 768 Pixeln, in der Vertikalen für den „Kopfbereich“ (der beispielsweise das Adressfeld enthält) um die 100 Pixel veranschlagt, für den „Fußbereich“ runde 80 Pixel. Gerade der Kopfbereich kann dabei auch größer ausfallen – beispielsweise, wenn die „Google“-Toolbar installiert ist, die nochmals mit runden 20 Pixeln Höhe zu veranschlagen ist. Die eigentliche „Inhaltsfläche“ eines Browserfensters reduzierte sich also in der Vertikalen um gute 180 bis 200 Pixel.

Diese Annahme führte mich zu dem Entschluss, dass im Experiment der Versuchstext in keiner der getesteten Schriftkonstellationen über eine Höhe von etwa 550 Pixeln hinausragen durfte. Maßstab war hier der potenzielle Text, der gesetzt werden würde in

- der schmalsten Zeilenbreite (mit der daraus resultierenden größten Zeilenanzahl von allen)
- in der größten Schriftgröße
- mit dem größten Zeilenabstand.

In diversen eigenen Vorstudien gewann ich die Erkenntnis, dass, wenn ich Schriften in der x-Höhe von 8 Pixeln und mit bis zu 23 Pixeln Zeilenabstand messen wollte bei gleichzeitiger durchschnittlicher Zeilenbreite von 28 Anschlägen (also der vertikal „platzintensivsten“ der geplanten Messkonstellationen), Texte mit deutlich mehr als 100 Wörtern (und damit etwa 800 Anschlägen insgesamt) die oben hergeleitete Bildschirmhöhe von 550 Pixeln überschreiten würden. So nahm ich diese griffige Marge von grob 100 Wörtern als Maßstab für die Konzeption der Experimentaltextlängen.

Als vollkommen unproblematisch erwiesen sich in diesem Zusammenhang im Übrigen die maximal resultierenden Textblock-Breiten, die sich aus den von mir projizierten typografischen Messkonstellationen ergeben konnten. Selbst in der höchsten zu messenden Schriftgröße und der breitesten angenommenen Zeilenbreite, so zeigte sich in meinen Vorab-Versuchen, würden die Texte horizontal nicht die physische Breite von etwa 750 Pixeln überschreiten. Sie würden sogar an einem mittelhoch aufgelösten 15-Zoll-Monitor noch in Gänze und ohne Scrollen sichtbar sein.

Es sei allerdings nicht unterschlagen, dass ich mit der projizierten Textlänge „100 Wörter“ deutlich unter Tinkers und Patersons oben erwähnter Vorgabe blieb. Geht man grob überschlägig von vier rezipierten Wörtern pro Sekunde bei durchschnittlich routinierten Lesern aus⁸⁴⁸, würde jeder meiner Einzeltexte vermutete 25 bis 30 Sekunden Lesezeit erfordern – also nur ein grobes Sechstel dessen, was Tinker und Paterson als Wert nahe legen. Im Umkehrschluss hätte dies freilich bedeutet, dass ich den Teilnehmenden meiner Studie bis zu sechs Einzelseiten pro Text hätte zumuten müssen – oder ein häufiges aktives vertikales Scrollen. Beide Varianten hielt ich, wie erwähnt, für nicht erstrebenswert; die reine gemessene Lesezeit wäre sowohl im „Scrolling-“ wie im „Paging-Modell“ durch unkalkulierbare Störeinflüsse korrumpiert worden, die ich unbedingt heraushalten wollte aus dem eigentlichen Erhebungsprozess.

4.7.1.3 Anzahl der Texte

Bevor es an die eigentliche inhaltliche Textgestaltung ging, war zu klären, wie viele Texte im beschriebenen Umfang von 100 Wörtern ich potenziellen Teilnehmenden würde „zumuten“ können, ohne allzu viele Versuchs-Abbrüche zu riskieren. Wie lange würden sich – optimistisch unterstellt – gutwillige, aber zu einem nicht quantifizierenden Anteil sicherlich auch ungeduldige Testteilnehmer „bei der Stange“ halten lassen?

Dabei war nicht nur zu berücksichtigen, dass die Lesezeit pro Text bei einer vermuteten Spanne von einer halben Minute liegen würde. Ins Kalkül war auch zu ziehen, dass ein gewisses Maß an personenbezogenen Daten erhoben werden sollte, dass den eigentlichen Testtexten eine Erläuterung sowie ein „Probetext“ voranzusetzen war, in dem die Probanden ohne aktivierte Zeitmessung das Identifikationsprinzip erlernen und trainieren konnten (dazu später mehr), und schließlich, dass das eigentliche Lese-Experiment durch eine abschließende Befragung ergänzt werden sollte.

Ich entschied, dass das Experiment die Teilnehmenden möglichst nicht mehr als vier, höchstens jedoch fünf Minuten ihrer Lebenszeit kosten sollte. Damit waren aber bereits gewisse Grundzüge des Versuchsdesigns vorgegeben: Ich kalkulierte für Begrüßung, Abfrage der per-

⁸⁴⁸ Vgl. Kösler 1992, S. 99

sönlichen Daten sowie die Kurzeinführung ins Experiment eine runde halbe Minute, für den Probetext – da hier die erste Konfrontation der Teilnehmenden mit dem Messprinzip erfolgen würde – etwa 40 Sekunden. Ebenfalls dreißig Sekunden veranschlagte ich für die abschließende Befragung. Für die eigentlichen Experimentaltexte unterstellte ich – inklusive computerseitiger Berechnungszeiten und selbstgewährten Erholungspausen der Teilnehmenden zwischen je zwei Texten – jeweils etwa 40 Sekunden. Ein Experimentaldurchgang nach diesem vermuteten Muster würde in der Summe also etwa 260 Sekunden beanspruchen, also vier Minuten und zwanzig Sekunden. Dies erschien mir als guter Wert; auf dieser Basis stellte ich meine weiteren Überlegungen an. Es stand also als vorläufige Vorgabe im Raum, dass ich fünf Experimentaltexte zu entwerfen hätte: Einen für den Probedurchgang und vier eigentliche Messtexte.

4.7.1.4 Anzahl und Mess-Sensibilität der Stolperwörter

Wenn also, wie beschrieben, die zu entwerfenden Experimentaltexte um die 100 Wörter umfassen sollten – wie viele Stolperwörter waren dann in einem jeden Text zu platzieren, um einen vermeintlich „alltagsähnlichen“, authentischen Lesevorgang zu induzieren?

Diese Frage war keineswegs unerheblich. Denn die Stolperwörter sollten immerhin nicht nur Indiz der Qualität eines gemessenen Rezeptionsprozesses insgesamt sein, nein, mehr noch, sie waren Messauslöser: Der Zeitpunkt ihrer Identifizierung durch den Probanden würde nicht mehr und nicht minder liefern als jene zentralen Messresultate, durch die ich auf die Qualität von typografischen Anordnungen Rückschlüsse zu ziehen beabsichtigte.

Unabhängig von der beabsichtigten Anzahl Stolperwörter in den Texten standen daher zunächst zwei Modelle auf dem Prüfstand:

- Der Text würde als „erfolgreich rezipiert“ qualifiziert – und also die Mess-Uhr gestoppt – in dem Moment, da alle im Text verborgenen Stolperwörter identifiziert sind. Oder:
- Der Text würde als „erfolgreich rezipiert“ qualifiziert in dem Moment, da nicht alle, aber ein definiertes Quantum an Stolperwörtern im Text identifiziert sind.

Angesichts der wirklichen Simplizität und Augenfälligkeit der Stolperwörter nach Metzses Modell neigte ich von vorneherein der ersten Variante zu – schließlich hatte ich mich ja gerade wegen des frappierenden Unterbrechungs-Charakters für dieses Modell entschieden. Wer also selbst die krassen kontextual-grammatischen Brüche nach Metzses Ansatz „überlesen“ würde, so meine letztliche Überzeugung, konnte keinen sinnentnehmenden Rezeptionsprozess mehr für sich in Anspruch nehmen.

Damit war freilich immer noch keine Entscheidung getroffen hinsichtlich der Anzahl der zu platzierenden Stolperwörter. Letztlich gelangte ich zur Entscheidung, dass ein einzelnes Stolperwort je Text selbst bei Auffindung ein zu unsicheres Indiz sein würde für einen erfolg-

reichen Leseprozess, drei Stolperwörter hingegen – also rechnerisch eines alle 25 Wörter – eventuell zu stark den erwünschten Lesefluss behindern würden. So gelangte ich zum goldenen Mittelweg: Ich würde zwei Stolperwörter in jedem der Experimentaltexte unterbringen.

4.7.1.5 Textinhalte

Nun also konnte es ans Schreiben der Experimentaltexte gehen. Eine vergleichsweise angenehme „Recherche“ in den Geschenkbuchabteilungen zweier hannoverscher Buchhandlungen mündete zunächst darin, dass zu ausgesprochen erschwinglichem Preis zwei Büchlein in meinen persönlichen Besitz übergingen, die mir das Rohmaterial für die finalen Experimentaltexte liefern sollten: „Woher kommt der Ausdruck ...“ leitet Redewendungen wie beispielsweise „Ein Stein vom Herzen fallen“ in humorvoller Diktion etymologisch und historisch her⁸⁴⁹, „Brockhaus! Was so nicht im Lexikon steht“ berichtet dagegen „Kurioses und Schlaues aus allen Wissensgebieten“ – zum Beispiel, wo das erste Kaffeehaus Europas stand⁸⁵⁰.

Ich sah beide Bücher unter den oben konstruierten inhaltlichen Kriterien für die Testtexte durch. Insgesamt entstanden in einem ersten Durchgang neun vorläufige Texte, die ich nach den Motiven der Bücher um- und teilweise neu formuliert hatte. Besonderen Wert legte ich darauf, dass alle Texte die exakt gleiche Wortanzahl aufwiesen (102). Aus diesen neun Texten ermittelte ich in einem zweiten Pretest (siehe Kapitel 4.7.1.8) schließlich die fünf vermeintlich tauglichsten; jene, die letztlich im Experiment zum Einsatz kamen.

Die Texte werden im Folgenden vorgestellt. Enthalten sind bereits pro Text zwei Stolperwörter, mit denen auch die Probanden der Online-Studie konfrontiert waren.

- Text 1 (Probetext): „England“

Dies ist ein Text, in dem Sie üben können, wie es geht. Ihre Zeit wird noch nicht gewertet!

Also: Klicken Sie bitte auf die „Stolperwörter“, die Sie in diesem Text finden!

In ganz Europa fahren die Autos auf der rechten Straßenseite – nur in England nicht. Und zwar aus Tradition: Schon im Mittelalter Galopp hielten sich die englischen Reiter meistens an der linken Seite des Weges. So waren sie besser auf Angreifer vorbereitet, die ihnen entgegen kamen, denn mit der rechten Hand konnten sie ein Schwert führen, eine Lanze oder später eine Pistole. So machte das Linksreiten den Verkehr sicherer. Und die Engländer sehen Kupplung bis heute keinen Grund, daran etwas zu ändern.

AUFLÖSUNG: Die Wörter „Galopp“ und „Kupplung“ ergeben in diesem Text keinen Sinn.

Wenn Sie es noch nicht getan haben, klicken Sie jetzt auf diese beiden Wörter - dann geht es weiter zum nächsten Text!

⁸⁴⁹ Krone 2003

⁸⁵⁰ F. A. Brockhaus GmbH 2000

- Text 2: „Glocke“

Eine harte Strafe verhängte die Stadt Venedig im Jahre 1498 gegen eine Glocke. In der Stadt war damals ein Prediger unterwegs, der der Ketzerei verdächtigt wurde. Als ihn die Polizei gefangen Italien nehmen wollte, läutete die Glocke, und der Angeklagte konnte fliehen. Der Große Rat der Stadt sprach die Glocke daraufhin des Verrats für schuldig. Zur Strafe wurde sie vom Turm gerissen und durch die Straßen geschleift, der Henker peitschte sie aus. Dann wurde sie aus der Stadt verbannt. Noch ärger Bischof traf es einen Hahn, der 1471 in Basel ein Ei gelegt haben sollte. Er wurde als „Teufel in Verkleidung“ verbrannt.

- Text 3: „Olympia“

Bei den olympischen Spielen der Antike traten die Athleten nackt an. Allerdings erst ab dem Jahr 720 vor Christus - vorher hatten die Sportler immer einen Lendenschurz getragen. Im Jahr 720 Kriegsgott jedoch verlor ein Läufer während des Rennens seinen Schurz - und gewann den Lauf. Danach gingen auch seine Gegner ohne Hüllen an den Start. Und diese Tradition hielt sich, bis die Spiele im Jahre 393 nach Christus vom Kaiser in Rom verboten wurden. Große Scham Arena mussten die nackten Sportler übrigens nicht empfinden vor den Frauen im Publikum. Es waren nämlich gar keine da: Im Stadion waren nur Männer zugelassen.

- Text 4: „ Mönche“

Die Deutschen nehmen im Schnitt 2000 Kalorien pro Tag zu sich. Das reicht, um satt zu werden. Die Mönche früherer Zeiten übertrafen diesen Wert jedoch um mehr als das Dreifache. Englische Forscher haben Bratwurst berechnet, dass in einem Orden täglich über 7000 Kalorien pro Kopf vertilgt wurden. Ein Großteil davon bestand aus Alkohol: Die Tagesration der frommen Brüder betrug unter anderem 4,5 Liter Bier. An 60 Feiertagen im Jahr kam noch einmal ein Liter Wein für jeden Mönch dazu. Man kann den Mönchen aber zugute halten, dass Rosenkranz eine mollige Wampe in den Klöstern ohne Heizung der einzige Schutz gegen Kälte war.

- Text 5: „Schlitzohr“

Manch einer empfindet es als Kompliment, wenn man ihn als „Schlitzohr“ bezeichnet. Vielleicht, weil es heute nicht mehr weh tut, eins zu sein? Im Mittelalter zumindest war „Schlitzohr“ alles andere als ein Ehrentitel. Ließ sich nämlich der Geselle eines Handwerkers etwas zu schulden kommen, hatte der Meister das Recht, den Mitarbeiter aus der Zunft auszuschließen. Als Zeichen für die Entlassung wurde Galgen dem Gesellen der Ohrring ausgerissen - das Symbol der Zunft. Wer dann als „Schlitzohr“ durch die Stadt ging, wurde Nase von allen als Missetäter erkannt. Später ging man dann dazu über, auch Betrügern das Ohr mit einem Messer zu schlitzen.

- Text 6: „Kartoffeln“

Wer macht schon das, was Politiker von ihm wollen? Zumindest waren die Bürger von Preu-

ßen sehr skeptisch, als Friedrich der Große die Kartoffel einführte, um seine Untertanen satt zu bekommen. Kaum einer rührte die Knollen an. Bis Friedrich auf eine Idee kam. Er Kochtopf ließ auf den Äckern Schilder aufstellen: „Das Stehlen der Kartoffeln ist bei Strafe verboten“. Der Verzehr von Kartoffeln nahm danach sprunghaft zu. Übrigens brauchte auch die Tomate recht lange, um den Gaumen der Europäer zu Bauernhof erobern. Genau wie die Kartoffel stammte die Tomate aus Amerika, und anfangs wurde sie nur als Zierpflanze gehalten: Sie galt als giftig.

- Text 7: „Hornberg“

Wenn etwas ausgeht „wie das Hornberger Schießen“, dann war alles umsonst, dann ist alles geblieben wie vorher. Der Spruch geht auf eine Geschichte zurück. So wird erzählt, dass die Bürger von Hornberg im Schwarzwald einen Herzog mit Salutschüssen begrüßen Kutsche wollten. Sie übten so lange für das große Ereignis, bis sie ihr ganzes Pulver verschossen hatten. Und so ritt der Herzog in aller Stille in die Stadt ein. Der Legende nach versuchten die Bürger aber, die Situation zu retten: Sie brüllten sich die Seele aus Pistole dem Leib, um die Schüsse nachzuahmen. Ob der hohe Gast darauf hereinfiel, ist jedoch nicht überliefert.

- Text 8: „Ägypten“

Der erste Streik der Welt ist über 3000 Jahre her. Er fand in Ägypten statt, und der Leidtragende war ein Pharao, Ramses der Dritte. Der Pharao ließ auf den Wänden seines Totentempels seine größten Schlachten darstellen. Leider Krokodil nahm er es mit der Bezahlung nicht so genau: Im Jahre 1156 vor Christus legten seine Arbeiter die Arbeit nieder, weil sie zwei Monate keinen Lohn bekommen hatten. Ramses bezahlte schließlich, und die Arbeiten gingen weiter. In den Totentempel zog Ramses übrigens früher ein, als er sich vorgestellt Nilwasser hatte: Drei Jahre nach dem Streik brach eine Verschwörung seines Harems aus, Ramses wurde ermordet.

- Text 9: „Croissant“

Das Croissant stammt nicht aus Frankreich, sondern aus Österreich. Als die Türken vor Wien lagen, versuchten sie, Tunnel in die Stadt zu Turban graben, da sie über die Stadtmauern nicht hinweg kamen. Um den Plan geheim zu halten, gruben die Türken bei Dunkelheit. Sie hatten aber nicht mit jenen Wienern gerechnet, die nachts immer wach waren: den Bäckern. Die bemerkten das Geschabe und schlugen Alarm. Zur Erinnerung backten sie das „Hörnchen“, dem sie die Form des türkischen Halbmondes gaben. Auf französisch heißt Hörnchen „Croissant“, und weil es in Frankreich längst beliebter Kilokalorien ist als in Österreich, hat sich dieser Name eingebürgert.

4.7.1.6 Die Positionierung der Stolperwörter in den Texten

Es dürfte bereits deutlich geworden sein, dass die Position der Stolperwörter in den Texten nicht identisch verteilt war. Im „Schlitzohr“-Text beispielweise bildete das zweite Stolperwort

(nämlich „Galgen“) das 63. von 102 Wörtern, während das zweite Stolperwort „Kilokalorien“ im „Croissant“-Text das 92. Wort ist.

Diese Varianz war gewollt. Sie sollte sicher stellen, dass sich bei den potenziellen Teilnehmenden kein Routine-Muster oder gar ein Lerneffekt in dem Sinne einstellen würde, dass die Probanden die Texte nicht mehr sinnentnehmend lesen, sondern in einem vagen und auf Basis der Erfahrung bereits rezipierter Experimentaltexte als „vielversprechend“ identifizierten Textbereich überfliegend nur nach dem Stolperwort „scannen“ würden.

4.7.1.7 Auswahl und Reihenfolge der Texte

Der Rückgriff auf verschiedene Experimentaltexte erfolgte, wie gezeigt, aus durchführungspraktischen Gründen. Unerfreuliche Randerscheinung dieser Entscheidung war jedoch, dass damit zu den eigentlich interessierenden typografischen Faktoren der vermutlich interagierende Faktor „Text“ getreten war.

Um diesen Störeinfluss zu minimieren, wurde, wie gezeigt, eine möglichst gleichförmige inhaltliche und stilistische Gestaltung der Texte angestrebt. Um darüber hinaus eine weitestmögliche Kontrolle über den Komplementäreinfluss der Variablen „Text“ zu gewährleisten, entschied ich, die Reihenfolge der darzubietenden Texte für alle Probanden identisch zu halten. Auf diese Weise war zumindest der Einfluss des vermuteten Lerneffekts, der sich im Verlaufe des Experimentaldurchgangs einer jeden Person ergeben und in erwarteten sinkenden Leseseiten manifestieren würde, auf die Messergebnisse zumindest als Störeinfluss kalkulierbarer, als dies der Fall gewesen wäre, wenn, zusätzlich zu den typografischen Variationen, auch die Abfolge der Textinhalte einer randomisierten Reihenfolge unterlegen hätten.

Bleib zu klären: Welche der neun weiter oben vorgestellten Texte empfahlen sich für die letztliche Verwendung im Experiment, und in welcher Reihenfolge? Dies versuchte ich in einem zweiten Pretest zu klären.

4.7.1.8 Der zweite Pretest

In Ablauf und Programmierung ähnelte der zweite Pretest, den ich im Verlaufe des Dezembers 2006 durchführte, bereits sehr der letztlichen Experimentalvariante. Wiederum bat ich Mitglieder einiger Foren, Bekannte und Kollegen sowie diesmal auch Studierende der Fachhochschule Gelsenkirchen, sich dem Vor-Experiment kritisch zu unterziehen. Enthalten im Experiment waren alle neun oben vorgestellten Texte; die Reihenfolge der Texte wie auch die typografische Konstellation, in der ein jeder Text angeboten wurde, wurde per Zufall generiert, lediglich der Übungstext behielt durchgehend seine „Pole Position“.

Insgesamt erhielt ich verwertbare Messdaten aus 68 Teilnahmedurchgängen. Ich gliederte die Resultate für die acht relevanten Texte (den Übungstext bezog ich nicht in die Auswertung mit ein) ab vor allem unter zwei Kriterien: der durchschnittlichen erzielten Lesedauer sowie der jeweiligen Standardabweichung der erhobenen Werte aus einem Text. Für das eigentliche Experiment wollte ich schließlich jene vier Texte auswählen, deren durchschnittliche Messzeiten einander am nächsten kamen – gleichzeitig achtete ich darauf, dass die internen Varianzen der Lesezeiten für ein und denselben Text möglichst gering ausfielen, da ich dies insbesondere als Indiz für eine stimmige Auswahl der Stolperwörter ansah, jedoch auch als Anhaltspunkt für eine intellektuell-stilistisch vergleichbare inhaltliche Textgestaltung.

Tabelle 4.4: Die Lesezeiten im 2. Pretest			
Text	n	Ø-Lesezeit	S
„Glocke“	64	36,57 sec.	20,69
„Olympia“	66	27,62 sec.	10,63
„Mönche“	66	30,04 sec.	10,55
„Schlitzohr“	64	28,07 sec.	8,38
„Kartoffeln“	65	31,30 sec.	11,97
„Hornberg“	65	29,15 sec.	9,67
„Ägypten“	63	28,94 sec.	9,30
„Croissant“	63	35,83 sec.	15,81
<i>Quelle: eigene Erhebung</i>			

Tabelle 4.4 zeigt das Ergebnis des zweiten Pre-Tests. Es erwies sich, dass – aus welchen Gründen auch immer – vor allem die Texte „Glocke“ und „Croissant“ den Testteilnehmern durchschnittlich deutlich mehr Mühe bereiteten als die übrigen. Insgesamt machten mir die in diesem zweiten Pretest erhobenen Daten die Aus-

wahl der letztlich vier eingesetzten Experimentaltexte relativ leicht; in Tabelle 4.4 sind jene Texte kursiv gesetzt, die ich unter den oben beschriebenen Kriterien letztlich auswählte für die Verwendung im eigentlichen Experiment. Es waren eben jene, die in den erhobenen Durchschnitts-Lesezeiten recht nahe beieinander lagen, dabei aber auch relativ vergleichbare Mess-Varianzen aufwiesen.

Wirklich verwertbare Daten übrigens im Sinne des Forschungsziels „Lesegeschwindigkeitsmessung“ erbrachte der Pre-Test übrigens nicht, konnte sie gar nicht erbringen; allein deswegen, weil keiner der vorgelegten Texte in allen 90 projektierten typografischen Testkonstellationen überprüft worden war.

4.7.2 Technische Ausgestaltung des Experiments

Da das Experiment bereits in einem sehr frühen Konzeptionsstadium als Online-Untersuchung angelegt war, stellte sich früh die Frage nach der bestmöglichen digitalen Plattform und programmiertechnischen Ausgestaltung der Studie.

Die Vorgabe war klar: Potenzielle Probanden im World Wide Web sollten in möglichst großer Anzahl in der Lage sein, das Experiment ohne größeren Aufwand, Wartezeit, tiefere Vorkenntnisse oder gar vorherige Installation von Spezialprogrammen an ihrem Rechner zu starten und durchzuführen. Gleichzeitig war sicherzustellen, dass die erhobenen Messdaten ihren zuverlässigen Weg in die Auswertungs-Datenbank finden würden.

4.7.2.1 Verworfen Alternative: autonome Programmdateien

Eine Möglichkeit war zunächst darin zu sehen, das Experiment programmiertechnisch auszugestalten in Form einer autonomen Programmdatei; in Gestalt einer Datei also, die per bekanntem Doppelklick aktiviert und abgespielt werden würde von den Teilnehmenden. Derlei autonome Dateien freilich, mithilfe welchen Programms auch immer konstruiert, fallen in der Regel nicht nur recht voluminös in der Dateigröße aus, sie hätten im vorliegenden Falle vor allem erst einmal den Weg auf die Festplatten potenzieller Probanden finden müssen.

Denkbar wäre gewesen als Modell, selbst startende oder auch per Doppelklick aktivierbare Dateien als E-Mail-Anhang in großer Menge an potenzielle Teilnehmende zu versenden. Dies aber hätte wiederum vorausgesetzt, dass die Empfänger den Anhang überhaupt erst einmal bewusst herunterladen – eine beachtliche Teilnahme-Schwelle, abgesehen davon, dass E-Mails mit umfangreicheren Anhängen (und vor allem solchen mit der für browserunabhängige Angebote charakteristischen Dateinamens-Endung „.exe“) häufig ungelesen im „Spam“-Filter von E-Mail-Programmen versanden. Eine Alternative war, die entsprechende Experimental-Datei zum Download auf einer Webseite bereit zu stellen; der Aufwand des Herunterladens wäre potenziellen Teilnehmerinnen und Teilnehmern aber auch in diesem Modell nicht erspart geblieben. Ohnehin führt diese Überlegung bereits zum naheliegendsten Szenario: Der Konstruktion des Experiments an sich als Webseite.

4.7.2.2 Verworfen Alternative: HTML/CSS

Da Webbrowser und die in ihnen zu verwirklichtbare Typografie nicht nur Gegenstand der Untersuchung als solche waren, sondern sicherlich auch als verbreitetste Programmgattung nicht nur auf deutschen Rechnern gelten dürfen, lag die Konstruktion des Experiments in Form einer „normalen“ Webseite nahe, aufrufbar über eine klassische URL („Uniform Resource Locator“, das ist die „Anschrift“ einer Webseite) im Adressfenster des Browsers und basierend auf einer HTML-/PHP-Programmierung. Zum Start des Experiments wäre dann nur noch die einmalige Eingabe einer Web-Adresse im Adressfenster des Browsers notwendig – oder die Aktivierung eines Links, der auf die Seite verweist, sei es in einer E-Mail, sei es auf einer anderen Website.

An dieser Stelle der Überlegung freilich treten wiederum die bereits ausführlich beschriebenen Unwägbarkeiten typografischer Konstruktion in Webbrowsern auf den Plan. Dass Betrachter von HTML-/CSS-basierten Webseiten beispielsweise die Schriftgrößen manipulieren können, ist, wie gezeigt, durchaus konzeptionelles Prinzip zeitgenössischer Webbrowser. Für das projektierte Experiment und dessen Ausgestaltung aber ergab sich nicht zuletzt aus dieser Eingriffsmöglichkeit ein Problem. Schließlich sollten die potenziellen Probanden ausdrücklich auch mit vermeintlich „unleserlich kleinen“ Schriftgrößen konfrontiert werden; und sie sollten diese Schriftgrößen eben nicht durch Manipulationen im Browser „korrigie-

ren“ können. Dass auch die kommunikatorseitige Vorgabe von Schriftarten zumindest im Browser *Firefox* aushebelbar ist, macht die Angelegenheit nicht einfacher.

Eine eindeutige und nutzerseitig unmanipulierbare Fixierung von Schriftgrößen und -arten auf HTML-basierten Webseiten ist eben im Grunde nur auf eine Weise verlässlich konstruierbar: indem der Text als Bild, als Bitmap im Format GIF, JPEG oder PNG, abgelegt und aufgerufen wird. Der Vorbereitungsaufwand dieser Lösung wäre allerdings schier unüberschaubar geworden. Schließlich sollten insgesamt fünf Texte in jeweils 90 Schriftanordnungen getestet werden; zur Einrichtung des Experiments wäre mithin die Erstellung von nicht weniger als 450 Bitmaps erforderlich gewesen. Innerhalb jedes dieser 450 Bitmaps wäre zudem – per HTML – die Definition zweier „sensitiver“, also anklickbarer Bereiche notwendig gewesen: jener Bereiche nämlich, in denen sich die beiden aufzufindenden Stolperwörter befanden. Doch auch damit wäre noch nichts erreicht gewesen: Schließlich sollte erst die Identifizierung des zweiten Stolperworts einen Fortgang im Untersuchungsprozedere und die eigentliche Zeiterfassung auslösen. Dies aber wäre mit reinen HTML/CSS-„Bordmitteln“ technisch nur sehr schwer realisierbar gewesen. Es hätte des Rückgriffs beispielsweise auf die Skriptsprache *JavaScript* bedurft, um eine befriedigende Lösung zu garantieren. Leider gilt für *JavaScript* in wesentlichen Teilen das selbe wie für HTML und CSS insgesamt: Browser stellen, je nach Hersteller, die Befehle dieser an sich sehr segensreichen Skriptsprache entweder gar nicht oder durchaus unterschiedlich dar. Der Einsatz von *JavaScript* hätte also erhebliche Risiken bezüglich der Verlässlichkeit der Messergebnisse mit sich gebracht.

So paradox es klingt: Für die Erforschung der Qualitäten verschiedener Schriftanordnungen auf HTML/CSS-Basis im Internet schied die Konstruktion eines Experiments auf HTML/CSS-Basis damit aus – zumindest, so weit Alternativen erkennbar waren. Und das war der Fall.

4.7.2.3 Die Alternative: *Adobe Flash*

Es bot sich durchaus eine Alternative: die Konstruktion des Experiments als *Flash*-Datei. Dateien dieses Typs garantieren über praktisch alle Browser-Grenzen hinweg ein so gut wie unverletzliches Erscheinungsbild eines einmal kreierten Designs⁸⁵¹, zudem: Schriften sind in *Flash*-Filme einbindbar⁸⁵², lösen sich also vom Diktat der „web-echten“ Schriften. Das verheißt potenziell volle typografische Kontrolle⁸⁵³. Was also sprach gegen diese Variante?

Vor allem dies eine: *Flash* ist, gelinde formuliert, umstritten. Ralf Siegel gelingt meines Erachtens eine prägnante Synopse der gängigen Vorbehalte gegenüber dem Programm, wenn er das Programm augenzwinkernd definiert als „(jenen) proprietären bunten Sack (...), der bei zappenden Zeitgenossen für leichte Kost und Pausenunterhaltung steht, den angeblich jeder

⁸⁵¹ Vgl. Maurice/Rex 2006, S. 264

⁸⁵² Vgl. Schweizer 2003, S. 162

⁸⁵³ Vgl. Weinman 2000, S. 267

aufmachen kann und der bei umstrittenen Druiden aus Übersee für hochgezogene Augenbrauen sorgt“⁸⁵⁴.

Zur Erläuterung sei einleitend bemerkt, dass Siegel mit dem „umstrittenen Druiden“ ganz augenscheinlich Jakob Nielsen meint, der sich bereits mehrfach vehement gegen die Verwendung von *Flash*-Filmen ausgesprochen hat⁸⁵⁵. Wie zu zeigen sein wird, steht der „Usability-Papst“ mit dieser Meinung nicht alleine.

Eindeutig nicht von der Hand zu weisen ist in jedem Falle, dass es sich bei *Flash* (beziehungsweise beim aus *Flash* heraus generierte Dateiformat „swf“) um ein sogenanntes „proprietäres“ Format handelt. Dieser Begriff bezeichnet Bild-, Datei- und Skript-Formate, die von kommerziellen Anbietern (in diesem Falle der US-amerikanischen Firma *Adobe*, die das Konkurrenzunternehmen *Macromedia* im Januar 2006 übernahm und dessen Produkt *Flash* dem *Adobe*-Portfolio einverleibte) aus Profitinteresse entwickelt und bereitgestellt werden und deren Quellcodes nicht oder nur rudimentär frei zugänglich sind. Die Anbieter dieser Formate handeln also, im Verständnis vieler Internet-Nutzer zumindest, dem offenen, demokratischen Charakter des World Wide Webs zuwider (Motto: „Das Internet gehört niemand“⁸⁵⁶). Wer einen *Flash*-Film erstellen will, muss in aller Regel das gleichnamige Programm käuflich erwerben – HTML dagegen ist, wie bereits gezeigt, ein prinzipiell „freies“ Format; wer eine Webseite auf HTML-Basis erstellen möchte, benötigt nicht mehr als einen ASCII-Texteditor, wie ihn jedes gängige Betriebssystem vorhält (z.B. *WordPad*, *Editor* oder *SimpleText*). Webdesigner auf HTML-/CSS-Basis zu werden, erfordert also auf technischer Basis nicht mehr als das Vorhandensein eines handelsüblichen PCs mit gängigem Betriebssystem – *Flash*-Produkte zu entwickeln, setzt die Eigentümerschaft eines Erstellungsprogramms voraus. Für die Verfechter eines demokratischen Netzes ist diese Hürde unverträglich. Ein ethischer Ansatz, das sei unbestritten.

Dass *Flash* nach wie vor vorrangig zur Erzeugung mehr oder minder sinnentleerter Web-Inhalte verwendet wird, dürfte Siegel zur Begrifflichkeit des „bunten Sacks“ geführt haben; auch die Unterstellung, *Flash*-Entwickler förderten vorrangig „leichte Kost“ und „Pausenunterhaltung“ zutage, wird durch einen Großteil von *Flash*-basierten Seiten keineswegs entkräftet. Auch Schweizer weist, bei aller Sympathie für *Flash*-basierte Websites, auf die vergifteten Verlockungen des Programms hin: „Die Möglichkeit, Dinge zu animieren, sind so vielfältig, dass der eine oder andere Entwickler dem Drang, möglichst viel Neues aus der Software herauszuholen, erliegt und die Inhalte auf der Strecke bleiben“⁸⁵⁷. Genauso sieht Hellbusch die Ära noch nicht beendet, da *Flash* „vor allem eine Spielwiese für Designer“⁸⁵⁸ war. Stocksmeier handelt *Flash* gar unter dem wenig schmeichelhaften Hauptkapitel „WWW-Zuckerguss“

⁸⁵⁴ Siegel 2002, S. 15

⁸⁵⁵ Vgl. Nielsen 2000b

⁸⁵⁶ Stocksmeier 2002, S. 15

⁸⁵⁷ Schweizer 2003, S. 163

⁸⁵⁸ Hellbusch 2005, S. 233

ab und warnt: „Solche Seiten sind erfahrungsgemäß sehr schlecht bedienbar, verwirren Besucher“⁸⁵⁹.

Freilich erliegt Stocksmeier hier – wie viele Zeitgenossen, mit denen ich die Ehre hatte, das Problem zu diskutieren – dem verbreiteten Fehler, optisches und Navigations-Konzept eines Web-Angebots in direkte Kausalverwandtschaft mit dem technischen Werkzeug zu setzen, in dem das Angebot in Bildschirmdarstellung übersetzt wurde. Dass aber das Programm *Flash* quasi aus sich selbst heraus jene wirren Navigationsstrukturen generiert, die zugegebenermaßen vielen *Flash*-Webseiten die Daseinsberechtigung rauben, ist eine geradezu abenteuerliche Behauptung. „Schlechte“ Webseiten mit rätselhaftem Navigationskonzept sind auch auf HTML-Basis erstellbar – und deren Anzahl dürfte die der *Chaos-Flash*-Seiten um ein mannigfaltiges überschreiten.

In diesem Zusammenhang können *Flash*-Befürworter den zitierten Kritikern fast schon dankbar sein, dass sie die exorbitante Nutzung von *Flash* für die Erstellung von Internet-Werbepannern nicht explizit erwähnen. Der Beliebtheit des Programms unter Internetnutzern ist die Marktführerschaft von *Flash* im Segment der digitalen Werbemittelerstellung sicherlich nicht förderlich.

Dass *Flash*, wie Siegel schreibt, „angeblich“ jeder aufmachen kann, ist eine Behauptung, die in ihrer restriktiven Diktion ebenfalls nicht vollkommen an der Wahrheit vorbei geht. Das Dateiformat „swf“, in dem *Flash*-Filme im Allgemeinen publiziert werden, ist kein vom W3C anerkanntes Dateiformat (wie übrigens auch „PDF“). *Flash*-Filme kann also nur betrachten, wer als interessierter Nutzer den *FlashPlayer* auf seinem Computer installiert hat – ein „Plug-In“, ein Mini-Programm also, das kostenlos von der *Adobe*-Webseite herunterladbar ist. Dieses Herunterladen ist allerdings heute kaum noch vonnöten, denn der *FlashPlayer* ist – obwohl, wie gesagt, kein W3C-Standard – seit Jahren bereits in allen gängigen Browsern vorinstalliert: „(Sie) können das Plug-In schon fast voraussetzen“⁸⁶⁰.

Dieses Plug-In kann allerdings – genau wie die Darstellung von Bildern, teilweise von CSS oder *JavaScript* auch – vom Nutzer in seinem Browser auch deaktiviert werden. In vielen größeren Firmen-Netzwerken ist dies der Fall – sei es, um die Belegschaft von Arbeitszeit verschlingenden Online-Spielen auf *Flash*-Basis fern zu halten, sei es, weil das *Flash*-Plug-In – zumindest auf Grundlage des Betriebssystems *Windows* – partout den Ruf eines potenziellen Einfallstors für Computerviren nicht los wird (diese durchaus mit Vokabeln eines Glaubenskampfes geführte Debatte zwischen Webdesignern und Systemadministratoren soll an dieser Stelle weder wiedergegeben noch inhaltlich vertieft werden).

Es kommt sogar noch etwas schlimmer: Das Vorhandensein des *Flash*-Plug-Ins im Browser an sich garantiert immer noch nicht das korrekte Abspielen eines Films. Auch die Version des

⁸⁵⁹ Stocksmeier 2002, S. 211

⁸⁶⁰ Schäffer 2001, S. 416

Plug-Ins muss stimmen. Schließlich entwickelt Hersteller *Adobe* das Programm *Flash* stetig weiter; Filme, die mit den jeweils neuesten Programmversionen erstellt werden, erfordern daher zum Abspielen in der Regel auch den jeweils aktuellsten *FlashPlayer*. In einem Browser aus dem Jahre 2000 laufen daher nicht unbedingt alle *Flash*-Filme, die im Jahr 2006 erstellt wurden. Wer in seinem Browser-Fenster also eine *Flash*-Datei abrufen, die in einer neueren *FlashPlayer*-Version bereitgestellt ist, als der Nutzer sie individuell auf seinem Rechner vorhält, wird aufgefordert, sich mit eben jener neueren Plug-In-Version zu versorgen (sie also herunterzuladen und zu installieren auf seinem Computer). Es ist dies ein nicht ganz unwichtiger Aspekt: Laut Fisch und Gscheidle nämlich folgen stolze 50 Prozent aller Online-User „nie“ einer Aufforderung, ein Programm herunterzuladen, um bestimmte Web-Inhalte betrachten zu können⁸⁶¹.

Dennoch sprach einiges für die Ausgestaltung des Experimental-Interfaces mithilfe von *Flash*:

- eine aktive Manipulation von Schriftgrößen, aber auch Zeilenabständen und -breiten, ist Nutzern in *Flash*-Publikationen praktisch nicht möglich⁸⁶². Was, wie gesagt, in klassischen Browsern durchaus als wünschenswert angesehen wird, ist hier ausgeschlossen.
- der klassische „Zurück“-Button in Browsern, der Probanden in einem HTML/CSS-basierten Experiment die Rückkehr zu einem bereits bearbeiteten Text ermöglicht hätte, entfaltet innerhalb von *Flash*-Filmen keine Wirkung; er führt höchstens zum Neustart des *Flash*-Films als Ganzem.
- *Flash* kann – entgegen landläufiger Vorstellungen – sehr kleine Dateien erzeugen und damit vor allem die Wartezeit des Online-Users angenehm kurz halten. Der *Flash*-Film, auf dem das Web-Experiment schließlich basierte, wies eine Gesamtgröße von gerade einmal 17,9 Kilobyte auf; selbst Nutzer von analogen 56K-Modems hatten damit das Experiment im Normalfall nach drei, höchstens fünf Sekunden Wartezeit startbereit – und dann auch absolut unterbrechungsfrei – auf dem Bildschirm. Zum Vergleich: ein einziges Bitmap von Text 1, gestaltet in der Schriftart *Verdana* in 12 Pixel Größe mit einem Zeilenabstand von 18 Pixeln und einer Zeilenbreite von 55 Anschlägen, nimmt im sparsamen Format GIF allein bereits 7,2 Kilobyte in Anspruch; der aufwändigere HTML-Code erhöht die Ladezeit zusätzlich.
- In *Flash* sind Seiten erstellbar, die – entgegen der Unkenrufe der versammelten Kritikerschaft – optisch genauso aussehen wie eine „typische“ HTML-basierte Seite. Das betrachtete ich als Chance: Die *Flash*-Gegner sollten möglichst gar nicht bemerken, dass sie mein Experiment in einem „verhassten“ Plug-In durchführten.
- *Flash* kann hervorragend mit der Skriptsprache PHP korrespondieren, die ihrerseits zuverlässig und sehr manipulationsresistent Datenbankeinträge (und, wie zu zeigen wird, noch einiges mehr) zu verwalten in der Lage ist.

⁸⁶¹ Vgl. Fisch/Gscheidle 2006, S. 437

⁸⁶² Vgl. Schweizer 2003, S. 162

Es stellte sich also die Frage: Wieviele potenzielle Teilnehmerinnen und Teilnehmer des Online-Experiments würden nicht an einem *Flash*-basierten Experiment teilnehmen können – sei es wegen fehlender oder veralteter *FlashPlayer*-Version oder deren Deaktivierung?

Die Zahlen zur Verbreitung des *FlashPlayers* auf den Browsern weltweit sind naturgemäß ähnlich variant wie alle anderen Web-Daten auch. *Flash*-Hersteller *Adobe* publiziert verständlicherweise recht optimistische Zahlen zur „Durchdringung“ von Browsern mit seinem Plugin. So stellte *Adobe* auf seiner Homepage eine „Verfügbarkeit von *Adobe Flash*“ auf europäischen Rechnern von 97,4 % für den *FlashPlayer 6* fest⁸⁶³ (für diese Version wurde das Online-Experiment publiziert). Andere Statistiken ergeben deutlich ernüchterndere Daten: Der Online-Statistik-Anbieter *WebHits* etwa ermittelte im Herbst 2005 eine Verfügbarkeit des *Flash-Players* von gerade einmal 64,1 % (ohne Versionsnummer)⁸⁶⁴. Angesichts dieser Zahlen, nimmt man sie vorläufig für voll, ist allerdings sehr fraglich, warum immer mehr Firmen viel Geld investieren in die Erstellung von *Flash*-basierten Bannern zur Online-Werbung (praktisch alle Kampagnen größerer Unternehmen auf Webseiten basieren heute auf *Flash*-Technologie). Ich persönlich vermute, dass insbesondere in größeren Firmennetzwerken die *Flash-Player* sehr flächendeckend deaktiviert sind (ich habe dies selbst bereits bei mehreren Kunden festgestellt); eventuell erklärt dies auch die extrem schlechten Zahlen von *WebHits*, die nämlich eventuell schwerpunktmäßig nicht auf privat genutzten PCs erhoben wurden, sondern auf „Bürorechnern“. Wie dem auch sei: Ich entschloss mich, allen Unsicherheiten zum Trotz das Projekt in Form eines *Flash*-Films zu realisieren.

Dabei hatte ich mich sehr rasch von der Erwartung und Hoffnung zu verabschieden, ich würde eine auf *Flash* basierende und für mein Vorhaben nutzbare Rohdatei auffinden in einem der diversen *Flash*-Diskussionsforen im Web, in denen ambitionierte Gestalter und Programmierer die Früchte ihrer Arbeit zur freien Nutzung und insbesondere Fortentwicklung bereitstellen. Einige „Code-Schnipsel“ der *Flash*-eigenen Skript-Sprache „*ActionScript*“ fand ich zwar, die sich als teilweise hilfreich erwiesen in der programmiertechnischen Ausgestaltung des Experiments. Insgesamt jedoch war früh klar: Ich würde die Realisierung meines Konzepts in *Flash*, die Konzeption der Experimentaldatenbank in *MySQL* und den Austausch der Messergebnisse zwischen *Flash*-Datei und Datenbank auf Basis der Skript-Sprache *PHP* eigenständig und ohne externe Hilfe besorgen müssen. Was ich letztlich auch tat.

4.7.2.4 Kriterien des Aufrufs in PHP

Eine der größten programmiertechnischen Herausforderungen dieses Experiments war zweifellos die Vorgabe, dass jede der 90 zu testenden Schriftkonstellationen sowie alle Testtexte insgesamt sich jeweils gleichverteilt in den erhobenen Messwerten wiederfinden sollten –

⁸⁶³ Vgl. Adobe Systems Inc. 2005

⁸⁶⁴ Vgl. Webhits Internet Design GmbH 2006

aber zusätzlich auch noch innerhalb der erfassten Daten eines jeden einzelnen Testtextes alle 90 getesteten typografischen Konstellationen in identischer Anzahl vorliegen sollten. Zudem sollte die Zuteilung von Konstellationen an die Teilnehmenden möglichst zufällig – randomisiert – erfolgen. Die Chancen eines jeden Probanden, jede beliebige der 90 Konstellationen im Verlaufe des Experiments virtuell vorgelegt zu bekommen, sollte immer noch so groß wie möglich sein.

Nach welchen Prinzipien – und unter Verwendung welcher technischer Methoden – waren also Probanden und Aufgaben im Experiment zusammenzuführen? Einerseits war eine möglichst zufällige Verteilung (Zuweisung) der einflussnehmenden Merkmale (also der zu messenden typografischen Konstellationen) an die Teilnehmenden in einem Zufallsverfahren⁸⁶⁵ erwünscht – andererseits würde die Zuweisung nicht gänzlich randomisiert zu verwirklichen sein, wenn die oben beschriebenen angestrebten Gleichverteilungen innerhalb des Gesamtergebnisses hergestellt werden sollten. Die Zuteilung musste also in Maßen „künstlich“ erfolgen. In verwandtem Sinne war also ein „Matching“⁸⁶⁶ der Versuchsteilnehmer erforderlich.

Im ursprünglichen Sinne bezeichnet „Matching“ in empirischen Forschungsdesigns die gezielte – und damit eben nicht mehr völlig zufällige – Zuordnung von Probanden und deren Messergebnissen in vorgegebene Gruppen nach definierten Kriterien⁸⁶⁷. Um beispielsweise eine soziodemografisch repräsentative Zusammensetzung von Teilnehmenden einer telefonischen Vor-Wahlbefragung zu gewährleisten, werden Forscher darauf achten, im zeitlichen Verlauf der Befragung ungefähr genau so viele Frauen wie Männer zu interviewen – weil die wahlberechtigte Bevölkerung sich hinsichtlich der Geschlechter ähnlich paritätisch zusammensetzt. Ist diese Parität gefährdet, weil sich in einem fortgeschrittenen Stadium der Befragung herausstellt, dass bislang deutlich mehr Männer als Frauen befragt wurden, werden die Forscher mehr Mühe auf die Akquirierung von weiblichen Teilnehmern verwenden; ab diesem Moment aber wäre keine reine Zufallsauswahl der Befragten mehr gewährleistet.

Das „Matching“ in meinem Experiment, so meine Überlegung, würde sich jedoch nicht auf die Befragten beziehen müssen, sondern auf die quantitative Gleichverteilungs-Situation unter den bislang erhobenen Messergebnissen.

Grundsätzlich in Betracht zur Lösung dieses Problems kam zunächst die *Flash*-eigene „Random“-Funktion. Diese erlaubt die gezielte Erzeugung von Zufallszahlen. Es wäre also möglich gewesen, automatisiert für jeden neu anzuzeigenden Text bezüglich Schriftart und -größe sowie Zeilenbreite und -abstand Zufallswerte erzeugen zu lassen und in anzuzeigende Schriftkonstellationen umzusetzen. Nach stochastischen Erwägungen wäre auf diese Weise absehbar eine Gleichverteilung der 90 Konstellationen unter den erhobenen Daten hergestellt worden.

⁸⁶⁵ Vgl. Zimbardo/Gerrig 2004, S. 35; vgl. ferner Brosius/Koschel 2005, S. 204f.

⁸⁶⁶ Klammer 2005, S. 279

⁸⁶⁷ Vgl. Mittenecker 1966, S. 26f.; vgl. ferner Kromrey 2006, S. 285f.

Als sonderlich exakt allerdings erwies diese Methode bereits nach wenigen Versuchen nicht. Daher entschied ich mich, einen etwas komplexeren Weg zu beschreiten. In welcher Schriftanordnung der jeweils nächste anstehende Text dem Probanden präsentiert wurde, entschied sich daher letztlich – und stets unmittelbar vor Freigabe des Textes – durch eine automatische, vom Teilnehmer nicht steuer- oder verhinderbare Aktivierung eines separat von mir entwickelten PHP-Skripts aus dem *Flash*-Film heraus. In diesem Skript wurde in einem vierstufigen Verfahren nach statistischen Verteilungs- und Zufalls-Kriterien die Schriftanordnung für den jeweils anstehenden Text ermittelt und an den *Flash*-Film rückübermittelt. Das Skript kam mithin bei allen Probanden, die den Übungstext sowie sämtliche vier Test-Texte durcharbeiteten, insgesamt fünf Mal zum Einsatz.

Die vierstufige Ermittlung der jeweils nächsten Schriftanordnung entwickelte sich im Skript dabei in dieser Abfolge:

Zunächst wurde, in einem ersten Schritt, durch Rückgriff auf die bis zur Skript-Aktivierung aufgelaufenen Datenbank-Einträge ermittelt, in welcher beziehungsweise welchen der 90 potenziellen Schriftanordnungen der anstehende Text allen bisherigen Teilnehmern insgesamt am seltensten präsentiert worden war. Dieses Kriterium traf dabei meist auf mehr als eine Schriftanordnung zu: Im Falle von Proband Nummer 1 beispielsweise war Text 1 notwendigerweise in allen 90 potenziellen Schriftanordnungen bislang 0-mal präsentiert worden. Die auf diese Art und Weise ermittelten Schriftanordnungen gingen in einen virtuellen Pool ein.

Dieser im ersten Schritt ermittelte Pool potenzieller Schriftanordnungen für den anstehenden Text wurde in einem zweiten Schritt um jene Schriftanordnungen bereinigt, mit denen der jeweilige Proband im bisherigen Verlaufe seines individuellen Durchlaufs bereits konfrontiert worden war.

Im dritten Schritt wurde ein zweiter Pool potenzieller Schriftanordnungen gebildet. In diesen gingen jene Anordnungen ein, die

- bis zum Zeitpunkt des Skript-Aufrufs insgesamt (also nicht, wie in Schritt 2, einzeltextbezogen) am seltensten zum Einsatz gekommen waren,
- und sich gleichzeitig in Pool 1 wiederfanden.

Aus Pool 2 potenzieller Schriftanordnungen schließlich wurde durch Aktivierung einer skript-internen Zufallsfunktion eine ausgewählt – diejenige, in der der anstehende Text schließlich dem Probanden präsentiert wurde.

Dieses vierstufige Verfahren stellte sicher, dass

- bis in die kleinste Erhebungseinheit „Schriftanordnung“ hinein eine zumindest näherungsweise perfekte Durchmischung der vier Texte gewährleistet war. Das oben beschriebene Skript stellte, wie beschrieben, sicher, dass, sobald eine Erhebungseinheit vier Einträge auf-

wies, keiner dieser vier Einträge aus der Bearbeitung zweier identischer Texte resultieren konnte. So war beispielsweise die Chance eines einzelnen Teilnehmers, Text 1 in der Anordnung „*Verdana* – 12 px – Zeilenabstand 18 px – Zeilenbreite 56 Anschläge“ vorgelegt zu bekommen, exakt gleich null, solange nicht die Texte 2 bis 4 in genau dieser Schriftanordnung von anderen Teilnehmern bereits in gleicher Häufigkeit wie Text 1 bearbeitet worden waren. Auf diese Weise wurden inhalts- oder erschöpfungsbedingte sowie durch die variierende Positionierung der Stolperwörter induzierte Unterschiede in den von den Probanden erzielten Bearbeitungszeiten der vier Texte innerhalb einer Erhebungseinheit zumindest ansatzweise neutralisiert.

- nicht nur die häufigkeitsmäßige Gleichverteilung der vier Versuchstexte innerhalb der kleinsten Erhebungseinheiten sicher gestellt wurde, sondern auch eine gleichmäßige Streuung der erfassten Daten über alle 90 Erhebungseinheiten hinweg. Der Vorteil dieser Vorkehrung liegt auf der Hand: Eine allzu ungleichmäßige Häufigkeits-Verteilung der Daten über die Erhebungseinheiten hinweg hätte die statistische Gültigkeit der datenmäßig weniger reich „bestückten“ Einheiten reduziert.
- jeder Proband jede Schriftanordnung im Verlaufe eines Durchgangs (und pro Proband war maximal ein Durchgang à vier Texte angestrebt) nur einmal präsentiert bekam. Auf diese Weise wurde zum einen verhindert, dass sich bei den Teilnehmern ein „Gewöhnungseffekt“ bezüglich der typografischen Anordnung der Texte einstellen konnte, der möglicherweise den Bearbeitungsprozess zeitmäßig begünstigt hätte: Mit jedem neuen Text mussten sich die Probanden auf eine neue, mehr oder minder unerwartete typografische Konstellation einstellen.

Zum anderen griff ich hier nochmals auf jene bereits erwähnte Erkenntnis aus dem ersten Pre-Test zurück, dass eine Präsentation der Texte in klar unterscheidbaren typografischen Anordnungen bei vielen Versuchsteilnehmern auch motivierenden Charakter entfalten würde; nämlich in dem Sinne, dass sich Sinn und Zweck der Untersuchung – die Erforschung unterschiedlicher typografischer Anordnungen – unmittelbar in den variierenden Schriftkonstellationen wiederfinden würde.

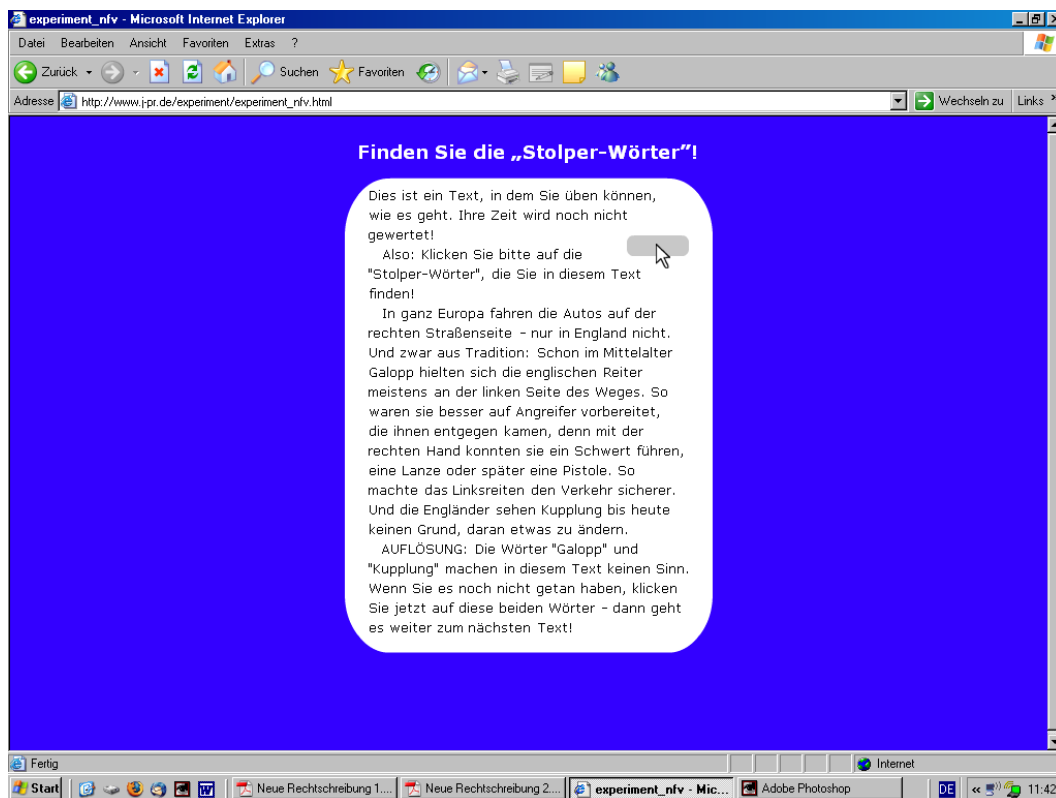
Kaum in seiner Wirkung messbar, aber nicht zu unterschätzen sein dürfte zudem der „Wachhalte-Effekt“, der bei den Probanden durch ein gewisses Maß an typografischer Abwechslung – zusätzlich zur inhaltlichen Varianz, die die vier Versuchstexte gewährleisteten – induziert werden konnte durch den Ausschluss identischer Textandordnungen innerhalb eines Erfassungsdurchgangs.

- der Einsatz der Zufallsfunktion in Schritt 4 des Skripts schließlich sorgte für jenen unbedingt gewünschten Zufallsfaktor, der bei aller „Vorsortierung“ der typografischen Varianten auch in meiner Studie wiederfinden sollte.

4.7.3 Die Programmierung

Die *Flash*-Datei samt zugehöriger MySQL-Datenbank und der relevanten PHP-Skripte hatte ich Ende November 2005 fertig gestellt und getestet (unter anderem auch im beschriebenen zweiten Pre-Test, in dessen Gefolge ich die letzten wichtigeren programmiertechnischen Unebenheiten in den relevanten Dateien beseitigte). Das Experiment konnte starten. Was an Ergebnissen herauskam dabei, schildere ich in den folgenden Kapiteln.

Das Aussehen des fertigen Experiments auf dem Bildschirm



Screenshot: Der Übungstext in der fertig programmierten Experiment-Version. An den Mauszeiger gekoppelt: eine halbtransparente Graufäche, mittels derer die Teilnehmenden die Stolperwörter zu identifizieren hatten.

ABBILDUNG 4.2

5 Allgemeine Ergebnisse

5.1 Eingesetzte statistische Methoden

Die Darstellung der quantitativen Ergebnisse meiner empirischen Untersuchung habe ich in zwei Kapitel unterteilt. Dabei liegt der Fokus dieses ersten der beiden Kapitel auf deskriptiver Datenbeschreibung. Das Folgekapitel 6 bedient sich hingegen schwerpunktmäßig der Methoden der Inferenzstatistik.

Während die deskriptive Statistik „nur“ Ergebnisse einer Erhebung bündelt⁸⁶⁸ – beispielsweise durch die Bildung von Mittelwerten oder die Ermittlung von Standardabweichungen –, geht die Inferenzstatistik (die auch als „schließende Statistik“ bezeichnet wird⁸⁶⁹) einen entscheidenden Schritt weiter: „Sie gibt Aufschluss darüber, welche Schlüsse (Inferenzen) aus den Stichproben ableitbar sind und welche Schlussfolgerungen man legitimerweise aus den Daten ziehen kann. Die Inferenzstatistik verwendet die Wahrscheinlichkeitstheorie um festzustellen, wie wahrscheinlich es ist, dass ein Datensatz durch Zufall entstanden ist“⁸⁷⁰. Dazu mehr im Kapitel 6.

5.2 Erhebungszeitraum

Die Online-Studie wurde im Zeitraum vom 1. Januar 2006 bis zum 28. Januar 2006 durchgeführt. Nach Abschluss des Experiments wurde die Webseite mit einem Passwort belegt und war mithin nur noch mir selbst zugänglich.

5.3 Akquirierung der Teilnehmenden

Akquiriert wurden die Teilnehmenden meiner Untersuchung im Wesentlichen über drei Kanäle:

- das „Flashforum“. In Abstimmung mit dem Betreiber dieses Online-Forums, Sascha Wolter, wurde am 1. und am 4. Januar 2006 in den Rubriken „Nachrichten“ und „Am Rande“ zur Teilnahme an der Untersuchung aufgerufen (die beiden Beiträge sind nach wie vor abrufbar unter <http://www.flashforum.de/forum/archive/index.php/t-188948.html> sowie <http://www.flashforum.de/forum/showthread.php?t=189202>).
- „heise online“. Im Web-Diskussionsforum des einflussreichen Verlags, der unter anderem das Fachblatt „c't“ herausgibt, wurde ebenfalls am 3. Januar zur Teilnahme aufgefordert (http://www.heise.de/extras/foren/go.shtml?read=1&msg_id=9569284&forum_id=44438).

⁸⁶⁸ Vgl. Kromrey 2006, S. 420

⁸⁶⁹ Vgl. Duller 2006, S. 215ff.

⁸⁷⁰ Zimbardo/Gerrig 2004, S. 63

Ein folgenreicher Fehler, wie sich rasch herausstellte. Zwar war auch zu mir das Gerücht vorgebracht, dass es sich bei den Stammlesern von „heise.de“ um ausgesprochen puristische, technikorientierte Menschen handle, die insbesondere „proprietäre“ Formate wie *Flash* im Internet geradezu aus Prinzip ablehnen – jene Bild-, Datei- und Skript-Formate also, die von kommerziellen Anbietern aus Profitinteresse bereitgestellt werden und deren Quellcodes nicht oder nur rudimentär frei zugänglich sind (vergleiche Seite 250). Zumindest war mein Beitrag nach fünf Minuten durch einen heise-„Aufseher“ in einen kaum aufgerufenen Unterordner abgeschoben – diese fünf Minuten hatten freilich ausgereicht, um ausreichend Forumsteilnehmer auf den Plan zu rufen, die mich mit einer Menge böser Anwürfe konfrontieren konnten. Die von mir daraufhin im selben Forum angestoßene Diskussion, warum *Flash* in dieser Nutzergemeinde so unbeliebt sei

(http://www.heise.de/foren/go.shtml?read=1&msg_id=9570363&forum_id=81866), erzeugte einen Rattenschwanz an Kommentaren, auf die an dieser Stelle nicht näher eingegangen werden soll. Festzuhalten bleibt lediglich: Insgesamt trugen Nutzer von „heise.de“-User nur sehr marginal durch Teilnahme zur Untersuchung bei.

- den Niedersächsischen Fußballverband (NFV), einen inzwischen langjährigen Kunden von mir. Der stellvertretende Verwaltungs-Direktor und Justitiar des Verbandes, Walter Burkhard, erklärte sich einverstanden damit, für das Experiment werbende E-Mails an all jene herauszugeben, die sich seit dem Januar 2002 an der traditionellen Online-Wahl zum „Fußballer des Jahres in Niedersachsen“ beteiligt hatten. In der betreffenden E-Mail wurde darauf hingewiesen, dass der NFV das Experiment ideell unterstütze und dass die Ergebnisse der Untersuchung nach Auswertung – und so weit verwertbar – in die anstehende Neugestaltung der Website des Verbandes einfließen sollten. In der Tat ist ein Relaunch der Site inzwischen angelaufen; Teilergebnisse dieser Arbeit wurden dabei im Gestaltungsprozess aktiv berücksichtigt und eingearbeitet. Technisch in Gänze umgesetzt, geschweige denn publiziert war der Entwurf zum Zeitpunkt der Fertigstellung der vorliegenden Arbeit jedoch noch nicht.

In der Summe war der Rücklauf auf diese Aufrufe sehr zufriedenstellend – auch angesichts der Tatsache, dass ich potenziell Teilnehmenden als motivationssteigerndes „Incentive“⁸⁷¹ lediglich interessante Erkenntnisse im Verlaufe des Jahres 2006 in Aussicht stellen konnte und nicht etwa Sach- oder gar Geldgeschenke. „Meine“ Teilnehmenden nahmen also durchweg freiwillig und ohne materielle Gegenleistung teil am Experiment.

5.4 Rücklauf

Es ist so gut wie unmöglich, für eine Online-Studie wie die hier diskutierte eine „Rücklaufquote“ zu bestimmen. Wie sollte dies auch funktionieren? Die vorliegende Studie wurde in

⁸⁷¹ Vgl. Couper/Coutts 2006, S. 226

zwei Web-Foren beworben sowie über eine E-Mail-Kampagne bekannt gemacht; und alle Teilnehmenden, die das Experiment bis zum Ende durchgeführt hatten, wurden auf der Abschluss- und Verabschiedungsseite gebeten, die Experiment weiterzuempfehlen (was offenkundig auch geschah). Jedoch gilt: „Wenn die Auswahlgesamtheit nicht identifiziert werden kann, ist das Ausmaß an Nonresponse schwer zu bestimmen“⁸⁷².

Nun lässt sich in den oben erwähnten Web-Foren prinzipiell nachvollziehen, wie viele Nutzer die Seite mit der Teilnahme-Empfehlung aufgerufen haben. Wie viele dieser Betrachter dann allerdings im Folgeschritt den Klick auf den Link zu meiner Studie riskierten, bleibt im Dunkel. Und wie viele derer, die die Seite aufriefen, dann das Experiment tatsächlich durchführten, ist schlichtweg nicht ermittelbar.

Eindeutig nicht identifizierbar ist gleichermaßen die Anzahl jener Nutzer, die von vorherigen Teilnehmenden per E-Mail ermuntert wurden, ebenfalls am Experiment teilzunehmen; zwar erreichten mich zahlreiche E-Mails, in denen meine Adresse zum Verteilerkreis gehörte und in denen das Experiment, augenscheinlich inspiriert durch meine Aufforderung zur Anwerbung Dritter, anderen Menschen empfohlen wurde – doch wie viele dieser Adressaten diese E-Mail überhaupt lasen, die Seite aufriefen und dann tatsächlich mitmachten, muss ungeklärt bleiben.

Fazit: Der Kreis jener, die im Prinzip Kenntnis von der Existenz dieser Studie hätten haben können und Zugang zu ihr hatten, ist nicht einmal annäherungsweise zu bestimmen.

Mit Einschränkungen bestimmbar ist dagegen der Rücklauf, macht man selbigen numerisch an den Aufrufen der Webseite fest, auf der das Experiment hinterlegt und aufrufbar war. Den Angaben des Web-Providers „1und1“ zufolge, auf dessen Servern das Experiment digital bereitgestellt („gehostet“) wurde, wurde die Seite etwa 1900 Mal aufgerufen innerhalb des Erhebungszeitraums. Nun sind derlei Web-Statistiken durchaus mit Vorsicht zu goutieren, da auch die ausgefeiltesten Erhebungsprogramme bis heute nicht in der Lage sind zu erfassen, ob beispielsweise eine Seite nach erstmaligem Aufruf im Arbeitsspeicher („Cache“) eines Rechners gespeichert wurde, danach ein zweites Mal von diesem Rechner aus ausgewählt, dann aber aus eben diesem Speicher und nicht vom Webserver direkt angefordert wurde.

Nimmt man jedoch die Angaben des Providers für bare Münze (dies soll der Einfachheit halber an dieser Stelle geschehen), so ist festzuhalten, dass das Experiment etwa 1875 Mal aufgerufen wurde (ich habe von der Provider-Angabe die schätzungsweise 25 Aufrufe subtrahiert, die ich selbst zur Statistik beigetragen habe durch Seitenaufrufe zur Überprüfung der Funktionstüchtigkeit des Programms).

⁸⁷² Couper/Coutts 2006, S. 223

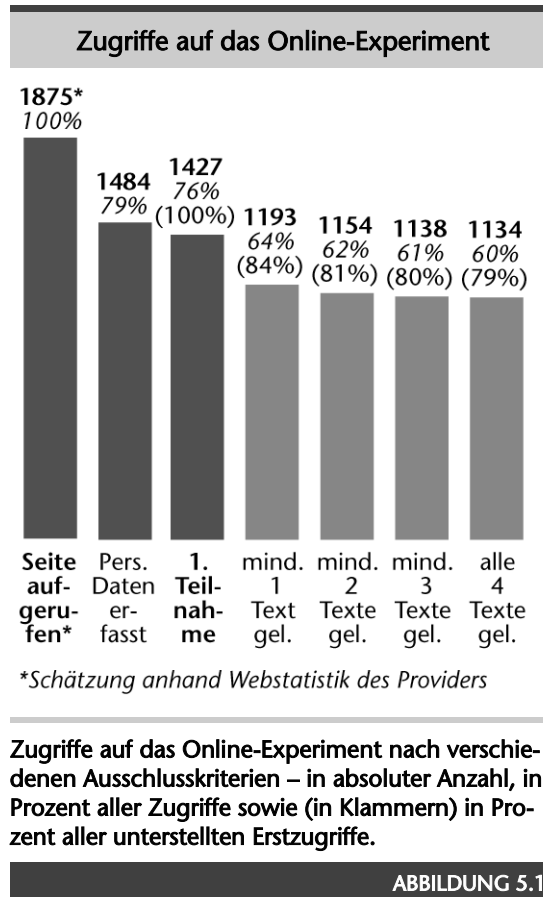
Wie der nebenstehenden Grafik zu entnehmen ist, wurden 1484 Mal persönliche Angaben gemacht (also die das Experiment einleitenden Fragen nach Alter und Monitorart beantwortet) und in der Datenbank gespeichert. Es existieren also 1484 Datenbank-Zeilen, in denen zumindest Alter und Monitorart festgehalten sind (in der Grafik: „Pers. Daten erfasst“).

Von diesen 1484 Angaben wurden 1427 als „Erstangaben“ gewertet; das heißt, 57 Einträge wurden identifiziert als von Nutzern stammend, die das Experiment bereits einmal durchgeführt hatten (in der Grafik „1. Teilnahme“).

Wie viele dieser 1427 Nutzer zumindest den Übungstext lasen und abschließend bearbeiteten, ist den erhobenen Daten nicht zu entnehmen. Ebenso wenig lässt sich aus den erhobenen Daten schließen, wie viele dieser Teilnehmer sich zumindest den ersten gemessenen Text digital „vorlegen“ ließen, ihn aber nicht bis zum zweiten Stolperwort durcharbeiteten.

Denn diese Daten erhob ich nicht aktiv; sie erschienen mir – bis heute – als zu wenig relevant.

So ist das nächste messbare (und, wie ich meine, auch relevante) Datum die Anzahl der Teilnehmenden, die als „Ersteilnehmer“ qualifiziert wurden und mindestens den ersten Text nach dem Übungstext im Sinne des Experiments bis zum Ende bearbeiteten – also beide Stolperwörter per Mausclick identifizierten. Dies waren 1193 Teilnehmende, mithin 84 % derjenigen, die als „Ersteilnehmer“ über den Übungstext hinausgekommen waren. Abbildung 5.1 verdeutlicht, dass diejenigen Teilnehmer, die diese Hürde genommen hatten, das Experiment in aller Regel nicht mehr abbrechen und bis zum 4. relevanten Text durchführten. 1134 von 1193 Teilnehmenden, die Text 1 erfolgreich bearbeitet hatten, hielten bis zum letzten Text durch – das entspricht einer Abbrecherquote von gerade einmal 4,9 % zwischen Text 2 und 4.



5.5 Strukturdaten der Teilnehmenden

Zu Beginn des Experiments wurden die Teilnehmenden, wie erwähnt, mit einer demografischen (Alter) und einer technischen (Monitorart) Einstiegsfrage konfrontiert. Zusätzlich wur-

de – ohne dass der oder die Teilnehmende aktiv Angaben tätigen musste – mithilfe einer *Flash*-Funktion, die systemische Auflösung des Bildschirms erfasst, an dem sich der Teilnehmende zum Zeitpunkt der Erfassung befand sowie das Betriebssystem, das zum Messzeitpunkt auf dem Rechner des Nutzers aktiviert war.

Auf die Abfrage weiterer persönlicher oder technischer Daten habe ich verzichtet. Dies vor allem aus zwei Gründen:

- Ich wollte, wie dargelegt, die Gesamtdauer des Experimentaldurchgangs möglichst kurz gestalten, um die Anzahl der Abbrüche möglichst gering zu halten.
- Ich wollte eventuellen Ängsten oder zumindest Vorbehalten potenzieller Teilnehmender vorbeugen, sie würden im Rahmen der Untersuchung über Gebühr in individuellen Merkmalen ausgeforscht. Je weniger persönliche Daten ich den Teilnehmenden abverlangen würde, so meine Überlegung, desto mehr Vertrauen seitens der Teilnehmenden bezüglich der vertraulichen Verwendung der erhobenen Daten würde mir geschenkt.

Aus diesen Gründen reduzierte ich die Eingangsfragen auf ein Minimum. Unter anderem erschien mir die „klassische“ Geschlechterabfrage im Rahmen dieser Studie irrelevant; auch Bildungsstand, Haushaltseinkommen oder Familienstand interessierten nicht vorrangig.

Durchaus von größerem Interesse wäre dagegen das allgemeine Medienkonsumverhalten der Teilnehmenden gewesen (wie oft und wie lange sie beispielsweise typischerweise im Web unterwegs sind und welche Arten von Seiten sie dort aufrufen – oder, generell, wie oft sie journalistische Texte rezipieren). Auch wäre durchaus die Frage stellbar gewesen, ob der individuelle Teilnehmende auf eine Sehhilfe angewiesen sei und diese zum Zeitpunkt des Experimentaldurchgangs nutze.

Ich habe letztlich auch von Fragen dieser Qualität abgesehen. Unter anderem wäre absehbar eine weitere Extra-Seite mit diesen Vorab-Fragen notwendig geworden – und zumindest das Medienverhalten hätte wahrscheinlich in einer eher komplexen Matrix erhoben werden müssen, die die Geduld der Teilnehmenden auf eine gewisse Probe gestellt und damit eventuell zu mehr Versuchsabbrüchen geführt hätte, als ich tatsächlich registrierte.

Die im Folgenden dargestellten und diskutierten Strukturdaten bezüglich Altersstruktur, Monitorart, Bildschirmauflösung und Betriebssystem der Teilnehmenden wurden nur aus einem ausgewählten Anteil der erfassten Daten gewonnen. In die Statistik einbezogen wurden dabei nur die Daten jener Teilnehmenden, die

- am Experiment zum ersten Mal teilnahmen und
- mindestens den ersten Text nach dem Übungstext abschließend bearbeiteten, also in diesem beide Stolperwörter per Mausklick identifizierten. Wie Abbildung 5.1 zu entnehmen ist, sind dies vermutete 1193 Probanden.

5.5.1 Altersstruktur

Zu Beginn des Experiments wurden die Teilnehmenden gebeten, ihr aktuelles Alter anzugeben. Als Alternativen zum Anwählen per Maus wurden angeboten: „unter 18 Jahre“, „18–35 Jahre“, „36–55 Jahre“ sowie „56 Jahre und älter“.

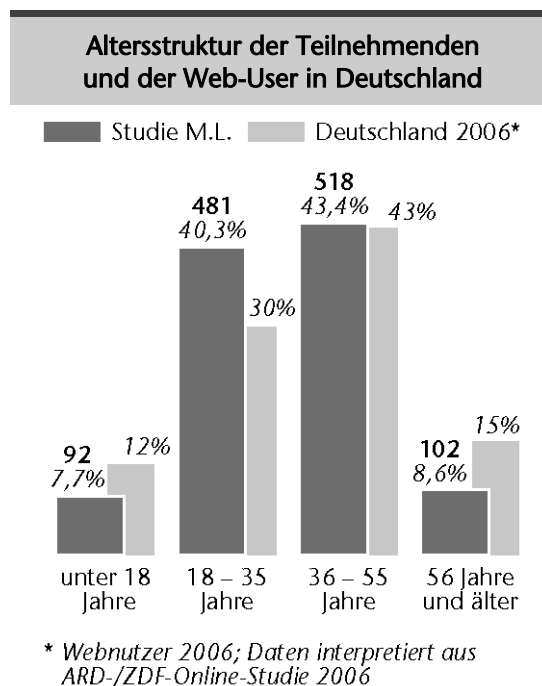
Das Alter fragte ich insbesondere zur groben Absicherung und Kontrolle der Repräsentativität der Studie zumindest unter einem demografischen Aspekt ab. Einen messbaren Einfluss des Alters der Probanden auf die Geschwindigkeitsmessungen vermutete ich dagegen nicht.

Mit der ersten Gruppe („unter 18 Jahren“) umriss ich allgemein die Gruppe der Minderjährigen. Gruppe 2 konstruierte ich aus der Überlegung heraus, dass es sich hier absehbar um eine Gruppe von Menschen handeln würde, die erwerbstätig oder in höherer Ausbildung befindlich sein würde – die vor allem jedoch das World Wide Web bereits als selbstverständlich existent oder im Alter von Anfang 20 als Innovation kennengelernt haben dürften. Hier würde es sich also um eine Gruppe handeln, so mutmaßte ich, die das Web durchaus schon länger als selbstverständlichen Teil des Alltags betrachten dürfte.

In der dritten Gruppe („36–55 Jahre“) vermutete ich dagegen bereits einen höheren Anteil von Menschen, für die das Aufkommen des Webs Anfang bis Mitte der 1990er Jahre einen Einschnitt gravierenderer Art bedeutet haben dürfte. Es würde sich nach meiner Überlegung um Menschen handeln, die das Web als Medium in einen Alltagsentwurf zu integrieren hatten, der zuvor von anderen Mediengattungen geprägt gewesen sein dürfte.

Die vierte Gruppe schließlich möchte ich, etwas unsauber zugegebenermaßen, als „Gruppe der Rentner und angehenden Rentner“ beschreiben – eine Gruppe, der nach wie vor eine erhebliche Berührungangst mit dem Web nachgesagt wird.

Die folgenden Erhebungsergebnisse zur Altersstruktur werden jeweils (in Klammern)



Altersangaben der Teilnehmenden, die beim ersten Durchgang mindestens Text 1 (also den ersten Text nach dem Übungstext) zu Ende gelesen haben (insgesamt und in Prozent). Zum Vergleich: Anteil der jeweiligen Altersgruppe an der Gesamtzahl der Internet-Nutzer in Deutschland.

ABBILDUNG 5.2

mern) mit dem Altersaufbau der bundesdeutschen Bevölkerung insgesamt abgeglichen⁸⁷³ sowie dem Anteil der Internet-Nutzer in Deutschland, die die jeweilige Altersgruppe stellt. Letztere Daten habe ich interpretiert aus den Ergebnissen der jüngsten ARD-/ZDF-Online-Studie⁸⁷⁴ (Stand August 2006). Da nach wie vor nicht von einer Volldurchdringung des Internets (und damit eben auch des Web) in bundesdeutschen Haushalten auszugehen ist (besagte Studie vermutet für 2006 36,7 Millionen Internet-User unter den knapp 71,7 Millionen in Deutschland Lebenden über 14 Jahren, das entspricht 51,1 Prozent), schienen mir diese Vergleichszahlen berücksichtigungswert. Die Daten der ARD-/ZDF-Studie resultieren aus Online- und Offlinebefragungen von 2664 bundesdeutschen Erwachsenen ab 14 Jahren und definieren als „Internet-Nutzer“ Personen, die im Verlaufe von vier Wochen vor dem jeweiligen Interviewtermin online gewesen waren. Da die Studie der Anstalten leider unter 14-Jährige nicht in die Befragung mit einbezog und zudem etwas andere Altersgruppen definierte, als ich dies in meiner Studie getan habe, sind die Vergleichsdaten mit ein wenig Vorsicht zu genießen (und werden daher im Folgenden auch nur in gerundeten Zahlen wiedergegeben).

Fast 84 % der Teilnehmer meiner Studie (51,5 % in der Gesamtbevölkerung, 73 % der Internetnutzer in Deutschland) waren – laut eigener Angabe – zum Zeitpunkt der Teilnahme zwischen 18 und 55 Jahre alt. Insgesamt gaben 40,3 % (20,2 % / 30 %) an, zwischen 18 und 35 Jahren alt zu sein, 43,4 % (31,3 % / 43 %) ordneten sich der Altersgruppe „36 – 55 Jahre“ zu. 7,7 % (17,7 % / 12 %) waren laut Eigenauskunft jünger als 18 Jahre, 8,6 % (30,8 % / 15 %) 56 Jahre oder älter.

Wie sind diese Daten zu interpretieren? Zunächst zeigte sich – im Vergleich zur ARD-/ZDF-Studie – eine eindeutige Überrepräsentation der 18- bis 35-Jährigen in meiner Untersuchung. Im Abgleich der nächsthöheren Altersgruppe ergab sich dagegen eine echte „Punktlandung“: Der Anteil der 35- bis 55-Jährigen unter den Online-Nutzern ist deutschlandweit fast exakt so groß, wie er in meiner Studie ausfiel. Die Altersgruppen der Minderjährigen und der Älteren dagegen fanden sich in meiner Studie unterdurchschnittlich repräsentiert. Eine Erklärung für diese Ergebnisse könnte sein, dass die ARD-/ZDF-Studie den Typus des „Online-Users“ sehr großzügig umreißt: Wer nur ein einziges Mal pro Monat online geht, ist laut der Studie noch ein „User“. Nun zeigt die Studie zwar, dass Internet-Nutzer unabhängig vom Alter durchschnittlich tatsächlich durchschnittlich etwa 4,8 Tage pro Woche online gehen – bemerkenswert ist aber, dass die (in meiner Studie überrepräsentierte) Gruppe der unter 30-Jährigen deutlich längere Online-Verweildauern pro Sitzung aufweisen (150 Minuten) als die 30- bis 49-Jährigen (116 Minuten) und die über 50-Jährigen (89 Minuten)⁸⁷⁵. Die höhere Nutzungsintensität der jüngeren, aber volljährigen Internet-Nutzer könnte also erklären, warum die Gruppe der 18- bis 35-Jährigen in meiner Studie überproportional vertreten war.

⁸⁷³ Werte siehe Anhang. Telefonisch angefordert und als „Excel“-Tabelle per E-Mail erhalten am 23.08.2006 vom Statistischen Bundesamt, Wiesbaden

⁸⁷⁴ van Eimeren/Frees 2006 S. 404

⁸⁷⁵ Vgl. van Eimeren/Frees 2006 S. 412

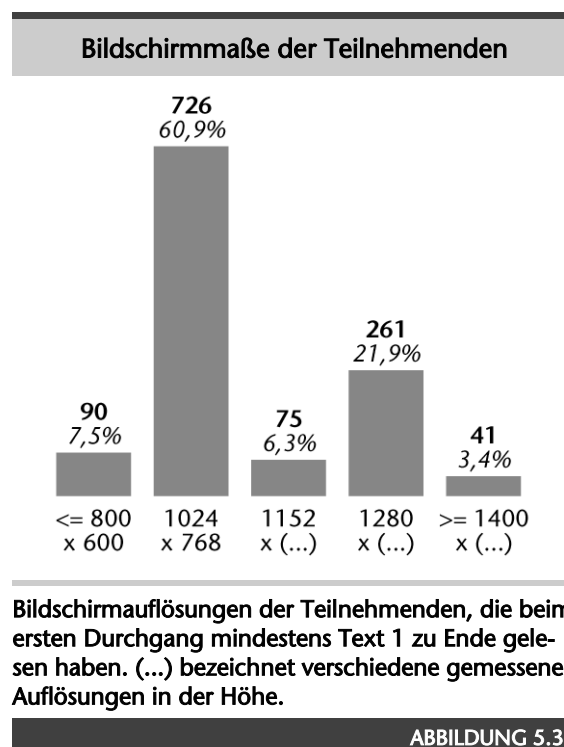
5.5.2 Bildschirmmaße

Die Auswertung der Bildschirmmaße der Teilnehmer erfolgte zeitgleich mit dem Aufruf der Webseite mit dem Experiment automatisiert und ohne Zutun der Probanden über eine Spezialfunktion in *Flash*. Die Ergebnisse wurden per PHP an die Datenbank ausgeliefert.

Die Ergebnisse sind für Webdesigner durchaus interessant. Denn entscheidet man sich gegen Jakob Niensens Dogma des „liquid designs“ und entschließt man sich, der zu gestaltenden Website eine fixe Breite zuzuordnen, so stellt sich die Frage: wie schmal ist eine Seite anzulegen, um einem größtmöglichen Betrachterkreis zumindest das lästige horizontale Scrollen zu ersparen?

Fakt scheint zu sein: Bildschirmbreiten von 640 Pixeln – das „Urmaß“ der ersten Webdesigner aus der späten Mitte der 90er Jahre – sind mit den 14-Zoll-Monitoren scheinbar so gut wie ausgestorben. In meinem Experiment jedenfalls kam dieses Maß nicht mehr vor, in einem einzigen von 1193 Fällen wurde eine Auflösung von 720 x 480 Pixeln festgestellt, in 89 Fällen eine Auflösung von 800 x 600 Pixeln – jene Breite also, die bis heute weitgehender Standard im Webdesign ist. Zu unrecht – nimmt man die Ergebnisse der hier diskutierten Studie zum Maß. Denn 92,5% aller Teilnehmenden saßen vor einem Bildschirm mit mindestens 1024 Pixeln systemischer Auflösung in der Breite. Zieht man allerdings in Betracht, dass durchaus einige Nutzer im Internet Explorer die so genannte „Favoritenleiste“ am linken Bildschirmrand beim Surfen dauerhaft geöffnet haben, spricht wiederum einiges für die Beibehaltung der „klassischen“ Breite „800 x 600“.

Was freilich das Entscheidende war: Die Wahrscheinlichkeit, dass ein Teilnehmender die Experimentaltexte nicht „auf einen Blick“ auf seinem Monitor zu sehen bekam, sondern vertikal zu scrollen hatte während des Rezeptionsvorgangs, war, wie spekuliert und erhofft, verschwindend gering in meiner Studie. Denn dieser unerwünschte Fall war ausschließlich gegeben, wenn ein Proband mit einem niedrig aufgelösten Bildschirm (unter 1024 x 768) mit einem Text in sehr großem Zeilenabstand und gleichzeitig sehr schmaler Zeilenbreite konfrontiert war. Auf Grundlage der obigen Daten schätze ich, dass im Verlaufe des gesamten



Experiments etwa 3 Prozent aller Teilnehmenden ein- oder mehrmals gezwungen gewesen sind, zu scrollen – eine vernachlässigbare Größenordnung, wie ich finde.

5.5.3 Betriebssysteme

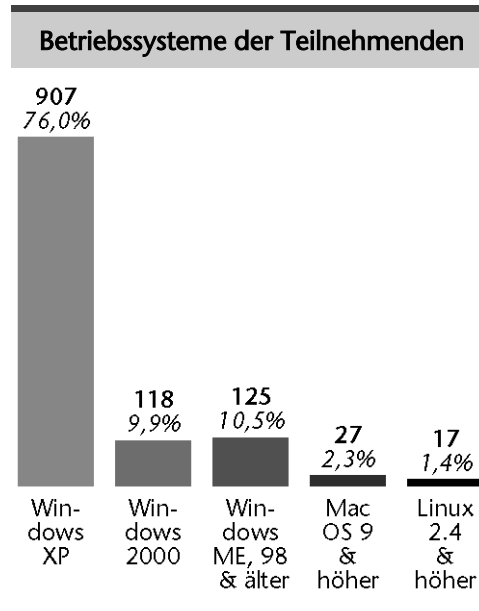
Ebenfalls automatisch und ohne Zutun der Probanden wurde während des Anmeldeprozesses das Betriebssystem ermittelt, auf dessen Grundlage der jeweilige Proband seinen Browser und darin das Experiment aktiviert hatte.

Das Ergebnis überrascht höchstens in seiner erdrückenden Eindeutigkeit: 96,4 Prozent der Teilnehmenden hatten ein *Windows*-System der Firma *Microsoft* aktiviert, als sie an dem Experiment teilnahmen. Die vermeintlich schärfsten Konkurrenten, das Betriebssystem OS der Firma *Apple Macintosh* und das Open-Source-Produkt *Linux* in verschiedensten Versionsstufen, verbuchten mit zusammen gerade einmal 3,7 Prozent einen sehr geringen Anteil.

Immerhin decken sich diese Ergebnisse recht weitgehend mit den aktuellen Statistiken der bereits zur Sprache gekommenen Firma *WebHits*⁸⁷⁶; zwar ist in deren Auswertungen das Betriebssystem *Windows XP* etwas schwächer, die Vorläuferversionen dagegen sind etwas stärker repräsentiert. In der Summe jedoch erfasste *WebHits* alle *Windows*-Systeme mit einem Anteil von 94 Prozent aller Fälle – also nahezu deckungsgleich mit den Anteilen meiner Studie.

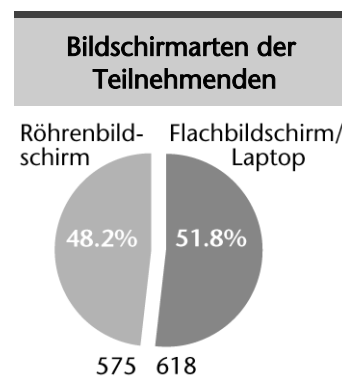
5.5.4 Bildschirmarten

Die Angaben der Probanden zur Art des Bildschirms, an dem sie das Experiment durchführten, bestätigt die Vermutung, dass wir uns an der Schwelle einer grundlegenden technischen Neuerung befinden. Eine knappe Mehrheit von 52 Prozent der Teilnehmenden gab an, vor einem Laptop oder einem Flachbildschirm zu sitzen,



Bildschirmarten der Probanden, die beim ersten Durchgang mindestens Text 1 zu Ende gelesen haben (absolut und in Prozent).

ABBILDUNG 5.4



Bildschirmarten der Probanden, die mindestens Text 1 zu Ende gelesen haben.

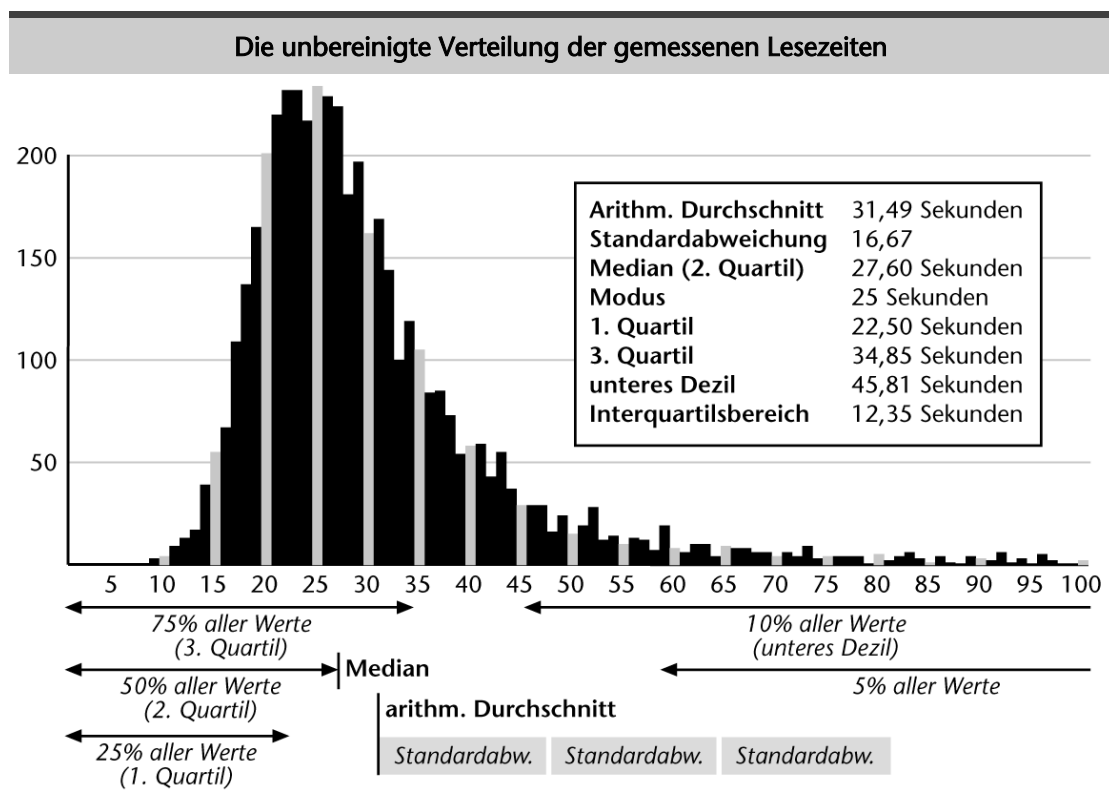
ABBILDUNG 5.5

⁸⁷⁶ Vgl. Webhits Internet Design GmbH 2006

„nur“ 48 Prozent gaben an, sie säßen vor einem „anderen“ Monitor – was nahe legt, dass diese Menschen auf einen klassischen Röhrenbildschirm blickten.

Dem Flachbildschirm scheint sich also auf einem Siegeszug zu befinden und dürfte die alt-hergebrachte „Röhre“ wohl bald Verängt haben. Auf diesen Aspekt wird im späteren Verlaufe dieser Arbeit noch detaillierter einzugehen sein.

5.5.5 Ein erster Blick auf die gemessenen Zeiten



Dieses Histogramm gibt wieder, wie sich die gemessenen Lesezeiten aufgliederten. Es zeigt, dass beispielsweise 201 Zeiten gemessen wurden, die sich im Bereich zwischen 20,000 und 20,999 Sekunden bewegten. Zeiten über 100,999 Sekunden wurden nicht in diese Grafik aufgenommen.

ABBILDUNG 5.6

Das Histogramm in Abbildung 5.6 verdeutlicht, in welchen Häufigkeiten sich die gemessenen Lesezeiten über das Gesamtergebnis verteilen. Die erhobenen Zeiten wurden dabei in 1-Sekunden-Schritten gebündelt und in der Häufigkeit ihres Auftretens abgetragen. Unter anderem ist ablesbar, dass Messwerte zwischen 21 und 28 Sekunden am häufigsten registriert wurden (der Messwert „25“, mithin erhobene Zeiten im Korridor zwischen 25,000 und 25,999 Sekunden, trat alleine 241 Mal auf). Erkennbar ist jedoch auch, dass sich die Daten eindeutig nicht normalverteilen im Sinne einer Gaußschen Kurve; allein Median, Modus und arithmetisches Mittel differieren doch allzu stark⁸⁷⁷.

⁸⁷⁷ Vgl. Zimbardo/Gerrig 2004, S. 63

Das Histogramm verdeutlicht mithin, dass die gemessenen Lesezeiten leicht rechts-schief verteilt waren. Grund sind die doch überraschend zahlreichen „Ausreißer“-Werte im oberen Lesezeiten-Bereich, die dafür sorgen, dass der arithmetische Durchschnitt (31,49 Sekunden) deutlich über dem Median der Werte (27,60 Sekunden) liegt.

5.6 Notwendigkeit einer Datenbereinigung

Bereits früh im Verlaufe des Experiments zeichnete sich ab, dass nicht alle erhobenen Lesezeiten in die End-Auswertung würden einbezogen werden können. Denn ein nicht unerheblicher Anteil der erhobenen Messwerte war augenscheinlich nicht Ausfluss eines unterstellten „normalen“ Leseprozesses, einer alltagsähnlichen Rezeption der vorgelegten Texte durch die Probanden.

Augenfällige Absonderlichkeiten offenbarten sich bereits bei der Betrachtung einzelner erhobener Lesezeiten – unabhängig von Alter, Schriftanordnung oder Textart. So betrug die durchschnittliche Lesegeschwindigkeit der Probanden 31,49 Sekunden pro Text oder 3,24 Wörter pro Sekunde. Da fällt es nicht leicht, Lesezeiten der doppelten durchschnittlichen Dauer von 63 Sekunden und länger – die immerhin noch 4,1 % aller erhobenen Fälle ausmachen – als Resultate eines „normalen“ Lesens zu werten. Diese Werte wären immerhin rein rechnerisch gleichzusetzen mit einer Lesegeschwindigkeit von 1,6 Wörtern pro Sekunde.

Ins Auge stachen auch jene Messergebnisse, die unter wahren „Klick-Orgien“ zustande gekommen waren. So „identifizierten“ einzelne Probanden die in den Texten versteckten Stolperwörter schlicht dadurch, dass sie offenkundig Wort für Wort anklickten; ein Proband trieb dieses „Trial-and-Error“-Prinzip mit atemberaubenden 284 Klicks innerhalb eines einzigen Textes zu fragwürdiger Meisterschaft. Ein Teilnehmer brüstete sich mir gegenüber sogar per E-Mail, diese vermeintliche „Lücke im System“ als solche entlarvt und die Texte schlichtweg inhaltlich nicht mehr zur Kenntnis genommen zu haben (meine E-Mail-Replik, dass er das Experiment dann leider umsonst durchgeführt habe, blieb unbeantwortet).

In diesen Fällen war die experimentelle Simulation eines alltagsähnlichen Lesevorgangs mithin augenscheinlich gescheitert; eine Berücksichtigung der Daten verbot sich daher im vorliegenden Versuch, der ja erklärtermaßen die Klärung der „Alltags-Tauglichkeit“ verschiedener Schriftanordnungen zum Ziel hatte. Es ergab sich also die Notwendigkeit, verschiedene erhobene Messwerte aus der Wertung zu nehmen. Eine Datenbereinigung vorzunehmen, war unumgänglich.

Im Folgenden sollen die Prämissen und Ergebnisse dieses „data minings“ erläutert und vorgestellt werden.

5.6.1 Kriterium Mehrfachteilnahme

Ein vergleichsweise einfacher Prozess war die Datenbereinigung nach dem Kriterium „Ersteilnahme“. Ich habe versucht, so weitgehend wie möglich sicherzustellen, dass ein und dieselbe Person nicht mehrfach an dem Experiment teilnahm. Es stand nicht nur zu befürchten, dass sich bei Probanden in einem zweiten Durchlauf bereits ein merklicher Trainingseffekt bezüglich der Messmethodik eingestellt haben würde, der die Lesezeiten nach unten beeinflusst hätte. Viel gewichtiger war die Annahme, dass sich Teilnehmende an die vorgelegten Texte und insbesondere an die in ihnen platzierten Stolperwörter erinnern und diese im zweiten Versuch gezielter suchen und schneller auffinden würden. Dieser verfälschende Effekt war naturgemäß unerwünscht.

So wurde zum einen zu Beginn des Experiments – für die Teilnehmenden unbemerkt – die IP-Adresse des Rechners abgefragt, an dem sie sich gerade befanden, und in der Datenbank hinterlegt. Zum anderen wurde auf dem Rechner eines jeden Probanden sowie in der Datenbank nach Eingabe der personenbezogenen Daten ein „Flash-Cookie“ hintergelegt; eine sehr kleine Datei, die lediglich die folgenden Informationen enthielt:

- Anzahl bisheriger Aufrufe des Experiments von diesem Rechner aus,
- Dauer vom Start des *Flash*-Films bis zur Eingabe aller personenbezogenen Daten in Millisekunden bei Erstaufwurf (im Folgenden als „Anlaufdauer“ bezeichnet).

Die Erfassung dieser drei Einzeldaten auf dem Rechner des Teilnehmenden wie in der Online-Datenbank erwies sich als ausgesprochen nützlich im Prozess der Datenbereinigung. So konnte ich über die Anzahl Aufrufe eine übersichtliche Vorauswahl derer treffen, deren Messergebnisse für eine Aussonderung infrage kamen. Die IP-Adressen sowie insbesondere die Anlaufdauer erlaubten in einem weiteren Schritt eine genaue Zuordnung der Durchgänge zu einem Teilnahme-Rechner. Im Zuge dieses Feinabgleichs zeigte sich beispielsweise, dass ein nicht unerheblicher Anteil Probanden offenkundig zunächst das Experiment aufrief, ohne überhaupt bis zum ersten Mess-Text vorzudringen; dass diese Probanden aber später einen (als eigentlich unerwünschter Durchgang „2“ erfassten) ernsthaften Aufruf folgen ließen, in denen sie das Experiment tatsächlich durchführten. In diesen Fällen (insgesamt 42 mit 156 Messergebnissen) deklarierte ich die Zweit- zu Erstdurchgängen um und löschte die unverwertbaren Erstdurchgänge aus der Datenbank. Ebenso verfuhr ich mit erfassten Dritt-, Viert- und weiteren Durchgängen.

Es ist freilich nicht auszuschließen, dass im Zuge dieser Bereinigung auch „ernsthafte“ Lese-durchgänge durch den Rost gefallen sind. Wer beispielsweise das Experiment im Erstdurchgang absolvierte und anschließend – in bester Absicht – zum Beispiel Lebensgefährtin, Gatten oder Kinder an den selben Computer bat und die Hinzugerufenen ebenfalls zur Teilnahme anhielt, tat dies umsonst. Denn diese weiteren Durchgänge wurden als Wiederholungen registriert und ausgesondert.

Umgekehrt ließ sich das Kontrollsystem natürlich umgehen, indem man beispielsweise am firmeneigenen Rechner den Erstversuch startete und am Abend, daheim, das Experiment unter der offiziellen Durchgangsziffer „1“ nochmals durcharbeitete.

Nun ist die Erfassung von Daten – insbesondere jener, die ohne Wissen der Teilnehmenden erhoben und gespeichert werden – ein zu Recht viel diskutiertes Thema. Ich möchte daher an dieser Stelle noch folgende Anmerkungen nachholen:

- außer mir selbst hatte und hat niemand Zugriff auf die Datenbank.
- unter den von mir erfassten Daten der Teilnehmenden war die IP-Adresse fraglos die heikelste. Mit gewissem Aufwand und unter Zugriff sowohl auf die von mir erfassten Daten als auch auf die Daten bestimmter Provider wäre es prinzipiell möglich festzustellen, welcher Vertragspartner wann auf das Experiment zugriff und wie weit er oder sie es durchführte. Ob es sich dabei um den Vertragspartner selbst handelte, ein anderes Haushaltsmitglied oder einen Gast, ist den Daten dagegen nicht zu entnehmen. Die IP-Adressen wurden nach dem Prozess der Datenbereinigung (in dem sie sich, wie gezeigt, als sehr nützlich erwiesen) sofort und unwiderruflich aus der Datenbank gelöscht.
- das „Flash-Cookie“ enthielt lediglich die beiden angesprochenen Daten „Durchgangs-Ziffer“ sowie „Anlaufdauer“. Ein Zugriff auf diese „Cookies“ ist ausschließlich mithilfe der von mir verfassten Experimentaldatei möglich; zudem dürften sie inzwischen, da ich sie mit einem „Verfallsdatum“ von acht Wochen belegte, von so gut wie allen Rechnern verschwunden sein, von denen aus das Experiment durchgeführt wurde.

5.6.2 Kriterium Fehlklicks

Wie gezeigt, wurde im Verlaufe des Experiments nicht nur für jeden einzelnen Text die Zeit zwischen dem Erscheinen des Textes auf dem Bildschirm und dem Anklicken des jeweils zweiten Stolperworts festgehalten, sondern auch die Anzahl der Klicks insgesamt erfasst, die ein Proband für die Bearbeitung des jeweiligen Textes tätigte.

Es sei daran erinnert, dass die Stolperwörter in den vorgelegten Texten lediglich als Vehikel zur Identifizierung der Qualitäten von Schriftanordnungen dienten, ihr Auffinden also nicht Selbstzweck der Untersuchung war, sondern Messmethode. Dass ein durchaus erheblicher Anteil von Probanden bei einzelnen Texten Schwierigkeiten bei der bloßen Identifikation der Stolperwörter hatte, muss nicht, kann aber durchaus auf einen grundsätzlichen Defekt der Messmethode hindeuten. In diesem Falle wäre nicht weniger als die Reliabilität der gesamten Untersuchung in Frage gestellt. An späterer Stelle soll daher eingehender auf mögliche Stör- oder gar Konzeptionsfehler der Versuchsanlage eingegangen werden.

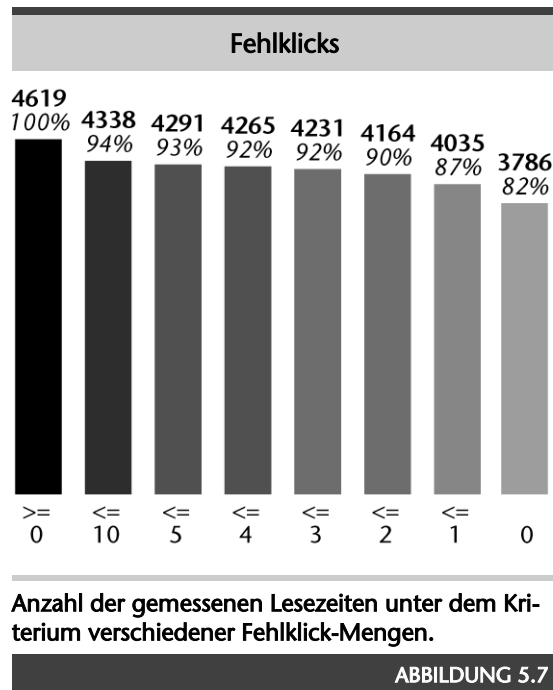
Geht man dagegen zunächst einmal von der grundsätzlichen Tauglichkeit des verwendeten Mess-Instrumentariums aus – und diese Annahme ist Grundlage der folgenden Überlegungen – so lohnt ein detaillierterer Blick auf die Ergebnisse unter eben dem Kriterium „Anzahl der Fehlklicks“.

Betrachtet man die Untersuchungsergebnisse im Lichte dieses Einzelkriteriums, so zeigt sich, dass ein erheblicher, von mir in diesem Ausmaße nicht erwarteter Anteil der Lesezeiten nicht ohne Komplikationen zustande kam. Durchschnittlich wurden pro Einzeltext beachtliche 3,93

Fehlklicks registriert (im zweiten Text sogar 5,09), 18 % aller erhobenen Zeiten gingen erst in die Wertung ein, nachdem die Probanden zuvor fehlerhaft auf mindestens ein „Nicht-Stolperwort“ geklickt hatten. Beachtliche 10 % aller Ergebnisse kamen erst nach mehr als zwei „Fehlklicks“ zustande, knapp 6 % aller erhobenen Zeiten weisen sogar mehr als 9 Fehlklicks aus. Der Fehlklick-„Spitzenwert“ lag für den ersten Text bei 184, für den zweiten sogar bei 284, für den dritten bei 144 und für den vierten bei 173.

Einzelne Probanden verfolgten augenscheinlich die Strategie, schlichtweg alle Wörter einzelner Texte nacheinander anzuwählen und auf diese Weise die Stolperwörter zu identifizieren – eine Umgehungsmethodik, mit der in der Konzeptionsphase der Untersuchung durchaus zu rechnen gewesen war und in deren Erwartung (allerdings, wie erwähnt, nicht in dieser quantitativen Ausprägung) die Erfassung der Fehlklicks überhaupt Eingang in die Auswertung gefunden hatte. Andere Teilnehmerinnen und Teilnehmer praktizierten die „Trial-and-Error-Methode“, so ist zu vermuten, aus einer gewissen Verzweiflung heraus, nachdem sie eines oder gar beide Stolperwörter auch nach mehrmaligem Ansatz nicht aufgefunden hatten.

Doch welche Gründe auch immer diese Probanden bewegten, sich gleichsam blind tastend auf die Suche nach den Stolperwörtern zu begeben; sie verzichteten offenkundig und aktiv auf den Prozess des inhaltsbezogenen Lesens, der immerhin Messziel der Untersuchung war. Die Messergebnisse dieser Teilnehmer waren daher zumindest teilweise qualitativ unbrauchbar geworden. Es spricht daher einiges dafür, erhobene Zeiten mit frapierend hohen Fehlklick-Raten in der Auswertung nicht zu berücksichtigen.



So nahe liegend jedoch die Disqualifikation dieser Lesezeiten als Prinzip ist, so schwierig gestaltet sich die Praxis des Ausschließens. Was darf in diesem Kontext als „viel“ gelten, als „zu viel“? Dass 284 Fehlklicks beim Auffinden zweier Stolperwörter in Texten von 102 Wörtern Länge gegen einen traditionellen Leseprozess sprechen, nehme ich als konsensfähig an. Doch immerhin kam, wie gezeigt, fast ein Fünftel der Messergebnisse nicht ohne Fehlklick zustande – von den übrigen 82 Prozent gemessenen Lesezeiten jedoch weisen wiederum 92 Prozent gerade einmal einen oder zwei Fehlklicks auf; dies sind aber sicherlich zu wenige Fehlklicks, als dass hier ein orientierungsloses Geklicke „auf gut Glück“ unterstellt werden könnte.

An welcher Stelle ist eine definitorische Grenze also sinnvollerweise zu ziehen? Disqualifiziert ein vereinzelter fehlerhafter Klick bereits das erhobene Ergebnis für den jeweiligen Einzeltext insgesamt, am Ende den Lesedurchlauf eines Probanden durch alle Texte als Ganzes?

Für orientierungsloses Geklicke „auf gut Glück“ spricht jedenfalls ein singulärer Fehlklick sicherlich nicht; eine um wenige Millimeter verrutschte Maus kann hier ebenso wahrscheinlich den Ausschlag dafür gegeben haben, dass ein Klick „daneben ging“, wie mangelnde Sorgfalt beim Lesen. Freilich kann genau diese spekulative Argumentation je nach Wohlwollen auch für Lesezeiten mit 2, 3 oder gar 5 Fehlklicks pro Text herangezogen werden, ohne dass sich an der subjektiven Stichhaltigkeit das mindeste ändern würde.

Mithin muss die Setzung einer Fehlklick-Obergrenze immer willkürlich bleiben, nicht zuletzt deshalb, weil die individuellen Gründe für die Fehlklicks nicht explizit und systematisch erfragt wurden im Experiment. Es ist müßig zu diskutieren, ob dies ein Manko des Versuchsdesigns ist; Im Nachhinein muss auch spekulativ bleiben, ob eine inhaltliche Erfassung der Fehl-Wörter an sich, die Registrierung ihrer geografischen Lage innerhalb des Textfeldes und ihre Nähe zum nächsten Stolperwort mehr Aufschluss über die Gründe der Fehlklicks erbracht hätten. Fakt ist, dass lediglich die schiere Menge der Fehlklicks erfasst wurde.

Wie aber wirkten sich die erfassten Fehlklicks auf die Lesezeiten aus?

Nicht von der Hand zu weisen ist, dass die Anzahl der Fehlklicks offenbar stark mit der Anzahl an auffälligen Ausreißern unter den gemessenen Lesezeiten korreliert, und, dass diese Korrelation wohl nur marginal durch den – eher unerheblichen – Zeitaufwand eines überflüssigen Klicks zustande gekommen sein kann. Der Anteil zeitlicher „Ausreißer“ unter den gemessenen Lesezeiten reduziert sich umso erheblicher, je mehr Einzelergebnisse man unter dem Kriterium hoher Fehlklick-Raten aus der Gesamtschau ausschließt. Dies wird deutlich, betrachtet man nicht nur die arithmetischen Zeitemittel, vor allem aber die Standardabweichungen, die sich aus diesem Ausschlussverfahren ergeben (siehe Tabelle 5.1). Die Streuung wächst überproportional in Relation zum Anstieg der arithmetischen Mittel und der Mediane, ein bemerkenswerter Anstieg der Standardabweichungen ergibt sich vor allem im Übergang von maximal einem zu maximal zwei Fehlklicks. Auf einen Anstieg der Ausreißer paral-

iel zur wachsenden Fehlklick-Anzahl deutet auch die Entwicklung der 95 %-Quantile hin – also jenes Wertes, oberhalb dessen sich 5 % aller gemessenen Zeiten befinden.

Tabelle 5.1: Fehlklicks und Lesezeiten					
Fehkl.	n	Ø-Lesez.	S	Median	95%-Qu.
0	3786	28,95 sec	12,07	26,72	47,50 sec
<= 1	4035	29,27 sec	12,30	26,85	48,60 sec
<= 2	4164	29,68 sec	12,93	27,04	50,49 sec
<= 3	4231	29,85 sec	13,10	27,16	51,21 sec
<= 4	4265	29,93 sec	13,15	27,21	51,42 sec
<= 5	4291	29,98 sec	13,26	27,22	51,75 sec
<= 10	4338	30,23 sec	13,81	27,28	52,60 sec
>= 0	4619	31,49 sec	16,67	27,60	59,04 sec

Quelle: eigene Erhebung

Insgesamt muss festgestellt werden:

Die Anzahl der Fehlklicks beeinflusste die Ergebnisse nicht in einem Maße, dass aus ihnen eine eindeutige Entscheidung für eine „Fehlklick“-Grenze deduzierbar wäre.

Höchstens zwischen den Kategorien „maximal 1 Fehlklick“ und „maximal 2 Fehlklicks“ tut sich eine ge-

wisse augenfällige Entwicklung auf. Der Statistiker Walter Krämer jedenfalls plädierte, als ich ihm das Problem schilderte und die Daten der Tabelle 5.1 vorlegte, sogar dafür, nur Zeiten in die Wertung einzubeziehen, die auf Grundlage von 0 Fehlklicks zustande gekommen sind⁸⁷⁸.

Ich habe mich letztlich entschieden, nur gemessene Zeiten in die Auswertung einzubeziehen, die mit maximal einem Fehlklick zustande gekommen waren. Diese Grenze ist, wie gesagt, notwendigerweise willkürlich gezogen.

5.6.3 Kriterium Lesezeit

Ähnliche Schwierigkeiten wie bei der Bewertung der Fehlklick-Raten ergaben sich in Betrachtung und Beurteilung der erhobenen Lesezeiten. Dass eine Aussonderung bestimmter Ergebnisse unumgänglich war, stand in meinen Augen fest; eine Lesedauer von 102 Sekunden und mehr, in immerhin 1,12 Prozent aller Fälle registriert, schien unverwertbar; schließlich entsprachen diese Zeiten einer Lesegeschwindigkeit von maximal einem Wort pro Sekunde. Ein Wert, der gewiss nicht auf den konventionellen Leseprozess eines des Lesens Kundigen gefolgt war. Eher steht zu vermuten, dass die Betreffenden schlicht unaufmerksam rezipierten und die Stolperwörter dabei überlasen.

Doch wiederum ergab sich die Frage: Wo war eine definitorische Grenze zu ziehen? Wo war der Beginn eines „Ausreißerbereichs“ zu setzen? War bereits eine Lesezeit von 46 Sekunden untauglich im Sinne des Experiments? Immerhin hatten nur zehn Prozent aller Teilnehmenden so lange oder länger für einen Text benötigt!

⁸⁷⁸ Gespräch mit Walter Krämer am 25.01.2005, 15.00 – 17.25 Uhr

Zusätzlich stellte sich die Frage, ob eine einheitliche Grenzziehung über alle vier Test-Texte hinweg sinnvoll sein würde. Immerhin war, bei allem Bemühen um Vergleichbarkeit, die Schwierigkeit der vorgelegten Texte offenbar nicht vollkommen gleichmäßig ausgefallen.

Tabelle 5.2 (bereits bereinigt um Fehlklicks und Mehrfachteilnehmer) veranschaulicht diese Differenzen. Während die durchschnittlichen Lesezeiten der Texte „1“ und „2“ noch in einem durchaus kompakten Bereich lagen, wurde Text „3“ bereits merklich rascher gelöst, Text „4“

	Ø-Lesezeit	S
Alle Texte	29,27 sec	12,30
Text 1 („Olympia“)	30,81 sec	13,75
Text 2 („Schlitzohr“)	30,57 sec	11,22
Text 3 („Hornberg“)	28,90 sec	11,60
Text 4 („Ägypten“)	26,95 sec	12,27
<i>Quelle: eigene Erhebung</i>		

nochmals erheblich schneller. Es liegt der Verdacht nahe, dass hier der Trainingseffekt durchschlug. Nicht in vergleichbarem Maße auffällig entwickelten sich dagegen die für die einzelnen Texte festgestellten Standardabweichungen.

Die festgestellten Unterschiede legten nahe, für die Datenbereinigung keine fixe „Ausschluss-Grenze“ bei den Lesezeiten zu definieren. Vielmehr sollten die Daten, so beschloss ich, unter dem Kriterium „Text“ individuell bewertet werden.

Einen Ansatzpunkt boten dabei die ermittelten Standardabweichungen – jene durchschnittlichen Abweichungen aller erhobenen Lesezeiten also vom jeweiligen arithmetischen Mittel.

Für klassische statistische Normalverteilungen gilt dabei, dass sich nahezu 96 % aller Werte im Bereich (Mittelwert +/- doppelte Standardabweichung) bewegen sowie 99 % aller Werte im Bereich (Mittelwert +/- dreifache Standardabweichung)⁸⁷⁹. Nun lag, wie gezeigt, im Falle des hier besprochenen Experiments eben keine klare Normalverteilung der Werte vor. Dies zeigt sich beispielsweise an Text „2“: Alle erfassten Lesezeiten unterhalb des Durchschnittswertes bewegen sich im Intervall zwischen Mittelwert und (Mittelwert – doppelte Standardabweichung) => $(30,57 - 22,44) = 8,13$ Sekunden. In der „Gegenrichtung“ dagegen fanden sich durchaus noch einige Messwerte jenseits sogar der dreifachen Standardabweichung über dem arithmetischen Mittelwert, also jenseits von $(30,57 + 33,66) = 64,23$ Sekunden.

Immerhin bewegten sich – über alle Texte hinweg betrachtet und nach den beiden Vorkriterien bereinigt – nur 5 Prozent aller erfassten Lesezeiten oberhalb von 48,50 Sekunden, noch 3 Prozent oberhalb von 54,63 Sekunden, lediglich gut 2 Prozent aller Werte oberhalb von 60 und 1 Prozent jenseits von 72 Sekunden.

Zu berücksichtigen war auch, dass angesichts von 90 untersuchten Konstellationen pro Einzelkonstellation nur etwa 40 Datensätze zur Verfügung stehen würden. Ein vereinzelter Messwert, der um 40 Sekunden über dem Mittelwert läge, würde hier also bereits eine Resultatsverzerrung von 1,00 Sekunde bewirken. Eine engere Grenzziehung beim Ausreißer-Bereich würde diesen Effekt zwar nicht beseitigen, aber doch merklich abmildern.

⁸⁷⁹ Vgl. Mittenecker 1966, S. 70; vgl. ferner Bortz 2004, S. 42f.

Letztlich habe ich nicht zuletzt aus diesem Grunde entschieden, für jeden Text gesondert den Beginn des nicht berücksichtigten Ausreißer-Bereichs bei dem Wert (Mittelwert + doppelte Standardabweichung) festzusetzen. Das hieß, dass alle Lesezeiten, die unter Text 1 erhoben und größer als $(30,81 + 2 \cdot 13,75 =) 58,31$ Sekunden ausgefallen waren, aus der Wertung genommen wurden. Für Text 2 lag die Obergrenze nach demselben Verfahren bei 53,01 Sekunden, für Text 3 bei 52,10 Sekunden und für Text 4 bei 51,49 Sekunden. Auf diese Weise fielen 3,3 Prozent der erhobenen Messwerte aus der Wertung.

5.6.4 Gleichverteilung der Konstellationen

Zu berücksichtigen war schließlich auch, dass für eine gelungene Auswertung jede der 90 untersuchten Schriftkonstellationen quantitativ einigermaßen gleichmäßig Eingang finden sollte in die Analyse. Hinzu trat, dass das gewählte Auswertungsverfahren der Varianzanalyse seine Tauglichkeit prinzipiell nur unter der Voraussetzung einer solchen Gleichverteilung der Messwerte in Gänze zu entfalten vermag⁸⁸⁰ (zum Verfahren selbst später mehr).

Wie bereits beschrieben, war im Experiment ein weitgehend gleichmäßiger Aufruf aller 90 Konstellationen für jeden Text sowie über alle vier Texte hinweg sicher gestellt. Diese Verteilung geschah freilich automatisiert und unter nur sehr marginaler Berücksichtigung von Durchgangs-Ziffer, Fehlklicks und eventuellen Zeit-Ausreißern. So wurden beispielsweise erfasste Lesezeiten von über 70 Sekunden zwar nicht als erfolgreiche Bearbeitungen registriert (diese Marge schien mir in der Konzeptionsphase des Experiments so großzügig gesetzt, dass sie eine spätere Datenbereinigung nicht substanziell antizipieren würde; dies darf als bestätigt angesehen werden).

Die Folge dieser „Vor-Selektion“ war jedoch, dass bestimmte Schriftkonstellationen nach erfolgter Aussonderung der Mehrfachdurchgänge und der Ergebnisse mit mehr als einem Fehlklick deutlich häufiger gemessen worden waren als andere. Es zeigte sich, dass Text „1“ in jeder der 90 Konstellationen mindestens 9 Mal innerhalb der zeitlichen Ausreißer-Marge (Mittelwert + doppelte Standardabweichung) bearbeitet worden war, die Texte „2“ und „3“ mindestens 10 Mal und Text „4“ mindestens 11 Mal.

Wie erklären sich diese Unterschiede? Immerhin existierten vor der Bereinigung nach den Kriterien „Erstteilnahme“ und „Fehlklick“ ja mehr Messdaten aus Lesedurchgängen von Text „1“ als von den drei anderen! Eine detailliertere Analyse zeigt, dass Text „1“ oft nur unter Aufwendung von mehreren Fehlklicks abschließend bearbeitet wurde; der Trainingseffekt über die Dauer des Experiments hinweg bewirkte, dass mit jedem weiteren Text die Fehlklicks abnahmen.

⁸⁸⁰ Vgl. Bortz 2004, S. 321ff.

Wie war nun zu verfahren? Das eindeutig „sauberste“ Verfahren hätte zweifellos darin bestanden, pro Text und jeder Schriftkonstellation 9 Werte in die Analyse einzubeziehen; damit wäre jede Konstellation in jedem Textzusammenhang in gleichem Gewicht in die Endauswertung eingegangen.

Der konjunktivische Stil des vorausgegangenen Absatzes deutet es bereits an: Ich habe diesen Weg nicht beschritten. Leiten ließ ich mich dabei einzig und allein von der Erwägung, dass auf diese Weise nochmals stattliche (und potenziell wertvolle und aussagekräftige) 360 Messwerte aus der Analyse herausgefallen wären; dieser Preis schien mir in der Abwägung zu hoch angesichts der Tatsache, dass die Unterschiedlichkeit der Texte ja erklärtermaßen nicht Untersuchungsgegenstand der Studie war, sondern quasi Randerkenntnis.

So entschloss ich mich, zwar eine Gleichgewichtung der 90 gemessenen Konstellationen innerhalb jedes Einzeltextes sicherzustellen, pro Einzel-Text jedoch die maximal mögliche Menge an Daten einzubeziehen. In die Wertung gingen also jeweils 9 Messwerte jeder der 90 Schriftanordnungen aus dem Lesen von Text „1“ ein, jeweils 10 Zeilen aus den Texten „2“ und „3“ sowie 11 Einzeldaten aus Text „4“.

5.6.5 Endauswahl der verwerteten Daten

So stellte sich abschließend nur noch die Frage: Welche Daten waren einzubeziehen oder nicht zu berücksichtigen, wenn – beispielsweise – 11 Messwerte zur Verfügung standen, aus den beschriebenen Gewichtungs-Gründen aber nur 9 in die Endwertung eingehen sollten?

Ich habe in diesen Fällen, salopp formuliert, jeweils die Ränder „geschliffen“. Standen also 11 Messwerte zur Verfügung, wo nur 9 berücksichtigt werden sollten, so wurde der jeweils kleinste und der jeweils größte der 11 Werte ausgesondert. War eine ungerade Anzahl von Werten auszusortieren, begann die „Auslese“ beim jeweils kleinsten Wert.

5.6.6 Ergebnis der Datenbereinigung

Zusammenfassend sei an dieser Stelle nochmals der vorgenommene Prozess der Datenbereinigung nachgezeichnet (Abbildung 5.8 verdeutlicht diesen zusätzlich).

Von insgesamt 4753 gemessenen Lesezeiten wurden im ersten Bereinigungs-Schritt 132 Messwerte beziehungsweise 2,8 Prozent ausgesondert, weil sie an einem Rechner erzeugt wurden, von dem aus bereits Messwerte geliefert worden waren. Es verblieben die Ergebnisse, die als „1. Durchgang“ gewertet wurden (in Abbildung 5.8: „davon 1. Durchgang“).

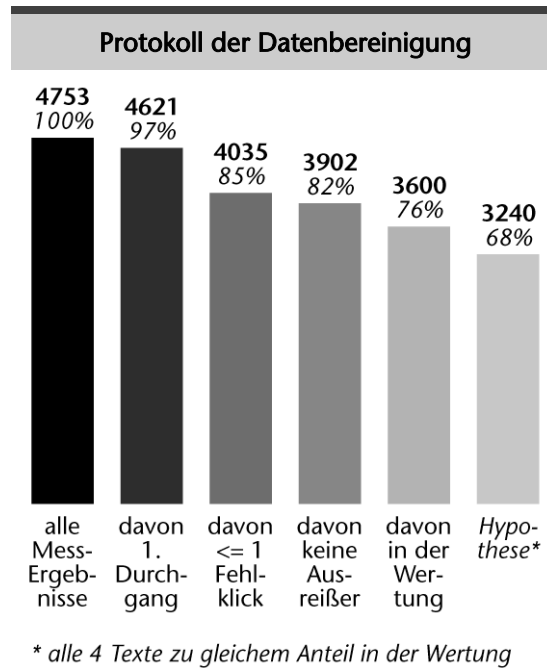
In einem zweiten Bereinigungs-schritt wurden Lesezeiten aussortiert, die zustande gekommen waren durch mehr als drei Klicks seitens des Teilnehmers – also mehr als einen Klick

mehr als notwendig („Fehlclick“). Diesem Ausschlussverfahren fielen weitere 586 Lesezeiten oder 12,7 Prozent der verbleibenden Messwerte zum „Opfer“. Dieser „Bereinigungsschritt“ war damit der „radikalste“ im Rahmen der Datenbereinigung.

Die verbleibenden Messwerte wurden wiederum um jene Lesezeiten bereinigt, die sich außerhalb des festgesetzten zeitlichen Toleranzbereichs bewegten. Diese als „Ausreißer“ identifizierten Lesezeiten waren per Definition größer als die (nach den obigen Kriterien bereinigte) durchschnittliche Lesezeit des Textes, aus deren Lektüre sie hervorgegangen waren, addiert mit der doppelten Standardabweichung vom Mittelwert eben dieses Textes. In diesem Schritt fielen nach diesem Kriterium weitere 133 oder 3,3 Prozent der erhobenen Werte heraus.

Im letzten Bereinigungs-Schritt wurde dafür Sorge getragen, dass jede der getesteten 90 Schriftanordnungen zu gleichem quantitativen Anteil Eingang fand in die Endauswertung. Dabei gingen jeweils 9 Messwerte einer Konstellation in die Wertung ein, die der Lektüre von Text „1“ entsprungen waren, jeweils 10 Messwerte aus dem Lesen der Texte „2“ und „3“ sowie jeweils 11 Ergebnisse, die nach dem Lesen von Text „3“ erfasst worden waren. Aus diesem Bereinigungs-schritt resultierten 3600 Einzel-Messergebnisse, die in die abschließende Wertung aufgenommen wurden. In der Summe wurden in diesem letzten Bereinigungs-Schritt nochmals 302 oder 7,7 Prozent der verbliebenen Werte aus der Wertung genommen. Die verbleibenden 3600 Werte repräsentieren 75,7 Prozent der Ausgangsdaten.

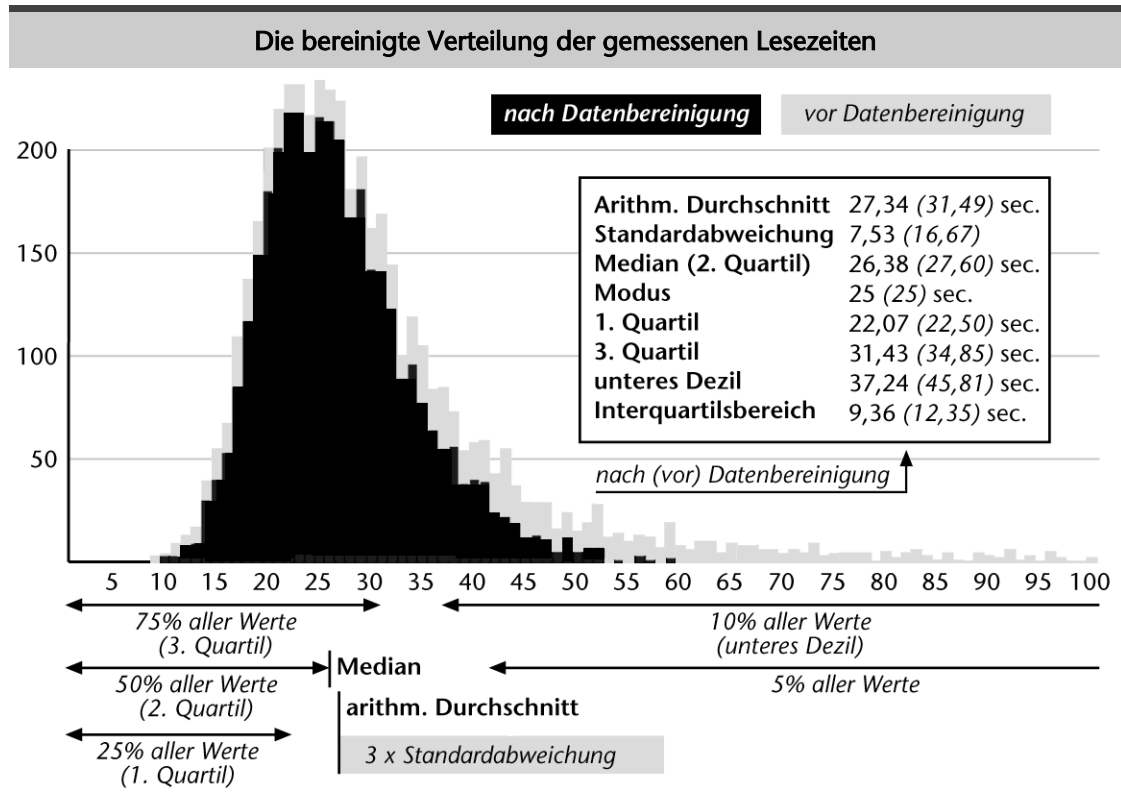
Das aus diesem letzten Schritt resultierende leichte Übergewicht der Messergebnisse aus Lesungen des Textes „4“ wurde dabei ebenso in Kauf genommen wie die leichte Unterrepräsentierung der Messergebnisse von Text „1“. Um eine absolute Gleichgewichtung aller 4 Texte herzustellen, wäre notwendigerweise eine Orientierung am kleinsten Nenner nötig gewesen; dies hätte bedeutet, dass aus allen 4 Texten nur 9 Ergebnisse pro Schriftanordnung in die Wertung hätten Eingang finden müssen. Das potenzielle Ergebnis ist Abbildung 5.8 ebenfalls zu entnehmen: statt 3600 Messwerten hätten nur mehr 3240 Ergebnisse Eingang in die Berechnungen gefunden (in Abbildung 5.8: „Hypothese“). Es wären also noch mehr Messwerte durch den Rost gefallen.



Die Datenbereinigung, aufgeschlüsselt nach den beschriebenen Bereinigungs-schritten.

ABBILDUNG 5.8

Abbildung 5.9 verdeutlicht abschließend in Form eines Histogramms, in welcher Form sich die gemessenen Lesezeiten nach Abschluss des Prozesses der Datenbereinigung verteilen.



Dieses Histogramm gibt wieder, wie sich die gemessenen Lesezeiten nach abgeschlossener Datenbereinigung aufgliederten.

ABBILDUNG 5.9

5.6.7 Alternativen der Datenbereinigung?

Nur gute drei Viertel aller erhobenen Daten fanden nach der oben beschriebenen Datenbereinigung noch Eingang in die Endauswertung. Das riecht, so mag mancher meinen, nach Kahlschlag, nach Datenfrisierung im Interesse eines plakativen Ergebnisses. Mögliche Kritikansätze seien daher an dieser Stelle aufgegriffen.

Zunächst: Die absolut überwiegende Anzahl der Daten fiel zwei Kriterien zum Opfer: den Fehlclicks einerseits und der nachträglichen Ausgewichtung der Schriftkonstellationen andererseits.

Eine Einbeziehung von Texten, die auch mit zwei Fehlclicks zustande gekommen waren, hätte eine Verbreiterung der Datenbasis um gerade einmal 3,2 Prozent bewirkt; eine Ausweitung auf 4 Fehlclicks eine Vergrößerung um 5,7 Prozent. Es sind dies in meinen Augen zu geringe Margen, als dass vom Prinzip hätte abgewichen werden sollen: Mehr als ein Fehlclick spricht gegen einen konventionellen Leseprozess, sondern für „trial-and-error“-Klicken.

Eine großzügigere Setzung der Grenze des „Ausreißer-Bereichs“ eines jeden Textes wäre eine weitere Alternative gewesen. Dieser war, wie beschrieben, angesetzt bei der erhobenen Veoppelten Standardabweichung eines jeden Testtextes, addiert zum arithmetischen Mittel der erhobenen Lesezeiten des jeweiligen Textes. Es waren jedoch lediglich 3,3 Prozent aller Werte, die diesem Kriterium zum Opfer fielen; eine Ausweitung der „Überschreitungsgrenze“ beispielsweise auf die dreifache Standardabweichung hätte also nur marginal mehr Werte in das Endergebnis gelangen lassen.

5.7 Welche Konstellation hat „gewonnen“?

Insgesamt 90 Schriftkonstellationen wurden im hier diskutierten Experiment auf ihre Lesbarkeit untersucht. Jede dieser Schriftkonstellationen findet sich in der bereinigten Datenauswahl gleichverteilt wieder, jede wurde von insgesamt 40 Probanden gelesen, von denen wiederum jeweils neun Versuchsteilnehmer Text 1, jeweils zehn Probanden Text 2 und 3 sowie elf Teilnehmende Text 4 rezipierten. Was liegt da näher, als dem allzu menschlichen Bedürfnis nach Ausrufung eines „Siegers“ zu erliegen, als dem Charme der Rangliste zu erliegen? Welche Konstellation hat also „gewonnen“? Welche Anordnung unter den 90 getesteten war die insgesamt am besten lesbare?

Das Verlangen sei befriedigt. Der Schnitt aller gewerteten Zeiten betrug, wie bereits gezeigt, 27,34. In Tabelle 5.3 sind die 18 typografischen Konstellationen mit den insgesamt besten Durchschnitts-Lesezeiten aufgelistet (mithin das „beste Fünftel“; im Anhang findet sich die vollständige Liste).

Tabelle 5.3: Die 18 Konstellationen mit den besten Durchschnitts-Messergebnissen								
	Ø-Lesez. (Sig)	n	Schriftart	Größe	x-Höhe	Zeilenabstd.	Relat. x-Hö./ZA	Zeilenbreite (Zeichen)
1.	23,49 sec. (4)	40	<i>Times</i>	17 px	8 px	19 px	2,4	42
2.	23,82 sec. (2)	40	<i>Verdana</i>	13 px	7 px	20 px	2,9	42
3.	24,37 sec. (0)	40	<i>Times</i>	17 px	8 px	21 px	2,6	42
4.	24,94 sec. (0)	40	<i>Times</i>	17 px	8 px	23 px	2,9	70
5.	25,02 sec. (0)	40	<i>Verdana</i>	14 px	8 px	20 px	2,5	42
6.	25,03 sec. (0)	40	<i>Times</i>	15 px	7 px	17 px	2,4	42
7.	25,21 sec. (0)	40	<i>Times</i>	17 px	8 px	21 px	2,6	56
8.	25,43 sec. (0)	40	<i>Times</i>	15 px	7 px	21 px	3,0	56
9.	25,49 sec. (0)	40	<i>Verdana</i>	13 px	7 px	20 px	2,9	56
10.	25,66 sec. (0)	40	<i>Times</i>	17 px	8 px	23 px	2,9	56
11.	25,72 sec. (0)	40	<i>Times</i>	17 px	8 px	21 px	2,6	70
12.	25,73 sec. (0)	40	<i>Verdana</i>	14 px	8 px	22 px	2,8	42
13.	25,81 sec. (0)	40	<i>Verdana</i>	14 px	8 px	18 px	2,3	42
14.	25,91 sec. (0)	40	<i>Times</i>	15 px	7 px	19 px	2,7	56
15.	25,95 sec. (0)	40	<i>Verdana</i>	13 px	7 px	20 px	2,9	84
16.	26,00 sec. (0)	40	<i>Verdana</i>	13 px	7 px	18 px	2,6	42
17.	26,06 sec. (0)	40	<i>Times</i>	15 px	7 px	17 px	2,4	84
18.	26,14 sec. (0)	40	<i>Verdana</i>	13 px	7 px	16 px	2,3	55

Quelle: eigene Erhebung

Insgesamt ist das Resultat bemerkenswert. Mit nämlich wirklich bemerkenswertem Vorsprung eroberte eine Schriftkonstellation die „Pole Position“, die vermutlich Wenige – den Autor eingeschlossen – ganz oben vermutet hätten. Es ist

- die Schriftart *Times*
- in der größten im Experiment gemessenen Schriftgröße 17 Pixel (entspricht ca. 13 Punkt)
- mit dem kleinsten gemessenen Zeilenabstand von 19 Pixel (entspricht etwa 14 Punkt)
- in der zweitschmalsten gemessenen Zeilenbreite von durchschnittlich 42 Anschlägen.

Im Vorgriff auf das nächste Kapitel muss allerdings nochmals hervorgehoben werden, dass die in Tabelle 5.3 niedergelegten Durchschnittszeiten sämtlich aus jeweils lediglich 40 Einzelmesswerten generiert wurden, was für einen statistisch tragbaren Vergleich offenkundig zu wenig ist. Im wechselseitigen paarweisen Vergleich der 90 getesteten Schriftkonstellationen – und das waren immerhin $(90 \times 89) = 8010$ Vergleichspaare – kristallisierten sich lediglich sechs signifikant unterschiedliche Pärchen heraus nach Einsatz des Post Hoc-Tests nach der „Dunnett T3“-Methode (die weiter unten noch vorgestellt wird). Wie Tabelle 5.3 ausweist (in Klammern hinter der gemessenen Durchschnittslesezeit in Spalte 2), fanden sich für die „Sieger-Konstellation“ gerade einmal vier wechselseitige Signifikanzen im Abgleich mit den 89 übrigen Mess-Varianten (und dies waren die Konstellationen auf den Rangplätzen 87, 88, 89 und 90).

Nichtsdestotrotz ist Tabelle 5.3 nicht ohne Reiz. Denn sie erlaubt bereits einige interessante, wiewohl zunächst explorative und statistisch nicht detaillierter hinterfragte Beobachtungen – zum Beispiel diese:

- *Times* und *Verdana* scheinen gleich gut lesbar am Bildschirm: Unter den besten 20 gemessenen Konstellationen waren beide Schriftarten jeweils 10-Mal vertreten.
- Schriften, die eine x-Höhe von 6 Pixeln aufweisen, scheinen schlechter lesbar zu sein als größere. Die durchschnittlich am schnellsten gelesene Konstellation mit einer Schriftgröße von 6 Pixeln x-Höhe landete auf dem 22. Platz.
- ein erstrebenswerter Zeilenabstand scheint im Bereich des 2,5- bis 3,0-fachen der x-Höhe der gewählten Schrift zu liegen (in der Tabelle: „Relat. x-Hö./ZA“).
- Zeilen, die durchschnittlich 42 oder 55 Anschlägen Platz bieten, scheinen die am besten lesbaren zu sein – 14 der 18 besten Konstellationen (78 Prozent) wiesen eine dieser beiden Durchschnittsbreiten auf.

Spannende Zwischenresultate – die im folgenden Kapitel inferenzstatistisch genauer untersucht werden.

6 Die Ergebnisse im Detail

6.1 Die statistische Signifikanz

6.1.1 Begriff und Bedeutungsgehalt

In den Schilderungen vorliegender Studien zur Lesbarkeit von Bildschirm-Typografie in Kapitel 3 tauchte der Begriff bereits in merklicher Frequenz auf, und in den Darlegungen dieses Kapitels wird die Verwendungsrate kaum geringer ausfallen in ihrer Häufigkeit: Signifikanz. Statistische Signifikanz, genauer gesagt. Zimbardo und Gerrig umschreiben nicht nur den Bedeutungsgehalt des Begriffs, sondern gleich auch dessen „magischen Wert“ in seiner Ausprägung so: „Auf der Basis einer Konvention betrachten Psychologen einen (gemessenen) Unterschied als bedeutsam, wenn die Wahrscheinlichkeit, dass er durch Zufall zustande gekommen ist, weniger als 5 von 100 beträgt“⁸⁸¹.

Hier wird also mithilfe statistischer Methoden die Möglichkeit quantifiziert, dass die Ergebnisse einer empirischen Vergleichs-Messung keinerlei Rückschlüsse und Verallgemeinerungen über die eigentliche Datenerhebung hinaus zulassen – dass also Messresultate nicht, wie erwünscht, ein Abbild der „Wirklichkeit“ zeichnen, sondern Ausfluss eines Zufall sind. Diese Zufalls- oder auch Irrtumswahrscheinlichkeit wird statistisch durch den sogenannten „p-Wert“ bestimmt (p = französisch „probabilité“), dessen durchaus zahlreiche Ermittlungsmethoden in dieser Arbeit nur insofern vorgestellt und diskutiert werden sollen, als sie der Würdigung der von mir erfassten Daten eine Hilfe sind⁸⁸².

Ob ein erhobener Messwert als bedeutsam, also statistisch „signifikant“ verschieden von einem anderen klassifiziert wird, hängt insbesondere von der Setzung des „Signifikanzniveaus“ (auch „Verlässlichkeitsniveau“ genannt⁸⁸³) durch den Forscher ab – mit dem Ausmaß an Irrtumswahrscheinlichkeit also, das dem Untersuchenden klein genug erscheint, um einem empirisch gemessenen Unterschied ein wissenschaftlich tolerables Maß an Verallgemeinerbarkeit zuzubilligen⁸⁸⁴. Diesen Wert setzen zu müssen, bleibt im Grunde nur jenen erspart, die sich auf empirische Vollerhebungen stützen können und damit die Ermittlung einer Irrtumswahrscheinlichkeit verzichtbar gestalten – und dies ist nicht nur in den Disziplinen der Sozialwissenschaft nur ausgesprochen selten der Fall. Jenseits von Vollerhebungen (und die gewährleisten nicht einmal Volkszählungen) verbleibt immer eine Zufallskomponente⁸⁸⁵.

Die klassische „Signifikanzschwelle“ wird dabei insbesondere im sozialwissenschaftlichen Bereich sehr oft bei der Irrtumswahrscheinlichkeit „5 %“ angesetzt (wie bereits oben bei Zimbardo und Gerrig angedeutet); bei als hochbedeutsam erkannten Erfassungen allerdings

⁸⁸¹ Zimbardo/Gerrig 2004, S. 65

⁸⁸² Übersichtsweise vgl. Bortz 2004, S. 52

⁸⁸³ Vgl. Hofstätter/Wendt 1974, S. 7

⁸⁸⁴ Vgl. Duller 2006, S. 228

⁸⁸⁵ Vgl. Zimbardo/Gerrig 2004, S. 65

wird das Signifikanzniveau auch gelegentlich schärfer definiert (bei 1 % oder gar 0,1 %⁸⁸⁶). Im weiteren Verlaufe der vorliegenden Arbeit wird, sofern nichts Anderweitiges beschrieben ist, ein Signifikanzniveau von 5 % unterstellt.

6.1.2 Notierungsmethoden der statistischen Signifikanz

Die im Folgenden dargestellten Ergebnisse meiner Untersuchung basieren in ihrer Mehrzahl auf Ermittlungen von Zufallswahrscheinlichkeiten, daher sei die (nicht nur) in dieser Arbeit verwendete Notierungsmethodik der statistischen Signifikanz zum besseren Verständnis an dieser Stelle noch kurz dargelegt.

- Der p-Wert als Kenngröße der statistischen Signifikanz kennt als Ausprägungen den Maximalwert 1 und den Minimalwert 0. Dabei ist der Wert „1“ gleichzusetzen mit einer „Irrtumswahrscheinlichkeit“ von 100 Prozent, der Wert „0“ mit einer Irrtumswahrscheinlichkeit von 0 Prozent. Das heißt: Wenn beispielsweise die Relation zweier statistisch erhobener Werte einen p-Wert von 0,12 aufweist, liegt die statistische Wahrscheinlichkeit, dass zwischen den beiden gemessenen Phänomenen nicht nur im Erhebungszusammenhang, sondern auch realiter ein signifikanter Unterschied besteht, bei 88 Prozent – oder, andersherum betrachtet, die Wahrscheinlichkeit, dass trotz der gemessenen Unterschiede in den Daten kein signifikanter Unterschied zwischen beiden Werten in der vermeintlich abgebildeten Realität besteht, bei 12 Prozent.
- Der p-Wert wird – auch in der vorliegenden Arbeit – in Form einer Gleichung notiert. So wird beispielsweise eine Irrtumswahrscheinlichkeit von 4,3 Prozent wiedergegeben als ($p=0,043$) oder, wenn nur die Erkenntnis einer Signifikanz oder Nicht-Signifikanz (Notierungsmethode für das Signifikanzniveau: „ $\alpha=0,05$ “) ausschlaggebend ist, als ($p<0,05$) oder ($p>0,05$).

6.1.3 Die „heilige Kuh“ Signifikanz

Alles eine Frage der Masse. Statistische Signifikanzen zu finden – diese Chance wächst, das haben die Gesetze der Stochastik so an sich, mit der Anzahl der Versuchsteilnehmer: „Durch Wahl einer hinreichend großen Stichprobe wird es in den meisten Fällen möglich sein, auch noch so geringfügige Unterschiede statistisch zu sichern“⁸⁸⁷. Inzwischen beklagen nicht wenige Experten das Überbordern dessen, was Krämer als „Signifikanztest-Ritual“⁸⁸⁸ brandmarkt: Der Fokus empirischer Forschung und Abhandlung gelte – offenbar nach einer Ära der Geringschätzung des Verlässlichkeits-Maßes – zunehmend nur noch der Frage, ob erho-

⁸⁸⁶ Vgl. Mittenecker 1966, S. 31

⁸⁸⁷ Hofstätter/Wendt 1974, S. 82

⁸⁸⁸ Vgl. Krämer 2006, S. 58

bene Werte signifikant ausgefallen seien. Ob jedoch diese Erkenntnisse auch ein gesundes Maß an Praxisrelevanz entfaltet, interessiert kaum noch⁸⁸⁹ – Hofstätter und Wendt sprechen hier treffend von der anzustrebenden „praktischen Signifikanz“: „Zu fragen bleibt, ob (...) für die Praxis etwas gewonnen wird“⁸⁹⁰.

In der Tat: Bezogen auf das von mir durchgeführte Experiment war, jenseits aller potenziell ermittelbaren und ermittelten statistischen Signifikanzen, die schlichte Frage zu stellen, ab welchem Punkt eine festgestellte Beschleunigung der Lesezeit durch Manipulation eines oder mehrerer typografischer Faktoren in der Ausprägung eine wirkliche Praxisrelevanz würde entfalten können. Wenn die Heraufsetzung einer Zeilenbreite um 20 Prozent im Experiment die Lesegeschwindigkeit um 4 Prozent erhöht, und diese Beschleunigung dann sogar noch als statistisch signifikant erkannt wird – rechtfertigt diese Erkenntnis dann eine apodiktische Empfehlung an die Gilde der Webseiten-Gestalter, eher die größere als die kleinere Zeilenlänge zu definieren? Rein rechnerisch bedeutet diese Differenz schließlich, dass ein Text in der längeren Zeilenbreite vermutlich in 6 Minuten gelesen würde, wo der Rezeptionsprozess in der kürzeren Breite eine Viertelminute länger dauern würde. Dies kann, muss man aber nicht praktisch bedeutsam finden. Ähnlich geschmacksabhängig ist freilich, ob eine Steigerung der Lesegeschwindigkeit von 10, gar 15 Prozent nach Manipulation einer typografischen Variable als bedeutend eingeschätzt wird oder nicht.

Obwohl ich in den folgenden Darlegungen selbstverständlich nicht auf die Erwähnung ermittelter statistischer Signifikanzen verzichten werde, versuche ich aus den obigen Erwähnungen heraus, mich nicht zu stark und insbesondere zu ausschließlich an selbige zu klammern in der Ergebnisdiskussion. Hervorheben und betrachten in ihrer zusätzlich potenziellen Praxisrelevanz werde ich insbesondere jene Vergleichswerte, die um mindestens 5 Prozent voneinander differierten in der Messung; es ist dies zugegebenermaßen ein wiederum willkürlich gesetzter Wert, aber, so meine ich, doch einer, der geeignet ist, die „statistisch“ von den „praktisch“ signifikanten ausreichend abzusetzen.

6.2 Die Varianzanalyse

Ziel der hier diskutierten Studie war, wie bereits dargelegt, nicht nur, isolierte Erkenntnisse zu gewinnen über die typografischen Faktoren Schriftart, Schriftgröße, Zeilenbreite und Zeilenabstand und deren Einfluss auf die Lesegeschwindigkeiten der Teilnehmenden. Von weiterem und besonderem Interesse war, ob und in welchem Maße die vier Faktoren interagieren, möglicherweise einander bedingen.

⁸⁸⁹ Vgl. Krämer 1997, S. 180f.

⁸⁹⁰ Hofstätter/Wendt 1974, S. 82

Gängiges Instrument zur Beantwortung dieser Fragen ist die statistische Varianzanalyse⁸⁹¹, die im englischsprachigen Raum auch mit dem Akronym ANOVA bezeichnet wird. Die vorliegende Varianzanalyse wurde mithilfe des Programms SPSS gestellt.

Die Varianzanalyse erlaubt die Klärung der Frage, ob und in welchem Maße einzelne gemessene Stimuli – im vorliegenden Falle also die vier gemessenen typografischen Faktoren – Einfluss auf die Entstehung eines Messergebnisses hatten⁸⁹². Allgemeine „Nullhypothese“ der klassischen Varianzanalyse ist, dass alle verglichenen Messwerte in allen Gruppen gleich groß seien⁸⁹³ – ein p-Wert unter dem Signifikanzniveau von 0,05 signalisiert in der Varianzanalyse also, dass dem nicht so ist, dass bestimmte Faktoren bestimmenden Einfluss auf die unabhängige Variable (im vorliegenden Falle: die Lesezeit) nahmen. Doch nicht nur der Einfluss einzelner Faktoren auf erhobene Ergebnisse ist einer Varianzanalyse entnehmbar – zusätzlich weist das Verfahren auch Interaktions-Einflüsse aus⁸⁹⁴. Es war also auch klärbar, ob beispielsweise das Zusammenwirken von getesteter Schriftart und Schriftgröße die Leseergebnisse maßgeblich beeinflusste. Gesucht – und hoffentlich auch gefunden – werden in Varianzanalysen mithin sogenannte „Haupteffekte“, jene Stimuli oder Stimulus-Konstellationen also, die die Messergebnisse in statistisch signifikanter Weise mitbestimmten⁸⁹⁵.

6.2.1 Die Resultate der Varianzanalyse

Die nachfolgende Tabelle gibt die Ergebnisse der Varianzanalyse auf Basis der von mir erhobenen und bereinigten Daten wieder. Dabei wurde – natürlich – der Faktor „Lesezeit“ („ZEIT“) als abhängige Variable eingesetzt, die vier zentral interessierenden typografischen Faktoren „Schriftart“ („SCHRIFTA“), „Schriftgröße“ („GROESSE“), „Zeilenabstand“ („ZEILENAB“) und „Zeilenbreite“ („ZEILENBR“) als abhängige Variablen. Zusätzlich wurden die Messwerte „Bildschirmart“ („BILDSCHI“) sowie „Alter“ („ALTER“) als sogenannte Ko-Varianten in die Auswertung mit einbezogen – das heißt, deren Einfluss aufs Leseergebnis wurde gleichfalls eruiert, jedoch nicht unter direktem und wechselseitigen Einbezug der vier zentralen unabhängigen Variablen; der typografischen nämlich.

⁸⁹¹ Vgl. Bortz 2004, S. 243ff.

⁸⁹² Vgl. Brosius/Koschel 2005, S. 200

⁸⁹³ Vgl. Hofstätter/Wendt 1974, S. 108f.

⁸⁹⁴ Vgl. Storm 1974, S. 72

⁸⁹⁵ Vgl. Mittenecker 1966, S. 70

Tabelle 6.1: Ergebnis der Varianzanalyse						
Tests of Between-Subjects Effects						
Dependent Variable: ZEIT						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Corrected Model	9812.115	91	107.825	2.105	.000	
Intercept	89171.640	1	89171.640	1740.453	.000	
ALTER	370.981	1	370.981	7.241	.007	
BILDSCHI	802.446	1	802.446	15.662	.000	
SCHRIFTA	101.125	1	101.125	1.974	.160	
GROESSE	1404.278	2	702.139	13.704	.000	
ZEILENAB	123.633	2	61.817	1.207	.299	
ZEILENBR	1075.359	4	268.840	5.247	.000	
SCHRIFTA * GROESSE	724.902	2	362.451	7.074	.001	
SCHRIFTA * ZEILENAB	13.991	2	6.995	.137	.872	
GROESSE * ZEILENAB	325.041	4	81.260	1.586	.175	
SCHRIFTA * GROESSE * ZEILENAB	308.612	4	77.153	1.506	.198	
SCHRIFTA * ZEILENBR	194.707	4	48.677	.950	.434	
GROESSE * ZEILENBR	416.353	8	52.044	1.016	.421	
SCHRIFTA * GROESSE * ZEILENBR	296.507	8	37.063	.723	.671	
ZEILENAB * ZEILENBR	454.812	8	56.851	1.110	.353	
SCHRIFTA * ZEILENAB * ZEILENBR	799.841	8	99.980	1.951	.049	
GROESSE * ZEILENAB * ZEILENBR	825.632	16	51.602	1.007	.445	
SCHRIFTA * GROESSE * ZEILENAB * ZEILENBR	1440.807	16	90.050	1.758	.031	
Error	179731.428	3508	51.235			
Total	2880296.805	3600				
Corrected Total	189543.543	3599				
a. R Squared = .052 (Adjusted R Squared = .027)						
<i>Quelle: eigene Erhebung</i>						

Die wichtigsten Spalten in der obigen Auflistung stellen die jeweils linke und ganz rechts platzierte dar. Links ist der gemessene typografische Faktor verzeichnet beziehungsweise die wechselseitige Kombination von gemessenen Faktoren; die rechte Spalte („Sig.“) gibt den p-Wert an, der für den jeweiligen Messfaktor beziehungsweise die angegebene Kombination ermittelt wurde. Der Tabelle ist zu entnehmen, dass beim festgelegten Signifikanzniveau von $\alpha=0,05$:

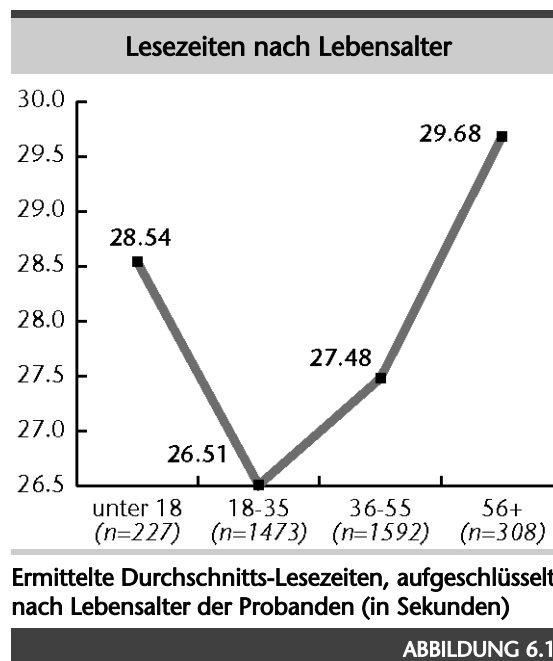
- im Experiment der Faktor „Schriftgröße“ (GROESSE) einen Haupteinfluss auf die gemessenen Lesezeiten ausübte ($p<0,001$).
- auch die Zeilenbreite (ZEILENBR) einen Haupteinfluss ausübte ($p<0,001$).
- eine sogenannte „Interaktion“ festgestellt wurde für die Kombination „Schriftart/Schriftgröße“ (SCHRIFTA * GROESSE). Diese beiden Faktoren bedingten einander im Experiment also messbar wechselseitig ($p<0,01$).
- ebenfalls eine Interaktion gefunden wurde für die Kombination „Schriftart/Zeilenabstand/Zeilenbreite“ (SCHRIFTA * ZEILENAB * ZEILENBR) ($p<0,05$).
- und eine Interaktion aller Einflussfaktoren ausgewiesen wird (SCHRIFTA * GROESSE * ZEILENAB * ZEILENBR) ($p<0,05$).

LENAB * ZEILENBR), ($p < 0,05$). Dass sich freilich diese Interaktion – in der vorläufigen „Rangliste“ in Kapitel 5.7 manifestierte es sich – in lediglich sechs wechselseitigen Signifikanzen ausdrückte, wurde bereits beschrieben; nicht zuletzt aus diesem Grunde soll dieser Befund „alles hängt mit allem zusammen“ im Folgenden nicht weiter verfolgt werden.

Zusätzlich zu den oben angeführten Signifikanzen stellte sich allerdings heraus, dass auch die beiden „Ko-Varianten“, nämlich Bildschirmart und Alter der Probanden, einen bedeutsamen Einfluss auf die erhobenen Messzeiten ausübten. Es galt also, hier noch einmal genauer hinzuschauen.

6.2.2 Bildschirmart und Lebensalter als „Überraschungsfaktoren“

Das Lebensalter der Probanden erwies sich, zu meiner Überraschung, in der Varianzanalyse als wesentlicher Einflussfaktor auf die erzielten Lesezeiten ($p < 0,01$). Abbildung 6.1 verdeutlicht dies: Die besten Lesegeschwindigkeiten – über sämtliche typografischen Konstellationen hinweg – erzielten Teilnehmende, die laut Eigenangabe zum Zeitpunkt des Experiments zwischen 18 und 35 Jahre alt waren. Probanden aus dieser Altersgruppe lasen damit um durchschnittlich 7,1 Prozent schneller als die minderjährigen Teilnehmenden, um 3,5 Prozent schneller als die Altersgruppe der 36- bis 55-jährigen sowie um beachtliche 10,7 Prozent schneller als die Gruppe der über 55-jährigen.



Erstaunliche Resultate – auch angesichts der Tatsache, dass der Literatur kaum Hinweise zu entnehmen sind, dass das Lebensalter von Web-Lesern einen potenziell so großen Einfluss nimmt auf die Lese-Leistung. Freilich verbirgt sich auch eine zweite Erkenntnis in Abbildung 6.1: Die Ergebnisse entstammen sehr unterschiedlich großen Stichproben. Schließlich verteilen sich, wie auch bereits gezeigt, die Altersgruppen im Experiment ausgesprochen heterogen auf die Teilnehmenden. Was daraus folgen kann und sollte, wird weiter unten besprochen.

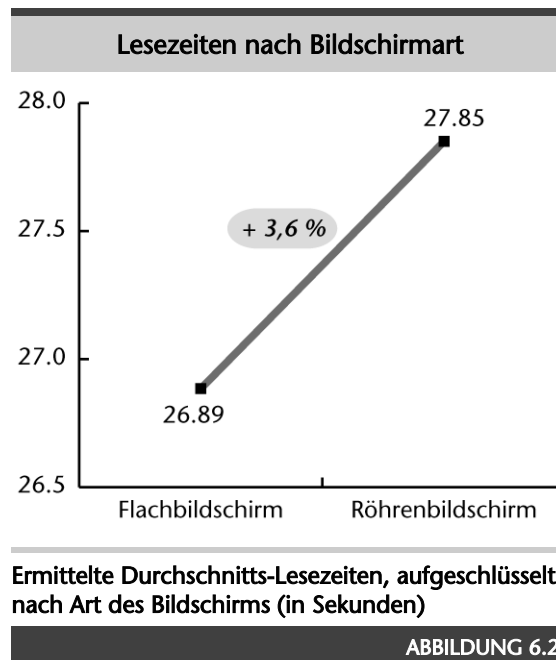
Denn zunächst soll auf den zweiten Überraschungsfaktor im Experiment hingewiesen werden, die zweite erhobene Ko-Variante. Auch die Konstruktionsart des Bildschirms nämlich, an dem die Probanden das Experiment durchführten, hatte laut Varianzanalyse offenbar einen merklichen Einfluss auf die erhobenen Lesegeschwindigkeiten ($p < 0,001$). Eine erste Auswer-

tung ergab das unten stehende Bild (Abbildung 6.2): Insgesamt lagen die Lesezeiten derjenigen Teilnehmer, die angaben, an einem Laptop oder vor einem Flachbildschirm zu sitzen, über alle Schriftkonstellationen hinweg eine knappe Sekunde unter denen der Probanden mit klassischem Röhrenbildschirm. Dies entspricht einer Beschleunigung der Lesegeschwindigkeit von immerhin 3,6 Prozent am Flachbildschirm im Vergleich zum Röhrenbildschirm.

Diesen Einfluss der Bildschirmart hatte ich in diesem ausgeprägten Maße genauso wenig erwartet wie die oben beschriebenen Unterschiede unter dem Kriterium „Lebensalter“ – zumal auch die Literatur in der Diskussion typografischer Faktoren am Bildschirm diesem Aspekt höchstens am Rande Beachtung schenkt. Freilich sind Flachbildschirme, aber auch Laptops erst in den vergangenen drei bis vier Jahren wirklich erschwinglich und damit zu Alternativen auch im privaten Gebrauch avanciert. Noch im Jahre 2002, davon ist auszugehen, unterstellten die meisten Autoren zu Recht klassische Röhrenbildschirme in den Büros und Arbeitszimmern – in der vorliegenden Untersuchung dagegen stellten die Inhaber von „flachen“ Schirmen die knappe, aber immerhin eben die Mehrheit.

Fakt war damit aber, dass zu den ursprünglich vier zu messenden Faktoren – Schriftart, Schriftgröße, Zeilenabstand und Zeilenbreite – recht unvermittelt potenziell ein fünfter und ein sechster hinzutreten; eben Bildschirmart und Lebensalter der Probanden. Wie war damit zu verfahren?

Grundsätzlich sprach einiges dafür, die Einfluss-Faktoren „Bildschirm“ und „Alter“ in die weiteren Analysen einzubeziehen. Dies war insofern problematisch, als – wie gezeigt – in der Vorbereitung und Datenbereinigung für die Varianzanalyse sehr sorgfältig auf eine absolute Gleichverteilung aller gemessenen Faktoren im einzelnen wie in wechselseitiger Kombination im Datensatz geachtet worden war. Ich hatte im Zuge der Datenbereinigung darauf Wert gelegt, dass jede der getesteten 90 typografischen Konstellationen in identischer Anzahl von Messergebnissen in die Endauswertung einging, die zusätzlich aus einer vergleichbaren Anzahl von differierenden Test-Texten hervorgegangen waren. Das Hinzutreten der unverhofften Faktoren „Bildschirmart“ und „Alter“ warf dieses Konstruktionsprinzip aus der Bahn.



Denn im Prinzip wäre es nun beispielsweise notwendig gewesen, auch den Faktor „Bildschirmart“ nach dem Muster der ursprünglichen vier gemessenen Faktoren gleichmäßig über die vier Textinhalte und die 90 gemessenen typografischen Faktoren zu verteilen. Aus den ursprünglich $(4 * 90) = 360$ Messeinheiten waren plötzlich $(4 * 90 * 2) = 720$ Messeinheiten geworden. Eine erste überschlägige Analyse ergab, dass, wenn ich auf Basis der vorhandenen Daten eine absolute Gleichverteilung aller Messeinheiten im Gesamtergebnis sicherstellen wollte, sich die Gesamtzahl der einzubeziehenden Messwerte von 3600 auf deprimierende 2160 verringert hätte – denn nur in dieser Anzahl lagen Messergebnisse vor, in denen sich die vier typografischen Faktoren in ihren 90 wechselseitigen Kombinationen, die vier Test-Texte und die beiden Bildschirmarten absolut gleichverteilt wiedergefunden hätten. Hätte ich zusätzlich den Faktor „Alter“ gleichverteilt in die Varianzanalyse eingebracht, wären sogar nur noch 450 Messwerte übrig geblieben.

Eine Alternative aus diesem Dilemma wäre eine Nacherfassung gewesen. Diese hätte jedoch erhebliche, im Rahmen dieses Projekts nicht rechtfertigbare zeitliche Verzögerungen nach sich gezogen – nicht nur, weil eine neuerliche, absehbar schwierigere Akquisitionsphase notwendig geworden wäre, sondern auch, weil eine sehr tief gehende Umprogrammierung des Zufallsalgorithmus' angezeigt gewesen wäre, der der Ursprungsuntersuchung zugrunde gelegen hatte.

Zumindest in Bezug auf den Faktor „Lebensalter“ und angesichts dessen höchst unproportionaler Verteilung über die Probanden hinweg habe ich mich entschieden, es vorläufig bei der obigen Kurzdarstellung zu belassen. Eben auf Grund der stark differierenden Stichproben-Größen in diesem Betrachtungsbereich versprach eine tiefer gehende Analyse – also die Detailbetrachtung beispielsweise der Messwerte zum Faktor „Zeilenbreite“ unter dem Aspekt des Alters –, nur sehr wenig aussagekräftige Daten zu erbringen. Festzuhalten aber bleibt, dass der Aspekt „Lebensalter“ in jedem Falle auf der Agenda verbleiben sollte; eine Studie, die sich diesem Thema gezielter zuwendet, scheint in jedem Falle angezeigt.

Anders bewertete ich den vermeintlichen „Randfaktor“ Bildschirm. Im Vergleich zu den Daten zum Lebensalter verteilten sich dessen beide gemessenen Ausprägungen (Flachbildschirm, Röhrenbildschirm) deutlich harmonischer über die Messzeiten, wenn auch, wie erwähnt, nicht völlig gleichverteilt über alle typografischen Konstellationen hinweg.

Trotz dieser Unwägbarkeiten habe ich mich letztlich entschieden, die Varianzanalyse ein zweites Mal durchzuführen – und zwar diesmal unter Einbezug des Faktors „Bildschirmart“ als Hauptvariable. Um die dabei eindeutig verletzte Gleichverteilung der Konstellationen in der Gesamtheit der Messdaten zumindest ein wenig aufzufangen, entschloss ich mich, das Signifikanzniveau in dieser zweiten Varianzanalyse auf 0,03 anzuheben (dieses recht einfach

Behelfsverfahren erwähnen Hofstätter und Wendt⁸⁹⁶). Hier also die Ergebnisse der zweiten Varianzanalyse:

Tabelle 6.2: Ergebnisse der zweiten Varianzanalyse					
Tests of Between-Subjects Effects					
Dependent Variable: ZEIT					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	14661.951	179	81.910	1.602	.000
Intercept	2626146.15	1	2626146.1	51357.1	.000
GROESSE	1443.420	2	721.710	14.114	.000
ZEILENAB	124.582	2	62.291	1.218	.296
ZEILENBR	1115.312	4	278.828	5.453	.000
SCHRIFTA	39.073	1	39.073	.764	.382
BILDSCHI	812.446	1	812.446	15.888	.000
GROESSE * ZEILENAB	357.086	4	89.272	1.746	.137
GROESSE * ZEILENBR	434.785	8	54.348	1.063	.386
ZEILENAB * ZEILENBR	503.963	8	62.995	1.232	.276
GROESSE * ZEILENAB * ZEILENBR	908.243	16	56.765	1.110	.339
GROESSE * SCHRIFTA	714.056	2	357.028	6.982	.001
ZEILENAB * SCHRIFTA	4.086	2	2.043	.040	.961
GROESSE * ZEILENAB * SCHRIFTA	315.370	4	78.843	1.542	.187
ZEILENBR * SCHRIFTA	122.681	4	30.670	.600	.663
GROESSE * ZEILENBR * SCHRIFTA	311.052	8	38.881	.760	.638
ZEILENAB * ZEILENBR * SCHRIFTA	774.315	8	96.789	1.893	.057
GROESSE * ZEILENAB * ZEILENBR * SCHRIFTA	1371.800	16	85.737	1.677	.044
GROESSE * BILDSCHI	153.992	2	76.996	1.506	.222
ZEILENAB * BILDSCHI	42.951	2	21.476	.420	.657
GROESSE * ZEILENAB * BILDSCHI	77.953	4	19.488	.381	.822
ZEILENBR * BILDSCHI	120.673	4	30.168	.590	.670
GROESSE * ZEILENBR * BILDSCHI	863.216	8	107.902	2.110	.032
ZEILENAB * ZEILENBR * BILDSCHI	372.535	8	46.567	.911	.506
GROESSE * ZEILENAB * ZEILENBR * BILDSCHI	1034.696	16	64.668	1.265	.211
SCHRIFTA * BILDSCHI	62.229	1	62.229	1.217	.270
GROESSE * SCHRIFTA * BILDSCHI	360.354	2	180.177	3.524	.030
ZEILENAB * SCHRIFTA * BILDSCHI	366.145	2	183.073	3.580	.028
GROESSE * ZEILENAB * SCHRIFTA * BILDSCHI	227.677	4	56.919	1.113	.348
ZEILENBR * SCHRIFTA * BILDSCHI	88.039	4	22.010	.430	.787
GROESSE * ZEILENBR * SCHRIFTA * BILDSCHI	229.981	8	28.748	.562	.810
ZEILENAB * ZEILENBR * SCHRIFTA * BILDSCHI	423.080	8	52.885	1.034	.407
GROESSE * ZEILENAB * ZEILENBR * SCHRIFTA * BILDSCHI	805.394	16	50.337	.984	.471
Error	174881.592	3420	51.135		
Total	2880296.80	3600			
Corrected Total	189543.543	3599			
a R Squared = .077 (Adjusted R Squared = .029)					
<i>Quelle: eigene Erhebung</i>					

⁸⁹⁶ Vgl. Hofstätter/Wendt 1974, S. 138

Die Verschiebungen gegenüber der ersten Varianzanalyse waren nicht durchgehend merklich ausgeprägt, im Detail jedoch beobachtbar. So erwies sich, dass, unter Annahme des gesenkten Signifikanzniveaus von 0,03 ...

- der Faktor Schriftgröße (GROESSE) nach wie vor einen Haupteinfluss auf die gemessenen Lesezeiten ausübte ($p < 0,001$).
- auch die Zeilenbreite (ZEILENBR) gleich bleibend einen Haupteinfluss ausübte ($p < 0,001$)
- wiederum auch eine sogenannte „Interaktion“ festgestellt wurde für die Kombination „Schriftart/Schriftgröße“ (SCHRIFTA * GROESSE). Diese beiden Faktoren bedingten einander im Experiment also messbar wechselseitig ($p < 0,01$).
- die Kombination „Zeilenabstand/Schriftart/Bildschirmart“ (ZEILENAB * SCHRIFTA * BILD-SCHI) knapp unter der Signifikanzschwelle lag ($p = 0,028$)
- die Kombination „Schriftgröße/Schriftart/Bildschirmart“ (GROESSE * SCHRIFTA * BILD-SCHI) gerade auf der Signifikanzschwelle lag ($p = 0,030$).
- der Faktor „Bildschirm“ (BILD-SCHI) (natürlich) wiederum einen Haupteinfluss ausübte ($p < 0,01$).

Unter den Bedingungen des neu gesetzten Signifikanzniveaus, aber meines Erachtens immer noch betrachtungswürdig war darüber hinaus, dass ...

- gerade eben keine Interaktion mehr gefunden wurde für die Kombination „Schriftgröße/Zeilenbreite/Bildschirm“ (GROESSE * ZEILENBR * BILD-SCHI) ($p = 0,032$).
- auch eine Interaktion aller typografischen Einflussfaktoren (SCHRIFTA * GROESSE * ZEILEN-AB * ZEILENBR) knapp nicht mehr gefunden wurde ($p = 0,044$).
- die Kombination „Zeilenabstand/Zeilenbreite/Schriftart“ (ZEILENAB * ZEILENBR * SCHRIFTA) noch einen recht niedrigen p-Wert erzeugte ($p = 0,057$).

6.2.3 Exkurs: Bildschirmarten

Die vorliegende Arbeit ist keine ingenieurwissenschaftliche. Ein vertieftes Eindringen in technische Konstruktionsprinzipien von Computer-Monitoren wurde aus diesem Grunde bisher unterlassen und soll in diesem kurzen Exkurs auch nur so weit vorangetrieben werden, als sich aus diesen Konstruktionsbesonderheiten eventuelle Rückschlüsse finden lassen für den erwähnten Befund: Die Lesezeiten an Flachbildschirmen fielen in meinem Experiment merklich besser aus als an Röhrenbildschirmen.

6.2.3.1 Röhrenbildschirme (CRT)

Der Begriff des „Röhrenbildschirms“ leitet sich her aus dem Prinzip der Kathodenstrahlröhre (englisch „cathode ray tube“, kurz CRT). In diesem Konstrukt werden Lichtpunkte auf der

Bildschirmscheibe dergestalt erzeugt, dass Elektronenstrahlen, die von von einer sogenannten „Kanone“ im Innern des Monitorgehäuses abgeschossen werden, auf eine Rasterschicht aus Phosphorteilchen an der Innenseite des Bildschirms treffen, die dadurch zum Leuchten gebracht werden⁸⁹⁷. Dieser Phosphorschicht innen vorgeschaltet findet sich eine Maske, die der leuchtende Strahl passieren muss: Mit dieser Lochmaske wird das Gesamtbild in einzelne Bildpunkte aufgeteilt und eine zielgenaue „Ansteuerung“ des Leuchtstrahls auf einzelne Phosphor-Rasterflecken sichergestellt. Weitere vorgeschaltete Farbfilter sorgen für farbige Eindrücke auf dem Schirm.

Die Bildqualität an CRT-Monitoren ist zunächst abhängig von der Phosphorqualität und der Gleichmäßigkeit der „Körnung“ des Phosphor-Rasters⁸⁹⁸; entscheidend an Röhrenbildschirmen ist jedoch die sogenannte „Bildwiederholrate“; in welcher Frequenz also der „Beschuss“ der Phosphorteilchen an der Bildschirminnenseite erfolgt, wie oft mithin die unweigerliche sukzessive Verblässung beschossener Phosphorteilchen durch neue Aktivierung „aufgefrischt“ wird. Je geringer diese Frequenz ausfällt, desto stärker „flimmert“ das Monitorbild⁸⁹⁹. Dieses Flimmern, das trotz stetig verbesserter Bildwiederholraten der Monitorgenerationen jüngerer Jahrgänge nach wie vor allen CRT-Monitoren zu eigen ist⁹⁰⁰, gilt als Hauptursache für schnellere Ermüdung bei der Bildschirmarbeit; die Augen sind schlicht stärker belastet als beispielsweise beim Rezeptionsprozess von Papier⁹⁰¹.

6.2.3.2 Flachbildschirme

Der Begriff des „Flachbildschirms“ ist zunächst irreführend, weil er verschiedene Technologien verallgemeinernd bezeichnet und zudem – anders als beim „Röhrenbildschirm“ – nicht auf die Darstellungstechnik bezogen ist, sondern auf die Gehäusetiefe. Letztere Eigenart freilich ist Ausfluss aller Flachbildschirm-Technologien, die nämlich durch Verzicht auf die Elektronenstrahl-Kanone des CRT-Bildschirms weit geringere Gehäusetiefen gestatten. Die mutmaßlich weitestverbreitete Flachbildschirm-Variante ist der Dünnschichttransistor (englisch „thin film transistor“, kurz TFT); auf dieser Technologie dürfte heute nahezu jeder handelsübliche Laptop-Monitor und für Normalverbraucher erschwingliche Flach-Computerbildschirme beruhen⁹⁰². Dabei werden Lichteffekte nicht durch eine Kathodenstrahlröhre erzeugt, sondern durch mehrere Leuchtstoffröhren im Gehäuseinnern. Dieses Licht fällt gleichmäßig auf eine Schicht aus Flüssigkristallen (englisch „liquid crystal display“, kurz LCD – Flachbildschirme werden entsprechend auch als „LCD-Schirme“ bezeichnet⁹⁰³). Diese Flüssig-

⁸⁹⁷ Vgl. Lindsay/Norman 1981, S. 241

⁸⁹⁸ Vgl. Ziefle 2002, S. 50

⁸⁹⁹ Vgl. Böhringer et al. 2003, S. 218f.

⁹⁰⁰ Vgl. Erben 1998, S. 24

⁹⁰¹ Vgl. Thissen 2001, S. 43

⁹⁰² Vgl. Thissen 2003, S. 92

⁹⁰³ Vgl. Die Zeit. Das Lexikon Bd. 04, S. 574

sigkristalle werden durch gezieltes Anlegen von elektrischer Spannung lichtdurchlässig und erzeugen auf diese Weise das sichtbare Bild auf dem Bildschirm. Dadurch aber entfällt auch das Phänomen der Bildwiederholfrequenz; TFT- beziehungsweise LCD-Monitore sind aus diesem Grunde praktisch flimmerfrei. Plasmabildschirme, die Konkurrenztechnologie zu TFT-Monitoren, finden unter Computermonitoren zurzeit noch keine sehr weite Verbreitung; sie sind jedoch verbreitetes Konstruktionsprinzip von Flachbildfernsehern.

Bayer weist schließlich auch auf negative Eigenschaften von TFT-Monitoren hin: „Sie haben eine schlechtere Kontrastdarstellung als CRT-Monitore, das heißt, der Unterschied von hellen und dunklen Bildpunkten ist undeutlicher“⁹⁰⁴. Zudem scheint es, als seien Flachbildschirme auf Flüssigkristall-Basis empfindlicher für störendes Umgebungslicht als CRT-Bildschirme⁹⁰⁵.

Dennoch muss festgehalten werden, dass diese Negativa offenkundig den entscheidenden Vorteil von Flachbildschirmen – die praktisch völlige Flimmerfreiheit – nicht konterkarieren können.

6.3 Der Dunnett T3-Test

6.3.1 Prinzip des Tests

Das inferenzstatistische Verfahren der Varianzanalyse allein vermag in seinen Resultaten bereits einiges über Charakter und Eigenarten erhobener Daten zu verraten; und doch liefert sie isoliert nicht genug Aufschluss, um eine ausreichend tiefgehende Analyse von Messwerten im Detail zu gestatten. So ließ die oben beschriebene Varianzanalyse auf Basis der von mir erhobenen Daten zwar beispielsweise deutlich werden, dass der Faktor „Zeilenbreite“ erkennbar einen „Haupteinfluss“ ausübte auf die erhobenen Lesezeiten; in welchen Breiten-Ausprägungen im Detail jedoch sich dieser Einfluss besonders augenfällig manifestierte, wird erst im Einzelvergleich der gemessenen Zeilenbreiten deutlich.

Daher habe ich diverse meiner Messergebnisse zusätzlich sogenannten „Post Hoc“-Vergleichen unterzogen, in denen ich ausgewählte Einzelphänomene paarweise auf Signifikanzen hin abglich. Ich habe dabei wiederum sogenannte „Erwartungswerte“ ermittelt⁹⁰⁶, das heißt, ich berechnete für verschiedene Messwerte, wie groß die Wahrscheinlichkeit ist, dass diese aus identischen Populationen stammen⁹⁰⁷ – wie groß also, umgekehrt formuliert, die Wahrscheinlichkeit ist, dass zwei unterschiedliche Messwerte nicht rein zufällig in dieser Differenz zustande gekommen sind, sondern aufgrund einer vermutet tatsächlichen Unterschiedlichkeit der Ausprägung in realiter.

⁹⁰⁴ Vgl. Bayer 2003, S. 23

⁹⁰⁵ Kiehn/Titzmann 1998, S. 28

⁹⁰⁶ Vgl. Mathes 1993, S.381ff.

⁹⁰⁷ Vgl. Bortz 2004, S. 140

Klassisches Instrument derartiger paarweiser Vergleiche ist der sogenannte t-Test⁹⁰⁸. Problematisch im Rahmen der hier vorgestellten Studie ist allerdings, dass es sich beim t-Test um die Variante eines sogenannten „parametrischen Tests“ handelt⁹⁰⁹. Tests dieser Gattung unterstellen das Vorhandensein einer statistischen Normalverteilung der Messergebnisse⁹¹⁰, die, wie gezeigt, in der vorliegenden Untersuchung auch nach der Datenbereinigung bereits in der Gesamtschau nicht in Reinform festzustellen war (immer noch waren die Messresultate im Ganzen leicht rechts-schief verteilt, was vermuten ließ, dass es in den kleineren Mess-Einheiten nicht durchgehend anders sein würde). Wenn auch als anerkannt gilt, dass „der t-Test gegen Verletzungen relativ robust ist“⁹¹¹, wurden die von mir erhobenen Daten zusätzlich unter einem zweiten Gesichtspunkt den Voraussetzungen paarweiser t-Tests nicht gerecht: Die Varianzen der zu vergleichenden Detail-Stichproben erwiesen sich als zu wenig homogen in meinem Ergebnispool, um den Einsatz des t-Tests zu rechtfertigen⁹¹². Mit anderen Worten: Die Standardabweichungen innerhalb der meisten Detail-Stichproben, die ich zu vergleichen trachtete (zum Beispiel Zeilenbreite „0“ und Zeilenbreite „1“), variierten nach den Voraussetzungskriterien des t-Tests⁹¹³ zu stark (dies zeigte sich bereits nach wenigen Durchführungen des „F-Tests“, mittels dessen die Varianzhomogenität zweier Stichproben überprüft werden kann⁹¹⁴).

Es führte also kein Weg daran vorbei: Der t-Test schien für meine Zwecke das ungeeignete Instrumentarium. Nach einigem Suchen stieß ich jedoch auf ein elaborierteres Verfahren des paarweisen Abgleichs, das ausdrücklich für die Gegenüberstellung von Stichproben mit unhomogeneren Varianzen entwickelt wurde: den „Dunnett T3“-Test⁹¹⁵. Auf der Grundlage dieses Verfahrens wurden die im Folgenden dargestellten paarweisen Vergleiche angestellt.

6.3.2 Darstellung der Dunnett-Ergebnisse in dieser Arbeit

Insbesondere die Ergebnisse der paarweisen Vergleiche nach der „Dunnett T3“-Methode werden im Folgenden teilweise tabellarisch, teilweise ausschließlich verbal dokumentiert. Die tabellarische Darstellung der „Dunnett T3“-Ergebnisse habe ich aus Raum- und Übersichtlichkeitsgründen auf Abgleiche zwischen maximal 5 Einzelwerten beschränkt; diese tabellarische Darstellung erfolgt dabei nach einem gleichbleibenden Muster, das ich an dieser Stelle kurz in Struktur und Kürzelwahl erläutern möchte.

⁹⁰⁸ Vgl. Heinecke/Köpcke O.J.

⁹⁰⁹ Vgl. Heinecke/Köpcke O.J.

⁹¹⁰ Vgl. Mathes 1993, S.381f.

⁹¹¹ Vgl. Hüttner 2005

⁹¹² Vgl. Bortz 2004, S. 141

⁹¹³ Vgl. Mittenecker 1966, S. 79f.

⁹¹⁴ Vgl. Hofstätter/Wendt 1974, S. 100f.

⁹¹⁵ Vgl. SPSS Inc. 2003, S. 13

Tabelle 6.3: Paarweiser Vergleich der Schriftarten							
	VB –	Ø	VB +	n	S	p (Ve.)	p (Tim.)
Ve.	26,62	27,17	27,72	1800	7,40	–	0,250
Tim.	26,97	27,51	28,05	1800	7,11	0,250	–
Quelle: eigene Erhebung							

Als Beispiel ziehe ich die Ergebnisse des isolierten Vergleichs der im Experiment dargebotenen Schriftarten *Verdana* und *Times* hinzu.

In der jeweils linken Spalte sind in den tabellarischen Darstellungen stets die gemessenen unabhängigen Variablen verzeichnet. Im angeführten Beispiel sind dies die Schriftarten *Verdana* (hier abgekürzt als „Ve.“) und *Times* (hier abgekürzt als „Tim.“).

Die mit „Ø“ überschriebene Spalte gibt für jede Variable die ermittelte durchschnittliche Lesezeit in Sekunden an. Der Beispiel-Tabelle ist also zu entnehmen, dass Texte, die den Probanden in typografischen Anordnungen mit der *Verdana* als Schriftart angeboten wurden, im Schnitt in 27,17 Sekunden bearbeitet wurden, Texte in der Schriftart *Times* dagegen in 27,51 Sekunden. Die mit „n“ betitelte Spalte gibt wieder, aus wie vielen ermittelten Lesezeiten dieser Durchschnittswert ermittelt wurde (in diesem Falle waren es für beide Schriftarten 1800 gemessene Lesezeiten), die mit „S“ bezeichnete Spalte verzeichnet die Standardabweichungen, die sich aus den erhobenen Zeiten ergeben haben (also beispielsweise 7,11 für die Schriftart *Times*).

Zusätzlich ist in jeder Tabelle der sogenannte „Vertrauensbereich“ oder auch „Konfidenzintervall“⁹¹⁶ verzeichnet. Dieser „Vertrauensbereich“ bezeichnet die Marge, um die der „tatsächliche“ Wert in der angenommenen Gesamtpopulation mit einer Wahrscheinlichkeit von 95 % vom berechneten Mittelwert maximal abweicht⁹¹⁷. In den Tabellen der folgenden Teilkapitel sind diese Abweichungen – ebenfalls ermittelt nach der „Dunnett T3“-Methodik – rubriziert als „VB–“ (also „Vertrauensbereich minimal“) sowie als „VB+“ („Vertrauensbereich maximal“).

Schließlich – und dies sind die entscheidenden Werte einer jeden Tabelle, die im Folgenden präsentiert wird – wurden die ermittelten Werte eben besagtem wechselseitigen Abgleich nach „Dunnett T3“ unterzogen. Die aus diesen Tests ermittelten p-Werte sind in den jeweils rechten Spalten der Tabellen kreuztabellarisch verzeichnet. Sie bezeichnen den Grad der Wahrscheinlichkeit, dass ein gemessener Unterschied in den erhobenen Werten nicht durch puren Zufall zustande gekommen ist. Auch diese Darstellungsvariante sei näher erläutert.

In Zeile 2 der Beispiel-Tabelle sind die gemessenen und errechneten Werte für die Schriftart *Verdana* verzeichnet. Die achte und letzte Spalte dieser Zeile ist übertitelt mit „p (Tim.)“. In dieser Zelle findet also sich der p-Wert, der angibt, wie groß nach dem „Dunnett T3“-Test die Wahrscheinlichkeit ist, dass der gemessene Unterschied in den Lesezeiten zwischen der *Verdana* und der *Times* (abgekürzt „Tim.“) im durchgeführten Experiment (immerhin 0,34

⁹¹⁶ Vgl. Mathes 1993, S. 392f.

⁹¹⁷ Vgl. Bortz 2004, S. 101ff.

Sekunden) nur zufällig zustande gekommen ist. Diese „Irrtumswahrscheinlichkeit“ liegt bei statistischen 25 %, was sich im p-Wert „0,25“ wiederfindet. Dies aber bedeutet auch, dass der gemessene Unterschied zwischen den Schriftarten *Verdana* und *Times* statistisch nicht signifikant ist – diese Annahme nämlich setzt eine Irrtumswahrscheinlichkeit, also ein Signifikanzniveau von maximal 5 % oder eben einen p-Wert von „0,05“ voraus. Mit anderen Worten: Es gibt keinen verlässlichen Anhaltspunkt, dass die *Verdana*, wiewohl sie im Schnitt der von mir erhobenen Messdaten etwas schneller gelesen wurde als die *Times*, der Serifenschrift wirklich überlegen ist in der Qualität ihrer Lesbarkeit.

6.4 Schriftarten

6.4.1 Isoliertes Ergebnis

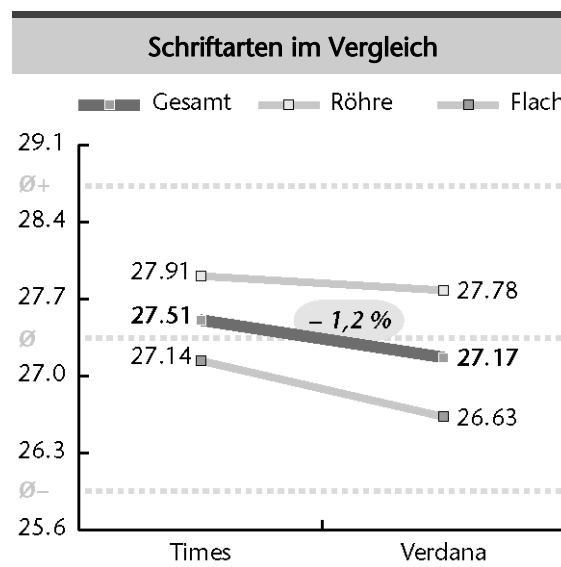
Wie gezeigt, finden sich in der Literatur nur wenige Verfechter des Einsatzes von serifentragenden Schriften im Web. Insbesondere wird argumentiert, dass die Abbildung subtiler Strichstärken-Differenzen und feiner Serifen an niedrig auflösenden Bildschirmen kaum möglich sei, zumindest nicht im Bereich kleinerer Schriftgrößen. Die Zeichengestalt von serifentragenden Schriftarten mit ausdifferenzierteren Strichstärken sei also am Bildschirm korrumpiert – weshalb von ihrem Einsatz abzuraten sei. Die Ergebnisse meiner Studie bestätigen diese herrschende Meinung allerdings nicht.

So fielen die Lesezeiten jener Probanden, denen ein Text in der Schriftart *Times*

vorlag, gerade einmal um 1,2 Prozent langsamer aus als die der Teilnehmer, die Texte in der Schriftart *Verdana* zu rezipieren hatten (siehe Abbildung 6.3).

Eine vernachlässigbare Differenz, auch in statistischem Sinn: Ein Haupteinfluss der Schriftart auf die Leseergebnisse war laut Varianzanalyse nicht vorhanden ($p=0,382$).

Der direkte Vergleich im „Dunnett T3“-Test bestätigt diese Erkenntnis. Auch hier konnte kein statistisch signifikanter Unterschied festgestellt werden ($p=0,25$; vergleiche Tabelle 6.4). Zusammengefasst: Ob man auf einer Webseite die *Times* einsetzt oder die *Verdana*, es beeinflusst die Lesbarkeit zumindest statistisch nicht signifikant.



Ermittelte Durchschnitts-Lesezeiten für die beiden Schriftarten, absolut und aufgeschlüsselt nach Monitorarten (in Sekunden; beide Datenpunkte „Gesamt“ $n=1800$, Datenpunkte „Röhre“ $846 < n < 873$, Datenpunkte „Flach“ $927 < n < 954$). (\emptyset = Durchschnitts-Messzeit über alle erhobenen Konstellationen hinweg [27,34 Sekunden], $\emptyset+$ = Durchschnitts-Messzeit über alle erhobenen Konstellationen hinweg plus 5 Prozent [28,71 Sekunden], $\emptyset-$ = Durchschnitts-Messzeit über alle erhobenen Konstellationen hinweg minus 5 Prozent [25,97 Sekunden]).

ABBILDUNG 6.3

Tabelle 6.4: Paarweiser Vergleich der Schriftarten							
	VB -	Ø	VB +	n	S	p (Ve.)	p (Tim.)
Ve.	26,62	27,17	27,72	1800	7,40	-	0,250
Tim.	26,97	27,51	28,05	1800	7,11	0,250	-

Quelle: eigene Erhebung

Auch das Zusammenspiel mit den beiden Bildschirmarten beeinflusste die Lesegeschwindigkeiten im Übrigen nicht signifikant

($p=0,270$); zwar fielen die Lesezeiten für die *Verdana* an den Flachbildschirmen etwas besser aus als an Röhrenbildschirmen; jedoch lag die *Verdana* auch an den Flachbildschirmen nur um knapp 1,9 Prozent und damit ebenfalls nicht signifikant besser als die *Times* (am Röhrenbildschirm waren es gar nur 0,5 Prozent). Auf eine tabellarische Detailbetrachtung der Schriftarten-Ergebnisse in Kombination mit den erfassten Bildschirmarten kann also an dieser Stelle verzichtet werden.

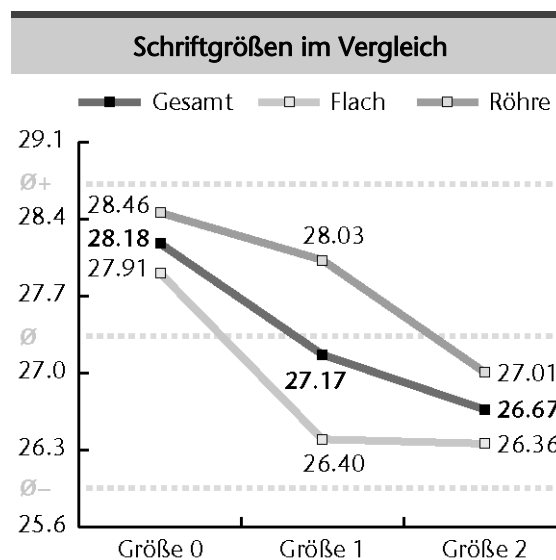
6.5 Schriftgrößen

6.5.1 Isoliertes Ergebnis

Bei den Schriftgrößen ergibt sich, auf den ersten Blick auch auf die nebenstehende Grafik, ein eindeutiges Bild: Je größer die Schrift, so scheint es, desto besser die Lesbarkeit. Eine detailliertere Betrachtung offenbart allerdings auch, dass das Optimum an Lesbarkeit an Flachbildschirmen augenscheinlich bei merklich kleineren Größen erreicht scheint als an Röhren-Monitoren.

Zunächst jedoch sei ein Blick geworfen auf die Resultate der Varianzanalyse. Dessen Ergebnis weist den Faktor „Schriftgröße“ tatsächlich als Aspekt mit statistischem Haupteinfluss ($p < 0,001$) auf die Leseergebnisse insgesamt aus. Hier ist jedoch zu berücksichtigen – der Vorgriff sei gestattet –, dass auch die Kombination von Schriftgröße und Schriftart als maßgeblicher Einflussfaktor in der Varianzanalyse identifiziert ist. Das heißt: Die Schriftgröße muss vorläufig unter Einschränkungen als Haupteinfluss-Faktor betrachtet werden. An der Bemerkenswürdigkeit der in Abbildung 6.4 visualisierten Ergebnisse ändert dies freilich wenig.

Tabelle 6.5 weist aus, dass sich im Gesamt-Abgleich der drei getesteten Schriftgrößen die Größe „0“ (Buchstaben mit der x-Höhe 6 Pixel) als um 3,6 Prozent langsamer und damit si-



Ermittelte Durchschnitts-Lesezeiten der 3 Schriftgrößen, absolut und aufgeschlüsselt nach Bildschirmarten (Angaben in Sekunden; alle Datenpunkte „Gesamt“ $n=1200$, Datenpunkte „Röhre“ $560 < n < 590$, Datenpunkte „Flach“ $610 < n < 640$).

ABBILDUNG 6.4

gnifikant schlechter lesbar erwies als Größe „1“ (7 Pixel Mittellänge, $p < 0,01$). Sogar um 5,4 Prozent schneller gelesen als Schriften in der Größe „0“ wurden Schriften in der Größe „2“ (8 Pixel x-Höhe, $p < 0,001$).

Tabelle 6.5: Paarweiser Vergleich der Schriftgrößen								
	VB –	Ø	VB +	n	S	p (Gr. 0)	p (Gr. 1)	p (Gr. 2)
Gr. 0	27,48	28,18	28,89	1200	7,30	–	0,002	0,000
Gr. 1	26,47	27,17	27,87	1200	7,19	0,002	–	0,240
Gr. 2	25,96	26,67	27,37	1200	7,20	0,000	0,240	–
<i>Quelle: eigene Erhebung</i>								

Zwischen Größe „1“ und Größe „2“ ergab sich dagegen ein durchaus messbarer, aber kein

statistisch signifikanter und wohl auch kaum praxisrelevanter Unterschied (Beschleunigung der gemessenen Lesezeiten um 1,8 Prozent, $p = 0,240$). Zwar verbesserten sich die gemessenen Zeiten nochmals mit der Schriftvergrößerung, aber eben nur unerheblich.

6.5.1.1 Ergebnisse am Flachbildschirm

Bereits ein flüchtiger Blick auf Abbildung 6.4 legt einen Schluss nahe: Welche Qualität eine installierte Schriftgröße auf Webseiten entfaltet, hängt durchaus von der Art des Monitors ab, auf der sie betrachtet wird. Zwar fand sich in der Varianzanalyse kein signifikanter Einfluss der Faktorenkombination „Schriftgröße – Bildschirm“ ($p = 0,222$); ein kurzer Blick auf die Aufschlüsselung nach Monitorbauart erscheint mir an dieser Stelle dennoch interessant.

Tabelle 6.6: Paarweiser Vergleich der Schriftgrößen (Flachbildschirme)								
	VB –	Ø	VB +	n	S	p (Gr. 0)	p (Gr. 1)	p (Gr. 2)
Gr. 0	27,33	27,91	28,49	610	7,20	–	0,001	0,000
Gr. 1	25,84	26,40	26,96	631	7,06	0,001	–	1,000
Gr. 2	25,80	26,36	26,92	640	7,06	0,000	1,000	–
<i>Quelle: eigene Erhebung</i>								

So legt bereits Tabelle 6.6 den Schluss nahe, dass sich die Lesbarkeit von Webtexten an Flach-

bildschirmen erheblich verbessert, richtet man sie nicht in einer Schriftgröße ein, die den Buchstaben eine Mittellänge von 6 Pixeln verschafft, sondern eine Stufe größer (die Lesegeschwindigkeit steigt um immerhin 5,4 Prozent). Der Übergang von Größe „0“ zu Größe „1“ am Flachbildschirm zeitigt damit eine statistisch hochsignifikante Beschleunigung der Lesegeschwindigkeit ($p < 0,01$).

Desweiteren zeigt die Auswertung aber auch, dass damit offenbar das Optimum erreicht ist: Eine weitere Vergrößerung der Schrift (von Größe „1“ auf Größe „2“) beschleunigt den Leseprozess augenscheinlich nicht mehr (Beschleunigung der Lesezeit um 0,15 Prozent, $p = 1,000$). Die *Verdana* erreicht nach dieser vorläufigen Erkenntnis ihre optimale Lesbarkeit am Flachbildschirm also in einer Größe von 13 Pixeln, die *Times* bei 15 Pixeln.

6.5.1.2 Ergebnisse am Röhrenbildschirm

Tabelle 6.7: Paarweiser Vergleich der Schriftgrößen (Röhrenbildschirme)

	VB –	Ø	VB +	n	S	p (Gr. 0)	p (Gr. 1)	p (Gr. 2)
Gr. 0	27,85	28,46	29,07	590	7,38	–	0,675	0,003
Gr. 1	27,42	28,03	28,64	569	7,25	0,675	–	0,057
Gr. 2	26,39	27,01	27,63	560	7,33	0,003	0,057	–

Quelle: eigene Erhebung

Etwas differenzierter stellt sich das Ergebnis dar, das die Auswertung der Ergebnisse ergibt, die

an Röhrenmonitoren erzielt wurden. Wie Tabelle 6.7 ausweist, fielen die Lesezeiten für Größe „2“ signifikant besser aus als die für die kleinste gemessene Größe – Größe „0“ wurde durchschnittlich 5,1 Prozent schneller gelesen als Größe „0“ ($p < 0,01$). Auch im Abgleich mit Größe „1“ fiel die höchste gemessene Schriftgröße um 3,6 Prozent schneller aus, allerdings knapp nicht signifikant ($p = 0,057$). Die Unterschiede zwischen Größe „0“ und Größe „1“ schienen dagegen überschaubar (Beschleunigung um 1,5 Prozent zugunsten der höheren Größe, $p = 0,675$).

Insgesamt scheint hier also tatsächlich die Faustformel zu greifen: „Je größer, desto besser lesbar“. Ein „Sättigungsgrad“ analog zu den Flachbildschirm-Ergebnissen ist aus diesen Ergebnissen freilich nicht deduzierbar – ob eine weitere Vergrößerung über die tatsächlich gemessenen „Größen 2“ an Röhrenbildschirmen eine weitere Beschleunigung bewirkt hätte, muss spekulativ bleiben.

6.5.2 Zwischenfazit: Schriftgrößen

Das Zwischenfazit für die gemessenen Resultate der verschiedenen Schriftgrößen muss differenziert ausfallen. Relativ eindeutig ist der Befund, dass die „Größen 0“ in beiden Schriftarten (das entspricht der *Verdana* in 11 Pixeln Größe und der *Times* in 13 Pixeln) zumindest relativ am schlechtesten abschnitten – wer also Schriften in diesen Größen-Maßen einrichtet auf einer Webseite, sollte sich wenigstens vor Augen halten, dass er sich – nach den vorliegenden Ergebnissen – unabhängig von der Monitorart eindeutig Optimierungspotenzial vor-enthält.

Schwieriger ist zu klären, inwiefern sich ein vermeintlicher Optimalwert aus den Ergebnissen deduzieren lässt. Tatsächlich scheint es, als entfalte an Flachbildschirmen eine Schriftkonstellation von „13 Pixel *Verdana*“ beziehungsweise „15 Pixel *Times*“ eine sehr gute Lesbarkeit, die durch weitere Vergrößerungen praktisch nicht mehr optimierbar ist.

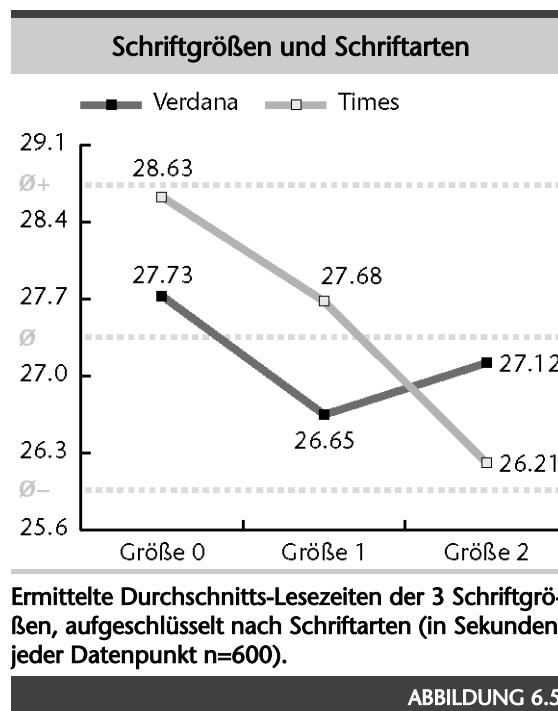
In Bezug auf Röhrenbildschirme stellt sich das Ergebnis schwieriger dar. Hier schnitten die jeweils größten gemessenen Schriftgrößen eindeutig besser ab als die beiden kleineren Abstufungen. Wo am Flachbildschirm das Optimum erreicht scheint (nämlich bei Größe „1“), ist am Röhrenbildschirm die durchschnittliche gemessene Lesezeit für die nächst größere Stufe

nicht signifikant, aber immer noch merklich verbessert. So lange also noch eine größere Anzahl von Nutzern mit Röhrenbildschirmen im Web unterwegs ist, sollten Schriften eher einen Ticken zu groß als zu klein gesetzt werden: Am Flachbildschirm schadet es nicht, am Röhrenbildschirm fördert es offenbar die Lesegeschwindigkeit.

6.5.3 Schriftgröße und Schriftart im Zusammenwirken

6.5.3.1 Ergebnisse an allen Monitorarten

Ich habe es bereits angedeutet: Die Varianzanalyse ergab nicht nur einen Haupteffekt des Faktors „Schriftgröße“ auf die erfassten Lesezeiten, sondern auch eine klare Interaktion der Faktoren „Schriftgröße“ und „Schriftart“, die also im Zusammenwirken ebenfalls maßgeblich die Lesbarkeit einer Schriftanordnung beeinflussen ($p < 0,01$). Doch damit nicht genug: Das kombinierte Zusammenwirken von „Schriftart“ und „Schriftgröße“ wird zusätzlich mit beeinflusst durch die Konstruktionsart des Bildschirms, an dem sie betrachtet wird (in der Varianzanalyse $p = 0,030$). Es empfiehlt sich also ein Detail-Blick auf die Interaktion dieser beiden typografischen und des monitorspezifischen Faktors; Abbildung 6.5 vermittelt ihn, zunächst unabhängig von der Bildschirmart.



Für diesen paarweisen Abgleich der 6 verglichenen Werte ergaben sich im „Dunnet T3“-Test folgende Signifikanzen:

- die *Times* in der Schriftgröße „0“ wurde signifikant langsamer gelesen als die *Verdana* in den Schriftgrößen „1“ und „2“.
- die *Times* in Schriftgröße „2“ wurde signifikant schneller gelesen als die *Times* in den Schriftgrößen „0“ und „1“ und die *Verdana* in Schriftgröße „0“.

Es ergaben sich im Gesamtabgleich folgende Lesezeit-Differenzen von mehr als 5 % in folgenden Konstellationen (die eben nur zum Teil signifikant ausfielen, aber an dieser Stelle doch erwähnenswert scheinen durch Überschreitung jener 5-Prozent-Grenze):

- die *Times* wurde in der Schriftgröße „17 Pixel“ (Größe „2“) um 5,3 Prozent schneller gelesen als in der Größe „15 Pixel“ (Größe „1“) und um 8,5 Prozent schneller als in der Größe

„13 Pixel“ (Größe „0“).

- die *Times* wurde in der Schriftgröße „17 Pixel“ (Größe „2“) um 5,5 Prozent schneller gelesen als die *Verdana* in „11 Pixel“ Größe (Größe „0“).
- die *Verdana* wurde in der Größe „13 Pixel“ (Größe „1“) um 6,9 Prozent schneller gelesen als die *Times* in der Größe „13 Pixel“ (Größe „0“).
- die *Verdana* wurde in der Größe „14 Pixel“ (Größe „2“) um 5,3 Prozent schneller gelesen als die *Times* in der Größe „13 Pixel“ (Größe „0“).

Insgesamt zeigt sich, dass die – wie gezeigt offenbar ohnehin nur bedingt richtige – Faustregel „größere Schrift = bessere Lesbarkeit“ nur für die Serifenschriftart *Times* Gültigkeit zu besitzen scheint. Die Lesezeiten der Schriftart *Verdana* dagegen verbesserten sich zwar merklich im Übergang von 11 zu 13 Pixeln Schriftgröße, erhöhten sich aber wieder bei einer weiteren Vergrößerung auf 14 Pixel. Mehr noch: Über alle Bildschirmarten hinweg erzielte die beste durchschnittliche Lesezeit die *Times* in der größten gemessenen Schriftgröße das absolut beste Resultat.

Das ist bemerkenswert – und womöglich ein Hinweis darauf, dass die *Verdana* in kleineren Größen zwar durchaus von ihren stabilen Strichstärken zu profitieren vermag, dass sich dieser Vorzug der „stämmigen“ Striche aber bereits ab 14 Pixel Größe in sein Gegenteil verkehrt. Eventuell wirkt die *Verdana* ab dieser Größe schon fast wieder zu „klobig“. Eine mögliche Interpretation der Messwerte für die *Times* ist hingegen, dass diese Schriftart mit wachsender Darstellungsgröße (und damit eben auch mehr Pixeln zur Darstellung der Lettern) ihre ausdifferenzierte Schriftgestalt immer besser „auf den Bildschirm“ und zur Geltung bringt.

6.5.3.2 Ergebnisse nach Bildschirmarten

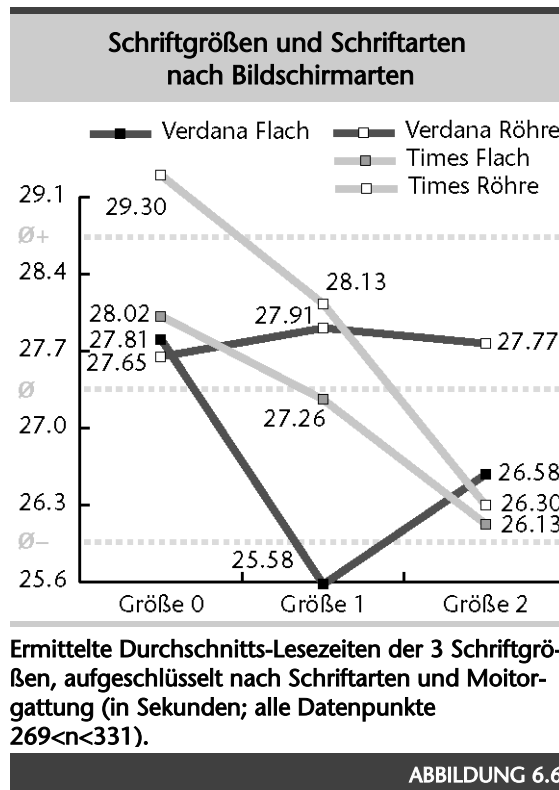
Freilich müssen die eben getroffenen Feststellungen teilweise wieder unter Vorbehalt gestellt werden, analysiert man die Ergebnisse derjenigen, die das Experiment an einem Röhrenbildschirm exerzierten, getrennt von jenen, die die Studie an einem Flachbildschirm durchführten. Dies geschieht in den Abbildung 6.6.

Denn diese Betrachtung legt eine nochmals differenziertere Bewertung insbesondere der *Verdana*-Ergebnisse nahe. So wurde die *Verdana* am Flachbildschirm, gesetzt in 7 Pixel x-Höhe (Größe „1“, numerische Schriftgröße 13 Pixel), durchschnittlich um mindestens 7,2 Prozent schneller gelesen als alle anderen Konstellationen in derselben oder der kleineren Schriftgröße – zunächst ohne Berücksichtigung der Bildschirmart. Nicht durchgehend erheblich, aber immer noch merkbar schlechter als die *Verdana* in Größe „1“ fielen auch alle größer gesetzten Konstellationen aus. Gegenüber der *Times*, gesetzt in 6 Pixel Mittellänge und dargeboten an einem Röhrenbildschirm, fiel die Lesegeschwindigkeit der mittelgroßen *Verdana* am Flachbildschirm sogar um stattliche 12,7 Prozent besser aus.

Tatsächlich erwies sich die *Verdana* am Flachbildschirm, gesetzt in 13 Pixel Größe, als signifikant schneller lesbar als sämtliche Konstellationen in der Größe „0“; ebenfalls signifikant besser fiel diese Konstellation aus im Abgleich mit allen *Verdana*-Anordnungen an Röhrenbildschirmen sowie im Vergleich mit allen *Times*-Konstellationen oberhalb von Durchschnittszeiten von 28 Sekunden.

Die Lesezeiten für die *Times* an Röhrenbildschirmen in Größe „0“ fielen zudem signifikant schlechter aus als alle Konstellationen in Größe „2“ mit Ausnahme der *Verdana* am Röhrenbildschirm. Eine Vergrößerung der *Times* am Flachbildschirm von 13 auf 17 Pixel ergab ebenfalls signifikante Lesezeit-Verbesserungen. Schließlich wurde auch die *Times* am Röhrenbildschirm in Größe „1“ signifikant schlechter gelesen als die *Times* am Flachbildschirm in Größe „2“.

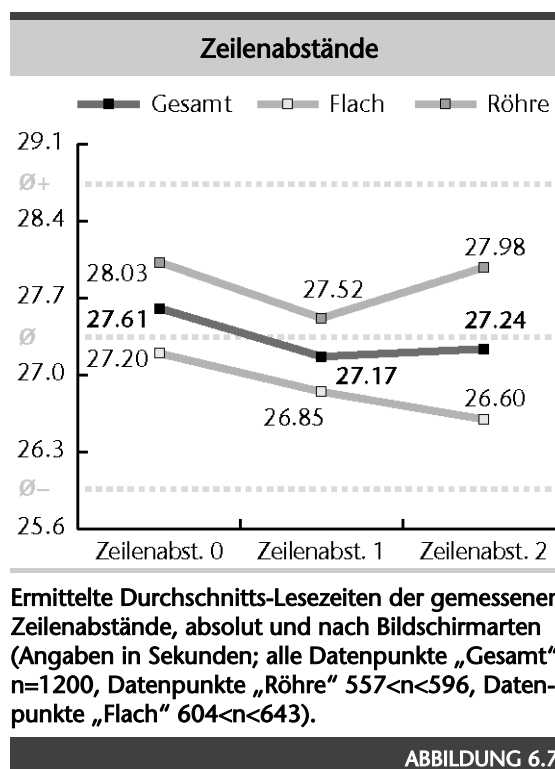
Insgesamt zeigt sich, dass die *Times* an beiden Monitorarten umso schneller gelesen wurde, je größer sie gesetzt war. Diese Entwicklung war an Röhrenbildschirmen noch merklich ausgeprägter zu beobachten als an Flachbildschirmen: An Röhrenbildschirmen verbesserten sich die Lesezeiten bei Vergrößerung von 15 auf 17 Pixel (Größe „1“ zu Größe „2“) um 6,5 Prozent, bei Vergrößerung von 13 auf 17 Pixel (Größe „0“ zu Größe „2“) gar um 10,2 Prozent. Ganz anders die *Verdana*: Bemerkenswert ist, dass diese Schriftart an den Röhrenbildschirmen (für die sie wohl konzipiert und geschnitten wurde, denn im Erstellungsjahr 1997 waren Flachbildschirme nahezu unbezahlbarer Luxus) ein erstaunlich konstantes Lesegeschwindigkeits-Niveau über alle gemessenen Schriftgrößen hinweg zu halten imstande war. Sollte dies die Absicht von *Verdana*-Schöpfer Matthew Carter gewesen sein, so muss man ihm tadellose Arbeit bescheinigen. An Flachbildschirmen dagegen entwickelten sich die durchschnittlichen Lesezeiten für die *Verdana* rückläufig (um knapp 4 Prozent) bei der Vergrößerung von mittlerer zu höchster gemessener Schriftgröße. In der größten Messgröße lag die *Verdana* hier praktisch gleichauf mit der *Times*. Wirklich herausragend in der Lesbarkeit erwies sich die *Verdana* an Flachbildschirmen also eigentlich nur in einer Schriftgröße (13 Pixel nämlich) – in eben dieser Ausprägung allerdings war sie den übrigen Testkonstellationen, wie gezeigt, mehr oder minder merklich überlegen.



6.6 Zeilenabstand

6.6.1 Isoliertes Ergebnis

Die gemessenen Zeilenabstände übten laut Varianzanalyse keinen Haupteinfluss auf die Leseergebnisse im Experiment aus ($p=0,296$). Auch im wechselseitigen Abgleich im „Dunnett T3“-Test ergaben sich keine statistischen Signifikanzen (siehe Tabelle 6.8) – was angesichts des bemerkenswert kleinen Korridors zwischen bester und schlechtester gemessener Durchschnittszeit (nur 0,44 Sekunden oder 1,6 Prozent lagen, über alle Bildschirmarten hinweg, zwischen „Zeilenabstand 0“ und „Zeilenabstand 1“) kaum überrascht. Ebenso zeigt die Varianzanalyse auch unter Einbeziehung des Faktors „Bildschirmart“ keine bedeutsamen Einflüsse des Zeilenabstands auf die Lesezeiten ($p=0,657$).



	VB -	Ø	VB +	n	S	p (ZA. 0)	p (ZA. 1)	p (ZA. 2)
ZA. 0	26,89	27,61	28,32	1200	7,20	-	0,342	0,497
ZA. 1	26,46	27,17	27,88	1200	7,21	0,342	-	0,994
ZA. 2	26,52	27,24	27,94	1200	7,35	0,479	0,994	-

Quelle: eigene Erhebung

Es lohnt also nicht, den Aspekt „Bildschirmart“ im Zusammenhang mit dem typografischen

Faktor „Zeilenabstand“ über Gebühr weiterzuverfolgen. Denn wie Abbildung 6.7 ebenfalls verdeutlicht, resultiert der scheinbare „Stillstand“ im Bereich der „Zeilenabstände 2“ aus einer geringen Diskrepanz zwischen den Ergebnissen, die an Flachbildschirmen erzielt würden und solchen, die an Röhrenbildschirmen zustande kamen – diese auseinanderklaffenden Teilergebnisse summieren sich in der Gesamtschau zu einer nur scheinbaren Stagnation.

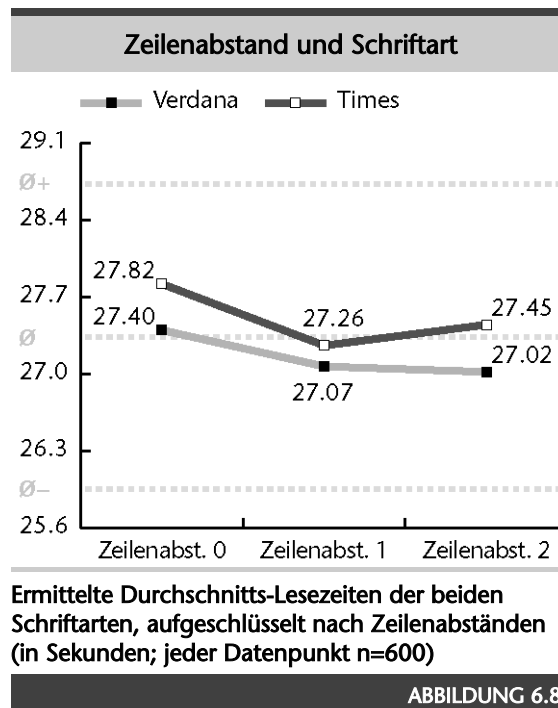
Wie Tabelle 6.8 insgesamt verdeutlicht, fanden sich für den Faktor „Zeilenabstand“ keinerlei signifikante Unterschiede in der paarweisen Betrachtung. Der einzige prozentual auffällige Unterschied war, dass die Zeilenabstände „2“ an Flachbildschirmen um 5,1 Prozent schneller gelesen wurden als die Zeilenabstände „0“ an Röhrenbildschirmen.

6.6.2 Zeilenabstand und Schriftart im Zusammenwirken

Abbildung 6.8 legt bereits in ihrer geringen Spektabilität nahe, was die Varianzanalyse nachwies: Das Zusammenwirken der typografischen Faktoren „Zeilenabstand“ und „Schriftart“ zeitigte keine signifikanten, geschweige denn praxisrelevanten Ergebnisse ($p=0,961$).

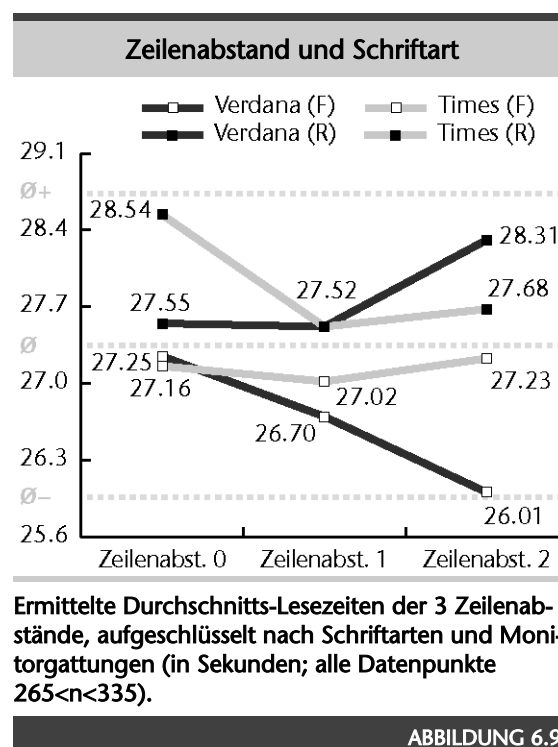
Entsprechend fanden sich auch keinerlei statistisch signifikante Unterschiede im paarweisen Abgleich der sechs Vergleichswerte. Es fanden sich auch keine wechselseitigen Abstände über 5 Prozent in dieser Konstellation.

Wohl aber ergab sich in der Varianzanalyse ein bedeutsamer Einfluss der Kombination „Zeilenabstand/Schriftart“ unter Einbezug des Faktors „Bildschirm“ ($p=0,028$). Daher werde ich im Folgenden diese augenscheinlich interessante „Dreier-Konstellation“ noch einmal detaillierter beleuchten.



6.6.3 Zeilenabstand, Schriftart und Bildschirm im Zusammenwirken

In der kombinierten Betrachtung der Faktoren „Zeilenabstand“, „Schriftart“ und „Monitorgattung“ ergeben sich wiederum auffällige Ergebnisse für die Schriftart *Verdana*. Während diese nämlich an Flachbildschirmen umso bessere Leseergebnisse induziert, je großzügiger der Zeilenabstand konstruiert ist, geht die Tendenz an Röhrenbildschirmen gerade in die andere Richtung: An diesen Monitoren finden sich keine Unterschiede zwischen den Zeilenabständen „0“ und „1“, eine weitere Vergrößerung des Durchschusses jedoch führt sogar zu einer leichten Verschlechterung der Werte um 2,8 Prozent.

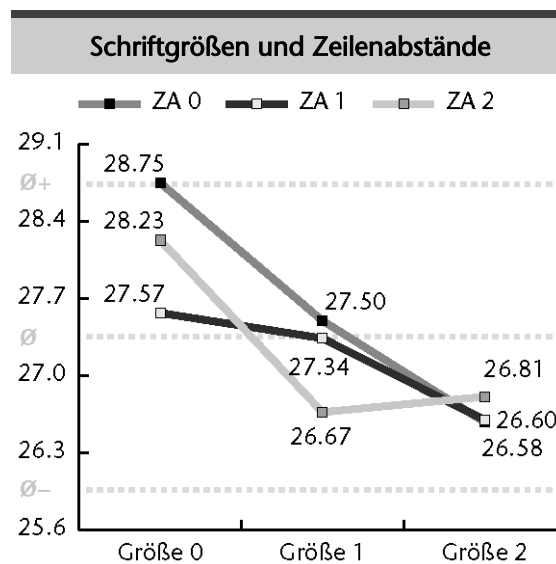


Der paarweise Abgleich aller Messergebnisse ergab, dass die *Verdana* in Zeilenabstand „2“ am Flachbildschirm signifikant schneller gelesen wurde als an Röhrenbildschirmen (und zwar um 8,1 Prozent schneller). Gleichfalls fielen die Messzeiten für die *Verdana* in Zeilenabstand „2“ an Flachbildschirmen signifikant besser aus als die der *Times* an Röhrenbildschirmen in Zeilenabstand „0“ (Beschleunigung: 8,9 Prozent).

Innerhalb der Messergebnisse für die *Times* hingegen waren so gut wie keine (zumindest, wie ich finde, wirklich praxisrelevanten, geschweige denn statistisch signifikanten) Einflüsse der Zeilenabstands-Setzungen feststellbar.

6.6.4 Zeilenabstand und Schriftgröße im Zusammenwirken

Das Zusammenwirken der typografischen Faktoren „Schriftgröße“ und „Zeilenabstand“ scheint, den Ergebnissen meiner Studie nach, keinen merklichen Einfluss zu nehmen auf die Lesbarkeit einer Schriftanordnung. Zumindest war der Varianzanalyse zufolge kein Hauptinfluss auf die Lesegeschwindigkeiten zu verzeichnen für diese Kombination ($p=0,137$). Die nebenstehende Grafik (Abbildung 6.10) gibt die kombinierten Messwerte wieder.



Ermittelte Durchschnitts-Lesezeiten der 3 Schriftgrößen, aufgeschlüsselt nach Zeilenabständen (in Sekunden; alle Datenpunkte n=400).

ABBILDUNG 6.10

Die Grafik freilich legt zumindest einige interessante Schlüsse nahe. So zum Beispiel die Feststellung, dass es spätestens bei einer Schriftgröße von 8 Pixeln x-Höhe keine Rolle mehr zu spielen scheint für die Lesbarkeit, wie groß der Zeilenabstand definiert wird. Nur marginal scheint der Effekt auch bei der nächstkleineren Schriftgröße „1“ zu sein; immerhin setzte sich hier der größte gemessene Zeilenabstand schon etwas deutlicher von den ersten beiden ab. Freilich ist die gemessene Verbesserung der Lesezeiten von rund 2,5 Prozent gegenüber den beiden anderen gemessenen Werten wohl zu marginal, um als praxisrelevant gelten zu können. Immerhin ist festzuhalten, dass alle Ergebnisse für die Schriftgröße „2“ um mindestens 5,0, maximal 7,5 Prozent besser ausfielen als für die Schriftgrößen „0“ in Zeilenabstand „0“ oder „2“.

Insgesamt stellte sich im paarweisen „Dunnett T3“-Abgleich heraus, dass Schriftgröße „0“ mit Zeilenabstand „0“ signifikant langsamer gelesen wurde als alle Konstellationen in Größe „2“ sowie in Größe „1“ im Zeilenabstand „2“. Weitere Signifikanzen fanden sich nicht.

Auch eine Detailbetrachtung des Zusammenwirkens von Schriftgröße und Zeilenabstand unter Einbeziehung des Faktors „Bildschirmart“ scheint nach den Resultaten der Varianzanalyse nicht lohnend ($p=0,822$) – daher wird an dieser Stelle darauf verzichtet.

6.7 Zeilenbreite

6.7.1 Isoliertes Ergebnis

6.7.1.1 An allen Bildschirmen

Ein durchaus überraschendes Ergebnis erbringt der Vergleich der Messwerte für die verschiedenen Zeilenbreiten, die, wie erläutert, im Experiment nicht über absolute Maße bestimmt wurden, sondern über die durchschnittliche Anzahl der Zeichen oder „Anschläge“ pro Zeile.

Die Varianzanalyse weist den Faktor „Zeilenbreite“ als starken, wenn nicht stärksten Einflussfaktor auf die erhobenen Lesezeiten im Experiment aus ($p<0,001$). Die durchschnittlich schlechtesten Lesezeiten erzielten dabei Teilnehmer, wenn sie mit der Zeilenbreite „0“ konfrontiert waren – dies entsprach Zeilen mit der Länge von im Schnitt 28 Anschlägen. Ein zunächst nachvollziehbares Ergebnis, werden Zeilenbreiten von unter 30 Anschlägen doch auch in der Literatur weithin als zu schmal klassifiziert.

Dass allerdings die eher sanfte Verbreiterung der präsentierten Zeilen um gerade einmal 14 Anschläge gleich die eindeutig besten Resultate aller gemessenen Zeilenbreiten zeitigen würde, war sicherlich nicht zu erwarten. Tatsächlich beschleunigte sich die Lesezeit im Übergang von Breite „0“ (28 Anschläge im Schnitt) zu Breite „1“ (42 Anschläge) um 6,0 Prozent, an Flachbildschirmen sogar um 6,4 Prozent. Wie Tabelle 6.9 verdeutlicht, fielen – über alle Monitorarten hinweg betrachtet – die Lese-Ergebnisse für die „42er-Zeilen“ statistisch signifikant besser aus als die Resultate für die schmalste und die längste gemessene Zeilenbreite – sowie merklich, aber gerade eben nicht signifikant ($p=0,056$) besser als für die Zeilen mit durchschnittlich 70 Anschlägen.

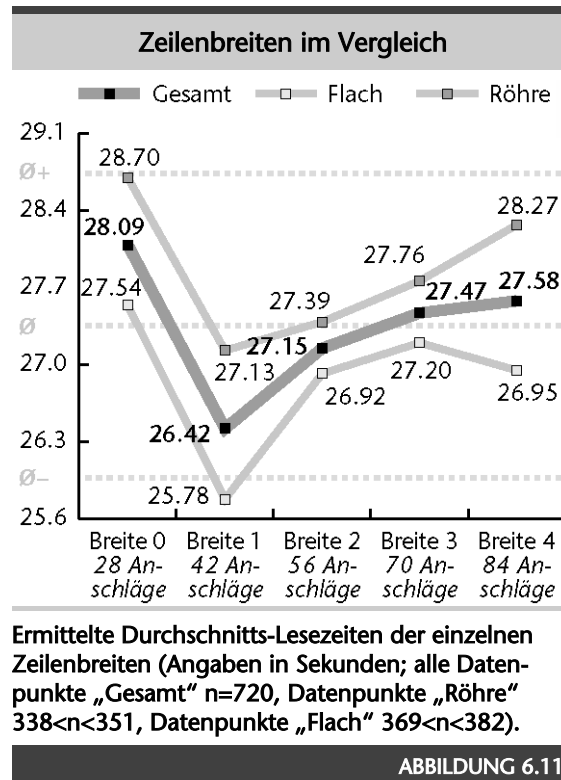


Tabelle 6.9: Paarweiser Vergleich der Zeilenbreiten										
	VB –	Ø	VB +	n	S	p (28 A.)	p (42 A.)	p (56 A.)	p (70 A.)	p (84 A.)
28 A.	27,03	28,09	29,16	720	7,31	–	0,000	0,114	0,685	0,880
42 A.	25,38	26,42	27,46	720	7,13	0,000	–	0,389	0,056	0,027
56 A.	26,10	27,15	28,20	720	6,88	0,114	0,389	–	0,993	0,949
70 A.	26,38	27,47	28,57	720	7,32	0,685	0,056	0,993	–	1,000
84 A.	26,48	27,58	28,68	720	7,53	0,880	0,027	0,949	1,000	–

Quelle: eigene Erhebung

Eine weitere Verbreiterung der Zeilenlängen über die angesprochenen 42 Anschläge hinaus führte mithin zu keiner weiteren Beschleunigung der Lesezeiten. Sowohl die in Breite „2“ (56 Anschläge) präsentierten als auch die in Breite „3“ (70) und Breite „4“ (84) dargebotenen Texte schnitten im Schnitt schlechter ab als Breite „1“. Allerdings fielen die durchschnittlichen Lesezeiten auch in diesen größeren Zeilenbreiten durch die Bank immer noch besser als die der Breite „0“.

6.7.1.2 Ergebnisse am Flachbildschirm

Auch für den Faktor Zeilenbreite lohnt ein isolierter Blick auf die Messergebnisse, die von Probanden erzielt wurden, die das Experiment an einem Flachbildschirm durchführten. Zwar ergab die Varianzanalyse keine signifikante Interaktion der Faktoren „Zeilenbreite“ und „Bildschirmart“ ($p=0,670$), was zunächst kaum verwundert: Im Wesentlichen ergibt sich schließlich sowohl für Flach- als auch für Röhrenbildschirme ein vergleichbares Bild wie in der Gesamtschau über alle Monitorgattungen hinweg. Im Detail ist jedoch bemerkenswert, dass sich die exponierte Position der Breite „1“ mit ihren 42 Anschlägen im Vergleich zu den drei größeren gemessenen Breiten an Flachbildschirmen noch etwas ausgeprägter darstellt als unter Einbeziehung aller Monitorarten (siehe Tabelle 6.10). So fiel beispielsweise die Lesegeschwindigkeit zwischen Breite „1“ und Breite „2“ über alle Bildschirmarten hinweg um 2,7 Prozent ab (26,42 gegenüber 27,15 Sekunden), an Flachbildschirmen jedoch um 4,2 Prozent (25,78 gegenüber 26,92 Sekunden).

Tabelle 6.10: Paarweiser Vergleich der Zeilenbreiten (Flachbildschirme)										
	VB –	Ø	VB +	n	S	p (28 A.)	p (42 A.)	p (56 A.)	p (70 A.)	p (84 A.)
28 A.	26,82	27,54	28,26	382	7,03	–	0,006	0,914	0,999	0,944
42 A.	25,06	25,78	26,50	379	7,00	0,006	–	0,228	0,071	0,208
56 A.	26,20	26,92	27,64	369	6,89	0,914	0,228	–	1,000	1,000
70 A.	26,42	27,20	27,98	372	7,48	0,999	0,071	1,000	–	1,000
84 A.	26,21	26,95	27,69	379	7,17	0,944	0,208	1,000	1,000	–

Quelle: eigene Erhebung

Insgesamt ergibt sich bei isolierter Betrachtung der Flachbildschirm-Ergebnisse eine Art „Zwei-Klassen-Gesellschaft“: Die Breite „1“ überragt alle übrigen vier gemessenen Breiten eindeutig. Innerhalb dieser übrigen vier Breiten sind die Unterschiede dagegen bemerkenswert marginal.

Im Wesentlichen ähnliche Resultate, wenn auch in den Ausprägungen leicht abgeschwächt im Abgleich mit denen, die an Flachbildschirmen erzielt wurden, fanden sich für die Resultate an Röhrenbildschirmen. Eine tabellarische Darstellung dieser monitorspezifischen Teilergebnisse erscheint mir daher an dieser Stelle verzichtbar.

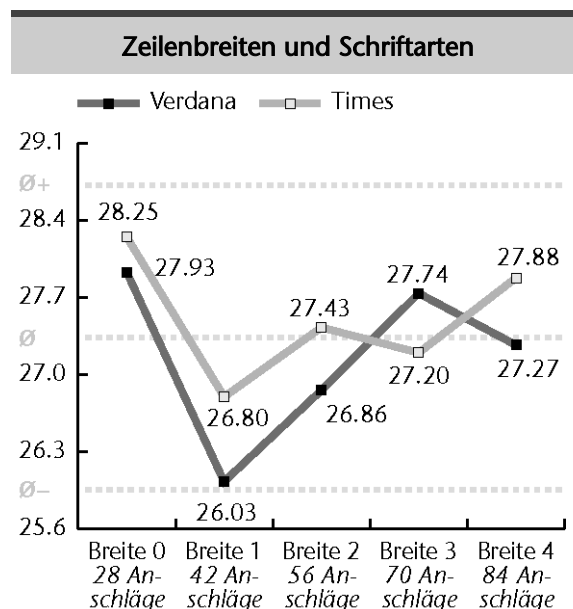
6.7.2 Zwischenfazit: Zeilenbreiten

Die Ausgestaltung der Zeilenbreiten auf Webseiten ist offenbar von wirklich zentraler Bedeutung für die Lesbarkeit von Texten. Es hat dabei den Anschein, als entfalte sich die bestmögliche Lesbarkeit in einem recht engen Werte-Korridor. Eine optimale Lesbarkeit entfalten wahrscheinlich Zeilen, die zwischen etwa 40 und 60 Zeichen Platz bieten; dabei scheint eher die untere Kante dieses Korridors empfehlenswert als die obere. Von sehr schmal bemessenen Zeilen (weniger als 40 Anschläge) ist augenscheinlich unbedingt und unabhängig von der Monitorart abzuraten.

Entscheidet man sich dagegen für Zeilenbreiten von mehr als 50 Anschlägen, so verschlechtert sich die Lesbarkeit an Flachbildschirmen sofort merklich im Vergleich zum empfohlenen Korridor, fällt aber bei immer weiterer Verbreiterung auf bis zu 84 Anschläge kaum noch ab. An Röhrenmonitoren dagegen verschlechtert sich die Lesbarkeit mit wachsender Zeilenbreite sukzessive.

6.7.3 Zeilenbreite und Schriftart im Zusammenwirken

Schlüsselt man die Ergebnisse der fünf verglichenen Zeilenbreiten nach dem Kriterium „Schriftart“ auf, so zeigt sich, dass der „Sprung“ zwischen Breite „0“ und Breite „1“ unabhängig von der Schriftart zu beobachten ist (und im paarweisen Vergleich für beide Schriftarten statistisch signifikant ausfiel) – allerdings ist er bei



Ermittelte Durchschnitts-Lesezeiten der 5 Zeilenbreiten, aufgeschlüsselt nach Schriftarten (in Sekunden; jeder Datenpunkt n=360)

ABBILDUNG 6.12

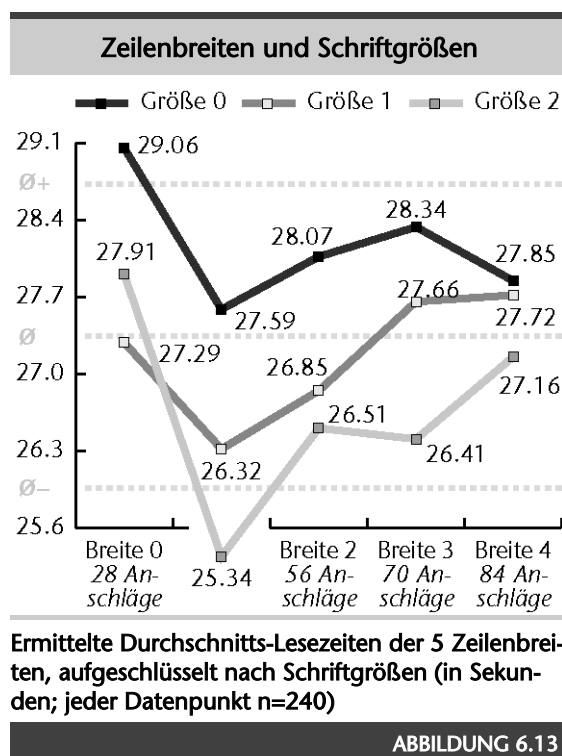
der *Verdana* mit 7,3 % Differenz ausgeprägter als bei der *Times* (5,4 %). Bei beiden Schriftarten steigen die Lesezeiten mit vergrößernder Zeilenbreite wieder an, ohne den Wert der Breite „0“ wieder zu erreichen. Die *Verdana* in Breite „1“ wurde dabei signifikant schneller gelesen als die *Times* in Größe „0“ und die *Times* in Größe „4“. Weitere Signifikanzen fanden sich nicht.

Insgesamt ist festzustellen, dass die Varianzanalyse der erhobenen Daten eindeutig keine signifikante Interaktion der Einflussfaktoren „Zeilenbreite“ und „Schriftart“ ausweist ($p=0,276$). Auch das Zusammenspiel der Faktoren „Schriftart-Zeilenbreite-Bildschirmart“ übte keinen signifikanten Einfluss auf die erhobenen Lesezeiten aus ($p=0,787$). Aus diesem Grunde kann an dieser Stelle auch auf eine Aufschlüsselung nach Bildschirmarten verzichtet werden.

6.7.4 Zeilenbreite und Schriftgröße im Zusammenwirken

Die Varianzanalyse ergab keine signifikante Interaktion der Faktoren „Zeilenbreite“ und „Schriftgröße“ ($p=0,276$). Die grafische Aufschlüsselung der Zeilenbreiten nach dem Kriterium „Schriftgröße“ (siehe Abbildung 6.13) verschafft einen Eindruck, warum dies so ist: Zwar wird wiederum deutlich, dass die Zeilenbreiten (isoliert als Haupteinflussfaktor auf die Lesezeiten erkannt) merklich wirken und auch die drei gemessenen Schriftgrößen (ebenfalls für sich betrachtet Haupteinflussfaktoren) sich eindeutig wechselseitig abgrenzen lassen in ihren Messwerten – aber beide Faktoren bedingen sich eben gegenseitig kaum.

Jenseits der Varianzanalyse fanden sich dennoch bemerkenswerte Resultate (was sich bereits in der außergewöhnlichen „Brechung“ der Abszisse in Abbildung 6.13 manifestiert): beispielsweise, dass Schriften, die mit einer x-Höhe von 8 Pixeln (Größe „2“) ausgestattet und in Zeilen mit durchschnittlich 42 Anschlägen eingerichtet sind, um stattliche 12,8 Prozent schneller gelesen wurden im Experiment als Schriften mit 6 Pixeln Mittellänge in einer Zeilenbreite von im Schnitt 28 Zeichen.



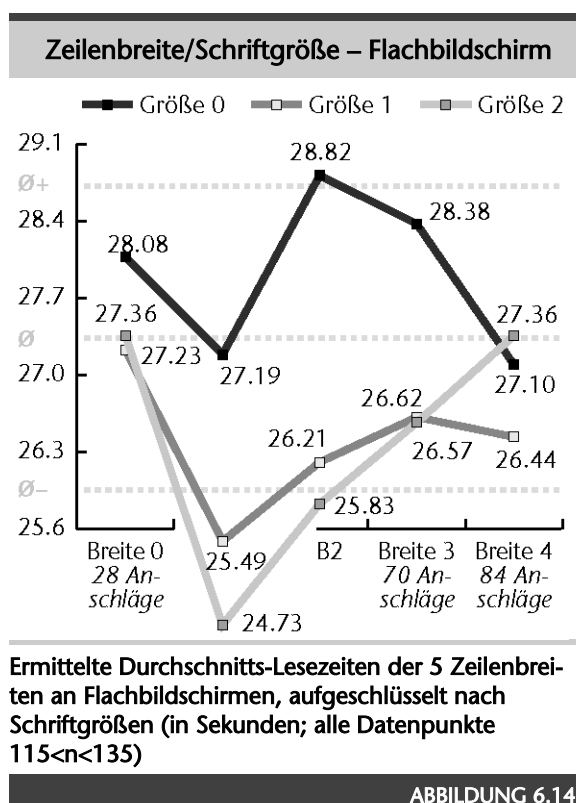
Insgesamt erwies sich im paarweisen Abgleich, dass Schriftgröße „2“ in der Breite „1“ signifikant schneller gelesen wurde als alle Konstellationen in der Schriftgröße „0“ sowie ebenfalls signifikant schneller als Schriften, die in der Schriftgröße „1“ gesetzt waren und mehr als 56 Anschlägen Platz boten. Schriftgröße „0“ in der Breite „0“ wurde dagegen signifikant langsamer gelesen als Schriftgröße „2“ in den Zeilenbreiten „1“, „2“ und „3“, außerdem langsamer als Größe „1“ in der Zeilenbreite „1“.

Ein genauere Blick auf die Varianzanalyse offenbart aber auch, dass die Kombination „Zeilenbreite – Schriftgröße“ augenscheinlich unterschiedliche Ergebnisse zeitigt, betrachtet man die Ergebnisse unter Einbeziehung des Faktors „Bildschirmart“. Für diese Dreier-Interaktion wurde unter den Bedingungen des reduzierten Signifikanzniveaus ganz knapp eine Bedeutsamkeit nicht nachgewiesen ($p=0,032$). Betrachtungswürdig erscheint sie mir dennoch.

6.7.4.1 Am Flachbildschirm

Die detaillierte Betrachtung des Zusammenwirkens von Zeilenbreite und Schriftgröße an Flachbildschirmen offenbart eine augenscheinliche „Zwei-Klassen-Gesellschaft“ unter den Messwerten. So entwickelten sich die Lesezeiten für die gemessenen Größen „1“ und „2“ nicht nur weitgehend synchron, sondern auch sehr niveaunählich. Die Lesezeiten für Schriften in der Größe „0“ dagegen fielen gerade in den drei mittleren gemessenen Zeilenbreiten um mindestens 6,4 und bis zu 11,3 Prozent schlechter aus als die Werte für die beiden größeren Schriftgrößen in der jeweils selben Zeilenbreite. Zwischen den Messwerten in Schriftgröße „2“ in Zeilenbreite „1“ einerseits und den Ergebnissen für Schriftgröße „0“ in Zeilenbreite „2“ lagen sogar stattliche 14,2 Prozent.

Im paarweisen Abgleich nach „Dunnett T3“ erwiesen sich die Schriften in Größe „2“, gesetzt in Breite „1“ (42 Anschläge) als signifikant schneller lesbar als die Schriften in Größe „1“ in den Breiten „0“, „2“ und „3“. Ebenfalls signifikant besser fielen die Messwerte in Schriftgröße „1“ in Zeilenbreite „1“ aus im Vergleich zu Schriftgröße „0“ in Zeilenbreite „2“. Dass keine weiteren Signifikanzen gefunden wurden, dürfte damit zusammenhängen, dass in diesem



„Dreier“-Vergleich schlicht nicht genug Messwerte in jede der 15 Einzelkonstellationen einfließen konnten.

Letztlich verstärken diese Ergebnisse jedoch insbesondere nochmals den Eindruck, dass von Schriften mit Mittellängen von 6 und weniger Pixeln dringend abzuraten ist im Browser-Schriftsatz.

6.7.4.2 Am Röhrenbildschirm

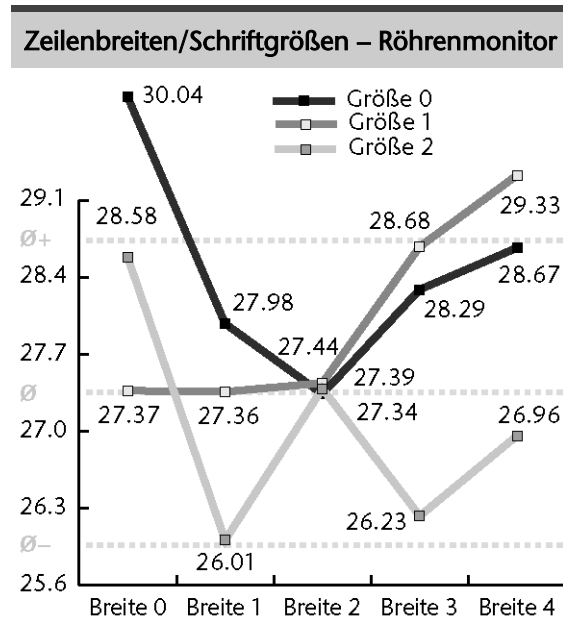
Ein zunächst sehr verwirrendes Bild ergibt die grafische Aufbereitung des Zusammenspiels von Zeilenbreite und Schriftgröße an Röhrenbildschirmen (Abbildung 6.15).

Es fanden sich lediglich zwei Signifikanzen im paarweisen Abgleich nach „Dunnnett T3“ (auch hier wohl bedingt durch die notwendigerweise eher kleinen Stichproben): Texte in den Schriftgrößen „2“ wurden in den Zeilenbreite „1“ und „3“ bedeutsam schneller gelesen als solche in Schriftgröße „0“ und Zeilenbreite „0“.

In der Summe kann gefolgert werden, dass zum einen auch die Auswertung für die Röhrenbildschirm-Resultate die Verwendung höherer Schriftgrößen nahe legt sowie andererseits wiederum die Beschränkung der Zeilenbreiten auf nicht mehr als 56 Anschläge.

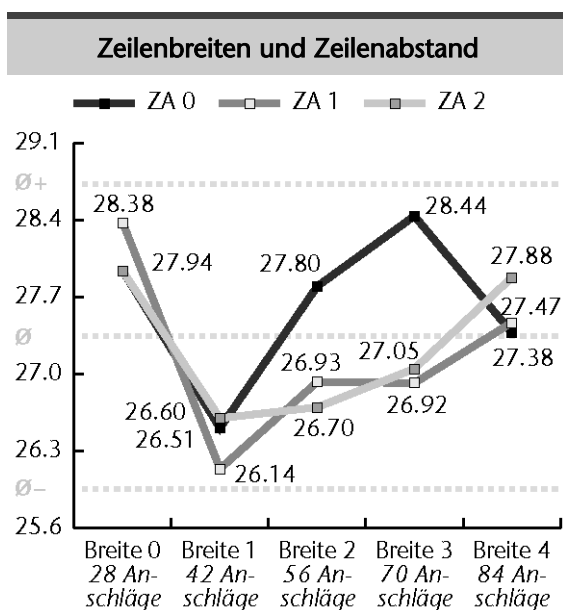
6.7.5 Zeilenbreite und Zeilenabstand im Zusammenwirken

Es ist eine der häufigsten Forderungen, dass mit der Zeilenbreite der Zeilenabstand zu vergrößern sei. Die Varianzanalyse der erhobenen Daten dagegen er-



Ermittelte Durchschnitts-Lesezeiten der 5 Zeilenbreiten an Röhrenbildschirmen, aufgeschlüsselt nach Schriftgrößen (in Sekunden; alle Datenpunkte 107<n<121).

ABBILDUNG 6.15



Ermittelte Durchschnitts-Lesezeiten der 5 Zeilenbreiten, aufgeschlüsselt nach Zeilenabständen (in Sekunden; alle Datenpunkte n=240).

ABBILDUNG 6.16

gab keine statistisch signifikante Interaktion zwischen den Faktoren „Zeilenbreite“ und „Zeilenabstand“ ($p=0,276$).

Auch in der Detailbetrachtung bestätigten sich die Forderung nach gezielter Koppelung von Zeilenbreite und -abstand nur sehr bedingt als stichhaltig: Wie Abbildung 6.16 verdeutlicht, fielen die Ergebnisse für die beiden schmaleren Zeilenbreiten unabhängig vom betrachteten Zeilenabstand sehr ähnlich aus.

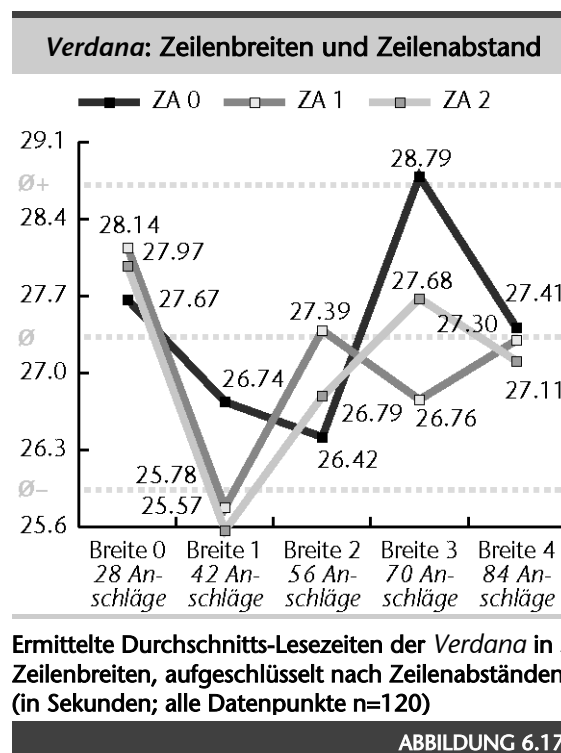
Bei den Breiten „2“ und „3“ dagegen fielen die Ergebnisse für den kleinsten Zeilenabstand „0“ sichtbar schlechter aus als für die beiden größeren Zeilenabstände, die auch in den größeren Zeilenbreiten quasi im „Gleichschritt“ vorankommen. Bei Zeilenbreite „4“, der größten, haben sich allerdings alle drei Zeilenabstands-Werte wieder stark angenähert. Insgesamt fanden sich im paarweisen Vergleich keine Signifikanzen unter den 15 verglichenen Werten.

6.7.6 Zeilenbreite, Zeilenabstand und Schriftart im Zusammenwirken

Eine „Dreier-Konstellation“ übte laut Varianzanalyse zwar keinen signifikanten, aber doch einen immer noch so bemerkenswerten Einfluss auf die gemessenen Zeiten aus ($p=0,057$), dass sie hier näher untersucht werden soll: das Zusammenwirken nämlich der Faktoren Zeilenbreite, Zeilenabstand und Schriftart. Um die grafischen Aufbereitungen dieser Detailbetrachtung nicht allzu verwirrend geraten zu lassen, wird dieses besondere typografische Zusammenspiel im Folgenden getrennt unter den Bedingungen der beiden verglichenen Schriftarten *Verdana* und *Times* aufbereitet und diskutiert.

6.7.6.1 Verdana

Auffällig ist zunächst, dass – wenn die *Verdana* als Schriftart angeboten wurde – die Beschleunigung der Lesezeiten im Übergang von Zeilenbreite „0“ zu Zeilenbreite „1“ unter den Bedingungen der beiden größeren gemessenen Zeilenabstände sehr merklich ausfielen (jeweils mehr als 8 Prozent), um in den drei breiteren Zeilenlängen wieder um mindestens 5 Prozent, eher noch stärker abzufallen – in Bezug auf die Zeilenbreiten ein inzwischen nicht mehr sehr überraschendes Ergebnis. Unter den Bedingungen des schmalsten Zeilenab-

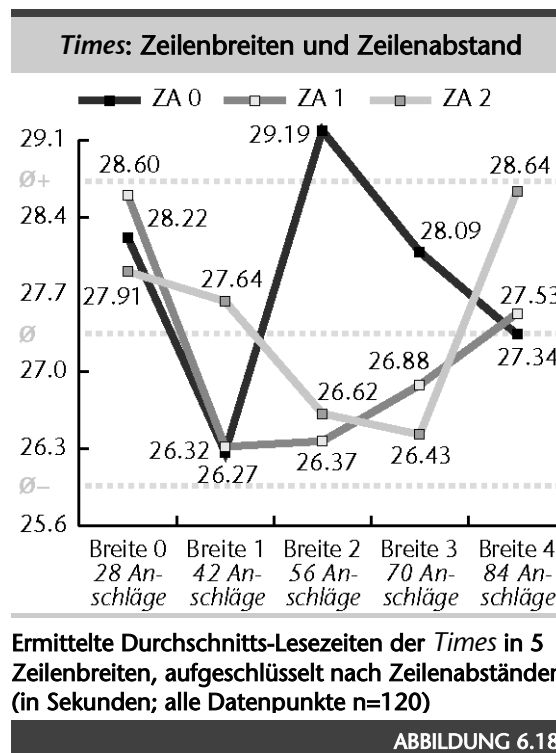


stands allerdings wurde die *Verdana* in breiter bemessenen Zeilen (56 Anschläge im Schnitt) am raschesten rezipiert, um bei einer weiteren Zeilenverbreiterung den schlechtesten *Verdana*-Wert überhaupt auszuweisen.

Signifikanzen wurden in dieser Konstellation (*Verdana*) nicht gefunden. In die Nähe einer Signifikanz geriet lediglich der Abgleich zwischen Breite „1“ in Zeilenabstand „2“ und Breite „3“ in Zeilenabstand „0“ ($p=0,080$) – aber eben nur in die Nähe.

6.7.6.2 Times

Recht überraschend fällt die Analyse des Zusammenspiels der typografischen Faktoren „Zeilenbreite“ und „Zeilenabstand“ unter den Bedingungen der vorgelegten Schriftart *Times* aus. Unter den Bedingungen des größten gemessenen Zeilenabstands nämlich erzielte diese Schriftart das beste durchschnittliche Leseergebnis in der sehr großen Zeilenlänge „3“ (70 Anschläge). Unter den Bedingungen der beiden schmalen Zeilenbreiten dagegen ergab sich das „gewohnte“ Bild: Die Setzung von Zeilenbreite „1“ zeitigte hier die besten Leseergebnisse.

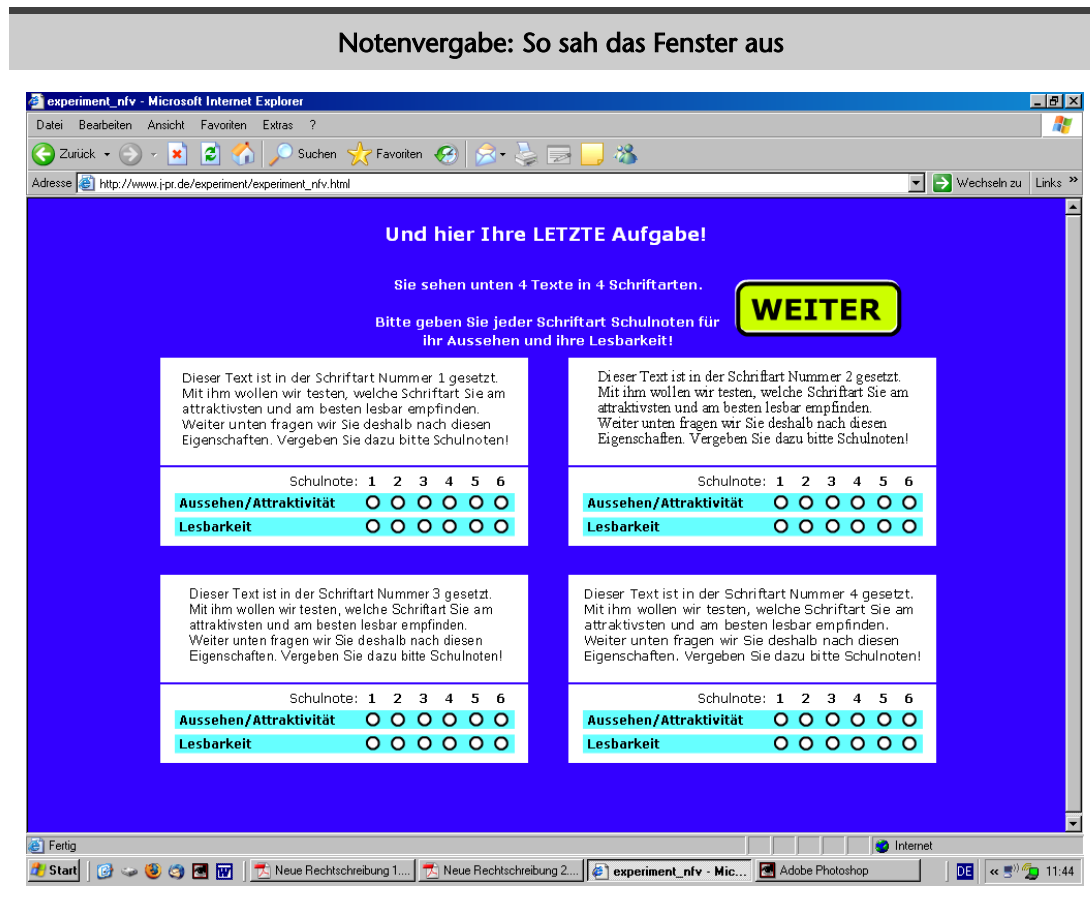


Signifikanzen wurden in dieser Konstellation nicht gefunden; angesichts von auch hier „nur“ 120 Messresultaten pro Detailwert überrascht dies nur bedingt.

6.7.6.3 Gesamtfazit: Zeilenbreiten

Es scheint etwas zu passieren bei der Verbreiterung von Zeilen im Internet von durchschnittlich 28 auf 42 Anschläge pro Zeile. Im Schnitt lasen jene Versuchsteilnehmer meiner Studie, die die breitere der beiden Zeilen vorgelegt bekamen, diese um 6,3 % schneller als diejenigen, die mit den schmalen Zeilen konfrontiert waren. Dieser Effekt war in leicht variierenden Ausprägungen über alle Texte zu beobachten, unabhängig von verwendeter Schriftart, -größe oder Zeilenabstand. Insbesondere ein Zusammenhang zwischen Zeilenbreite und Zeilenabstands-Bemessung, in der Literatur gelegentlich formuliert und unterstellt, scheint kaum zu existieren.

6.8 Die Schrift-Benotung



Screenshot der Abschlussaufgabe im Experiment.

ABBILDUNG 6.19

Im Anschluss an die vier (beziehungsweise, inklusive Übungstext, fünf) Textaufgaben wurden die Probanden im Experiment gebeten, sich vier Textblöcke anzusehen; diese waren in vergleichbaren Schriftgrößen gesetzt, jedoch in unterschiedlichen Schriftarten. Präsentiert wurden die im Web wohl am häufigsten anzutreffenden Schriftarten *Arial*, *Times* und *Verdana*. Als vierte Schriftart legte ich die *Frutiger* vor, und zwar aus folgendem Grund: Die inzwischen für den Jahresbeginn 2007 angekündigte neue *Windows*-Version von *Microsoft* wird gerüch-teweise (wie bereits dargestellt) mit einer neuen Standardschriftart ausgeliefert werden: der *Segoe*. Augenscheinlich hat *Microsoft* mit dieser Schrift allerdings – wie seinerzeit mit der *Ari-al* – alles andere als eine echte Neuschöpfung zu bieten, sondern lediglich eine Kopie eben der *Frutiger*⁹¹⁸, sicherlich einer der schönsten und stimmigsten Schöpfungen der zweiten Hälfte des 20. Jahrhunderts⁹¹⁹. Da mir selbst diese „neue“ Schriftart *Segoe* nicht vorlag, prä-sentiertere ich die *Frutiger*.

⁹¹⁸ Vgl. Klein 2005, S. 13

⁹¹⁹ Vgl. Friedl 1998, S. 40f.

Die Probanden wurden aufgefordert, für jeden Textblock zwei Schulnoten zu vergeben: Eine für „Aussehen/Attraktivität“ der jeweiligen Schrift, eine weitere für die von ihnen vermutete „Lesbarkeit“ des Fonts. Insgesamt wurden also acht Noten erbeten.

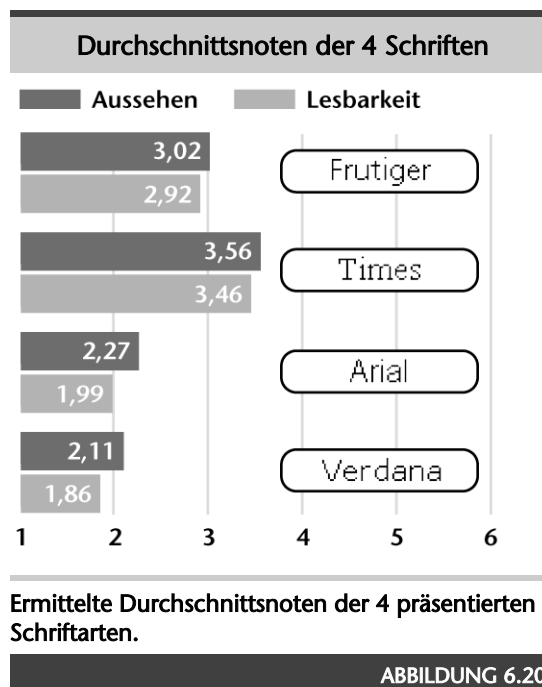
6.8.1 Bemerkung: Die Schulnote als ordinale Skala

Zur Grafik rechts ist folgende Anmerkung unumgänglich: Sie basiert auf einer statistisch eigentlich nicht haltbaren Berechnung.

Denn das deutsche System der Schulnoten von „1“ (sehr gut) bis „6“ (ungenügend) ist, wie wohl zahlenbasiert, kein nominal, sondern ordinal aufgebautes Konstrukt – und schon gar nicht auf regelmäßigen Intervallen aufbauend. Eine „3“ ist in diesem System eben nicht „halb so gut“ wie eine „6“, obgleich die Ziffern dies nahelegen. Auch gleichsam umgekehrt aufgepäuselt entzieht sich das System einer Rechnung: Eine „1“ ist per definitionem nicht „drei Mal besser“ als eine „3“. Vielmehr gilt als Standard, dass eine „1“ einem Ergebnis nahe der Höchstpunktzahl entspricht (in einer Mathematik-Klausur also steht eine „1“ einer Arbeit zu, in der der Prüfling alle Aufgaben korrekt gelöst hat), eine „4“ dagegen der halben erreichbaren Punktzahl. Daher ist eigentlich auch die Berechnung eines arithmetischen Durchschnittswertes aus Schulnoten nicht korrekt.

Dass im hier beschriebenen Experiment dennoch Schulnoten erbeten wurden, muss im Nachhinein als Schwäche des Untersuchungsdesigns eingestanden werden. Wenn hier dennoch zunächst eine Grafik mit Durchschnittsnoten gezeigt wird, so deswegen, weil sich die Tendenzen in selbiger sehr anschaulich manifestieren.

Die Ergebnisse sind durchaus auffällig. So bildeten *Arial* und *Verdana* das Spitzen-duo, wobei die Lesbarkeiten der beiden Schriften jeweils etwas positiver bewertet wurden als deren „Attraktivität“. Weit abgeschlagen rangiert die *Times*, nur knapp davor platzierte sich die *Frutiger*.



6.8.2 Paarweiser Abgleich der Noten

Nun interessierte natürlich insbesondere, ob und inwiefern sich die beobachteten Benotungsdifferenzen als statistisch signifikant erweisen würden. Um dies zu überprüfen, wandelte ich die von den Teilnehmenden vergebenen Schulnoten in Nominalwerte einer improvisierten gleichintervalligen Skala um. Dabei übersetzte ich die Schulnote „3“ – als von mir vermuteter Ausdruck einer Bewertung im Sinne von „in Ordnung“ oder „weder gut noch schlecht“ – in den Skalenwert „0“, die Schulnote „4“ in den Skalenwert „-1“, die Note „5“ in den Skalenwert „-2“ sowie die Note „6“ in den Skalenwert „-3“. Um die improvisierte Skala insgesamt symmetrisch in der Extremausprägung zu gestalten, ordnete ich der Schulnote „2“ den Skalenwert „1,5“ zu und der Schulnote „1“ den Wert „3“.

Auf Basis dieser improvisierten Skala war nun auch ein paarweiser Abgleich nach „Dunnett T3“ möglich. Dabei zeigte sich, dass die *Times* signifikant „unattraktiver“ und auch als signifikant schlechter lesbar empfunden wurde im Direktvergleich mit den übrigen drei Schriftarten. Die *Frutiger* wiederum wurde signifikant schlechter bewertet in Sachen „Attraktivität“ und „Lesbarkeit“ als *Verdana* und *Arial*. Zwischen *Verdana* und *Arial* stellte ich keinen signifikanten Unterschied fest in der Bewertung der „Attraktivität“ ($p=0,332$). Allerdings wurde die *Verdana* signifikant besser benotet in ihrer „Lesbarkeit“ als die *Arial* ($p<0,05$).

Diese Ergebnisse sind vor allem in der Hinsicht beachtenswert, als im Lesegeschwindigkeitstest ja gerade festgestellt worden war, dass sich *Times* und *Verdana* in ihren Lesbarkeitsqualitäten die Waage halten, sofern die erzielbare Lesegeschwindigkeit zum Maßstab genommen wird. Nimmt man nun jedoch die „Sympathiewerte“ als weiteren Einflussfaktor hinzu, so muss festgestellt werden: Die *Times* mag gut lesbar sein am Bildschirm, zumindest nicht schlechter als die *Verdana*. In dieser Qualität wird sie subjektiv allerdings mehrheitlich nicht erlebt; beliebt ist sie nicht.

6.8.3 Attraktivität = Lesbarkeit?

Weiterhin ist die Beobachtung interessant, dass die Durchschnittnoten für die Attraktivität einer jeden der 4 präsentierten Schriftarten nahezu identisch ausfielen wie die Noten für die „Lesbarkeit“ der jeweiligen Schrift.

Dass diese Übereinstimmungen in der Bewertung von Aussehen und Lesbarkeit ein und derselben Schriftart nicht Zufall statistischer Verteilungen sind, zeigt Abbildung 6.21. Hier ist, nach Schriftarten aufgeschlüsselt, die jeweilige Anzahl der Teilnehmenden abgebildet, die dem Aussehen einer Schriftart eine bessere, dieselbe oder eine schlechtere Note gaben als deren vermuteter Lesbarkeit. Mindestens die Hälfte aller Teilnehmer vergab dabei für ein und dieselbe Schriftart eine identische Note für die Kriterien „Attraktivität/Aussehen“ und „Les-

barkeit“; im Falle der *Verdana* benoteten sogar knapp 62 Prozent aller Teilnehmenden beide Kriterien gleich.

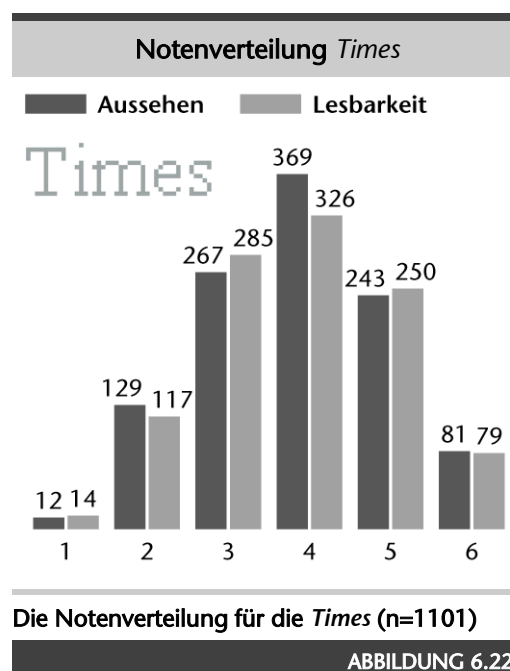
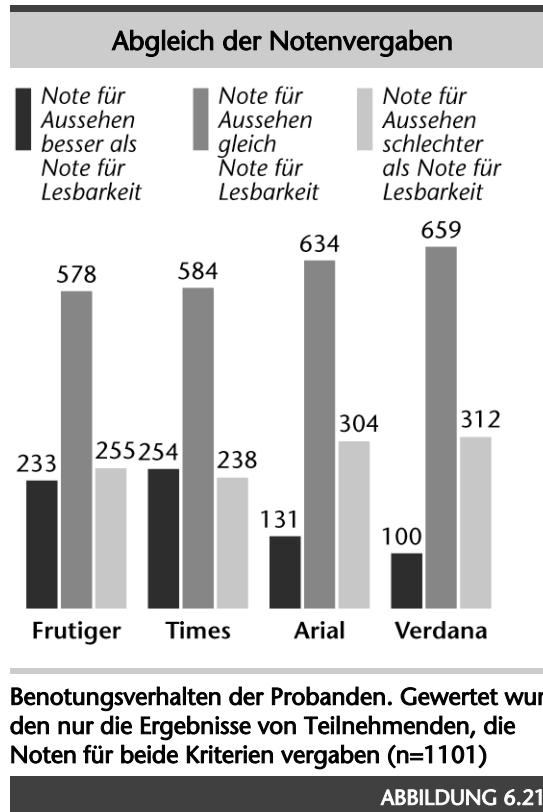
Über die Gründe dieser Übereinstimmungen kann notwendigerweise nur spekuliert werden. Eine mögliche Erklärung wäre, dass die Probanden zu einem hohen Anteil einen Kausalzusammenhang konstruierten zwischen der Lesbarkeit und der Attraktivität einer Schriftart. Ob dabei freilich im Urteil der Teilnehmenden die Güte der vermuteten Lesbarkeit die Attraktivitätsnote bedingte oder umgekehrt, muss ungeklärt bleiben.

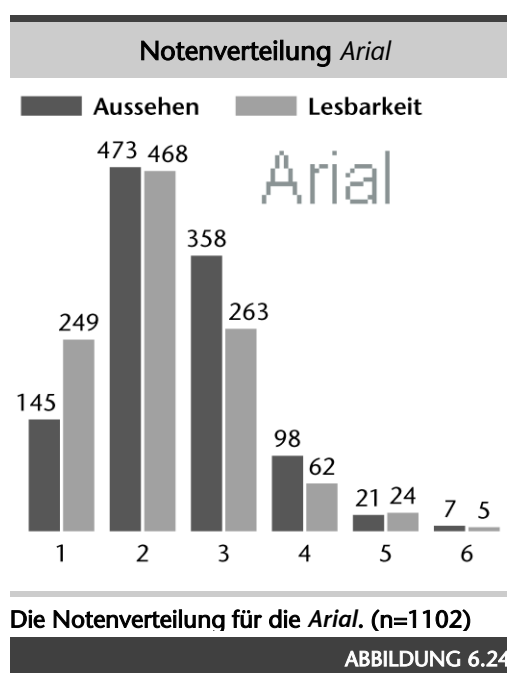
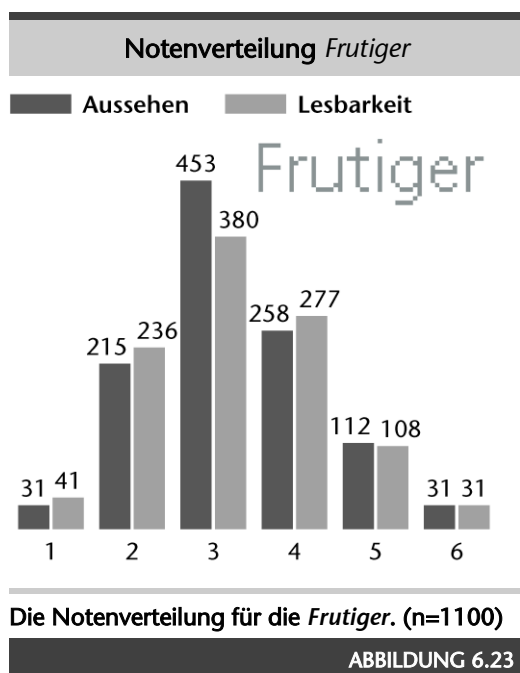
Für mögliche Folgestudien böte sich daher an, die Teilnehmenden ausdrücklich um die Einschätzung zu bitten,

- ob für die Beurteilung einer Schriftart eher das Aussehen das Hauptkriterium sein sollte oder ihre Lesbarkeit (oder beides gleichgewichtig)
- ob Lesbarkeit und Attraktivität einer Schriftart miteinander zusammenhängen beziehungsweise: ob eines der Kriterien die Güte des anderen determiniert.

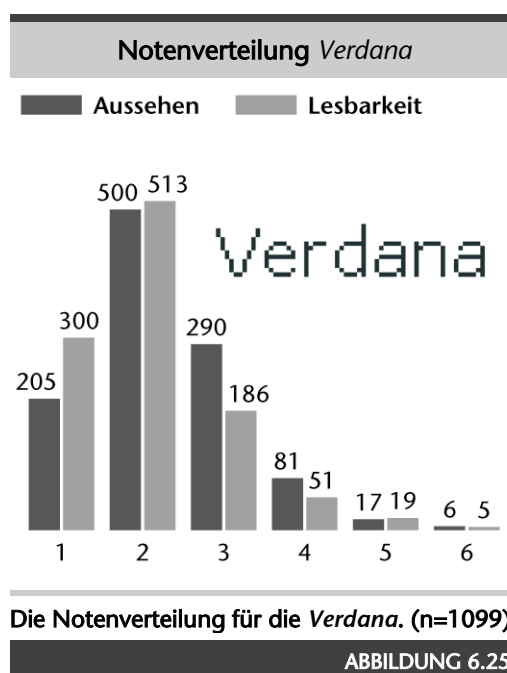
6.8.4 Die Noten im Detail

Die *Times*, die, wie gezeigt, in den Lesbarkeitstests keineswegs schlechtere Werte erzielte als die Konkurrentin *Verdana*, schnitt im Urteil der Probanden am eindeutig schlechtesten ab. Die Durchschnittsnote von 3,56 für „Aussehen/Attraktivität“ (Median = 4, Modus = 4) fiel für die *Times* merklich schlechter aus als bei der *Frutiger* (Durchschnitt = 3,02, Median = 3, Modus = 3) und deutlich schlechter als bei *Arial* (Durchschnitt = 2,27, Median = 2, Modus = 2) und *Verdana* (Durchschnitt = 2,11, Median = 2, Modus = 2).





Insgesamt dürfen die Schriftarten *Arial* und *Verdana* nach diesen Ergebnissen als die klaren „Publikumsfavoriten“ gelten, deutlich vor der *Frutiger*, die, wie erläutert, offenbar Pate gestanden hat für die vermeintlich neue *Microsoft*-Schriftart *Segoe*, und eindeutig vor der *Times New Roman*. Nicht nur, dass die beiden bestbenoteten Schriftarten in Sachen „Attraktivität/Aussehen“ klar vorne lagen – sie werden von der Leserschaft auch eindeutig als „besser lesbar“ eingestuft als *Times* und *Frutiger*.



6.8.5 Die These von der „gefühlten Lesbarkeit“

Dass diese Eindeutigkeit im Urteil – vor allem die Geringschätzung der *Times* nicht nur, aber auch in Sachen „Lesbarkeit“ – sich nicht deckt mit den Ergebnissen nicht nur meiner Lesbarkeits-Untersuchungen, sollte nicht als dem Laienurteil geschuldet und daher „wissenschaftlich unhaltbar“ abgetan werden; vielmehr könnte man hier durchaus, analog dem Wetterbericht, von einer „gefühlten“ Lesbarkeit sprechen, vom individuellen Eindruck der Lesenden, wie leicht und angenehm rezipierbar der aktuell dargebotene Text sei. Und wenn auch diese „gefühlte Lesbarkeit“ nicht unmittelbare Wirkung entfaltet auf die Lese- und Lesegeschwindigkeiten, so ist zumindest die Vermutung nicht abwegig, dass sie durchaus Einfluss besitzt auf die Rezep-

tionsbereitschaft von Lesenden; darauf also, ob Nutzer bereit sind, vom Betrachter zum Leser zu werden.

Diese These von der „gefühlten“ Lesbarkeit unterstellt freilich, dass dem Leseprozess zumeist ein Akt der beurteilenden Abschätzung eines vorliegenden Textblocks nach optisch-ästhetischen Motiven vorausgeht. Die These ist mithin zu verstehen als adaptierte Anleihe aus der Bildbetrachtungsforschung. Diese geht davon aus, dass informationshaltige Bildkonstrukte von Betrachtern in einer sogenannten „prä-attentiven“ Phase zunächst einer allgemeinen Gesamteinschätzung unterworfen werden. Fällt diese Voreinschätzung positiv aus, so beginnt, im sich anschließenden „attentiven Prozess“, die eingehendere Beschäftigung mit den Bildinhalten als solchen⁹²⁰.

Interpretiert man nun einen noch nicht gelesenen Textblock als visuelles Gesamt, als ein gar bildähnliches Konstrukt, so liegt zumindest die Vermutung nahe, dass auch hier ein „prä-attentiver“ und ein „attentiver Prozess“ stattfinden könnten. Wenn also, wie die Ergebnisse nahe legen, ein Gros der Teilnehmenden die *Times* als deutlich weniger leserlich empfand als die übrigen drei, so könnte dies zumindest dazu führen, dass die Bereitschaft, mit dem Leseprozess zu beginnen, bei der *Times* weniger ausgeprägt ist als bei den übrigen.

6.8.6 Die These von der Gewohnheit

Eine weitere Annahme könnte Anhaltspunkte liefern für das überlegene Abschneiden der Schriftarten *Arial* und *Verdana* im Benotungsteil: die Vermutung nämlich, dass Lesende gewohnte Schriftbilder neuartigen allgemein vorziehen. In der Disziplin des Printdesigns hat diese Theorie zahlreiche Befürworter⁹²¹. Bergner etwa stellt fest, dass „ein gewohntes Schriftbild (...) in jedem Falle das rasche Aufnehmen des Inhaltes (erleichtert). Lesen ist an Konvention gebunden“⁹²².

Nun ist freilich zu fragen, ob ein gerade einmal zwölf Jahre altes Medium wie das Internet, das noch dazu in den paar Jahren seines Bestehens als Massenmedium eine große Anzahl an auch optischen Wandlungen durchlebt hat (schließlich erblickte die *Verdana* erst Ende der 90er Jahre überhaupt das Licht der Diode), bereits derlei Gewohnheitseffekte zu erzeugen in der Lage gewesen ist. Wenn man allerdings die Liste verwendeter Schriftarten in nachrichten-orientierten Angeboten zu Rate zieht (siehe Seite 104), so ist nochmals festzustellen, dass die *Verdana* das Web zurzeit ziemlich im Griff hält. Dass besagte *Verdana* in der vorliegend diskutierten Studie die besten „Publikums-Noten“ erhielt, kann zumindest als Indiz dafür gelten, dass gewohnte Schriftbilder auch im Web ihre Langzeitwirkung nicht verfehlen.

⁹²⁰ Vgl. Weidenmann 1994, S. 86

⁹²¹ Vgl. beispielsweise Garcia 1987; vgl. ferner Evans 1974.

⁹²² Vgl. Bergner 1990, S. 21

7 Zusammenfassung und Ausblick

Die spannendste Frage muss vorläufig unbeantwortet bleiben: Wird das versammelte Expertentum praktizierender Webdesigner von diesem Forschungsbericht Notiz nehmen? Oder fällt auch die vorliegende Arbeit jenem Schicksal anheim, das zwangsläufig und die meisten Anläufe zu ereilen scheint, in denen bisher versucht wurde, den Prozess der Rezeption von Typografie in Webbrowsern empirisch zu ergründen: Selbsternannte Fachleute ignorieren (bestenfalls!) die Existenz dieser Studien, suchen sich einen Verleger und erklären mit dessen Autorität, der Unangreifbarkeit einer schmalen New-Economy-Vita und einer Serienbestellnummer im Rücken zum Webdesign-Dogma, was sie selbst für richtig halten. Schreiben sie zudem in flottem Stil und ist der Einband dickwandig genug, ist die Chance groß, dass ihre persönliche Meinung Dogma in Lehrbetrieben und Hochschulen wird. Ob dieses Dogma empirisch gesichert ist, spielt spätestens dann keine Rolle mehr, wenn ausreichend Studierende es in Hausarbeiten nachgebetet und andere Autoren es abgeschrieben haben. Denn wenn etwas in allen Büchern steht, also „herrschende Meinung“ ist – wer benötigt dann noch überprüfende Forschung?

Die Hoffnung ist klein, dass die vorliegende Arbeit einem besseren Schicksal entgegensieht. So wird wohl noch in fünf, in zehn Jahren in vermeintlich brandaktuellen „Webdesign-Ratgebern“ zu lesen sein, dass im Web serifentragende Schriften schlechter lesbar seien als serifenlose – eine Behauptung, die nicht nur im hier vorgestellten, sondern auch bereits in einigen älteren angelsächsischen Experimenten widerlegt wurde. Es wird höchstwahrscheinlich auch in diesen noch zu verfassenden Werken behauptet werden, dass die sehr fein gezeichneten Striche von klassischen Schrifttypen wie der *Times* am grob aufgelösten Monitor zu plumpen Einheits-Stumpfen verstümmeln – was unbestreitbar ist –, und es wird die Deduktion folgen, dass eben aus diesem Grunde die Lesbarkeit dieser Schriften am Bildschirm reduziert sei – was als Kausalschluss offensichtlich falsch ist. Denn nicht nur die *Times* erwies sich in meinem Online-Experiment als kaum merklich schlechter, also langsamer lesbar als die Vergleichsschriftart *Verdana* – in verschiedenen Vorgänger-Studien schnitt auch die Serifen-Schriftart *Georgia* alles andere als übel ab.

Interessanterweise wird – so befürchte ich zumindest – in den erwähnten und noch zu verfassenden Webdesign-Werken (wie in den bereits vorliegenden) mehrheitlich kaum eine Rede davon sein, dass aus einem ganz anderen Grunde der Einsatz von serifentragenden Schriften wie eben der *Times*, aber mit Abstrichen auch der *Georgia* im Web zumindest überdenkenswert erscheint: sie sind offenbar nicht sonderlich beliebt in der Nutzerschaft, zumindest merklich unbeliebter als viele serifenlose Fonts. Sie werden offenbar als weniger attraktiv und gleichzeitig als eher schlechter lesbar empfunden als „Groteske“. In der Abschlussbefragung meines Experiments zumindest erhielt die *Times* eine magere Durchschnittsnote von „4+“ in den Kategorien „Attraktivität“ und „Lesbarkeit“, während sich die serifenlosen Schriftarten *Arial* und *Verdana* auf einer stabilen und als signifikant besser ermittelten Durch-

schnittsnote „2“ in beiden Kategorien verfestigten. In einem vagen Zwischenreich verorteten die Teilnehmenden meines Experiments übrigens in derselben Befragung die 1976 entwickelte Schriftart *Frutiger*, die nach herrschender Überzeugung nicht nur Pate stand für, sondern verdächtig identisch ist mit der „neuen“ *Microsoft*-Schriftart *Segoe*, die mit dem für 2007 angekündigten neuen Betriebssystem *Vista* ausgeliefert werden soll: Diese Schriftart erhielt für „Attraktivität“ wie subjektiv empfundene „Lesbarkeit“ eine solide, aber keineswegs herausragende Schulnote „3“.

Diese Ergebnisse zeigen, und sie bestätigen damit die Erkenntnisse einiger früherer Studien: Objektiv gemessene und subjektive Einschätzung der Lesbarkeit einer Schriftart am Bildschirm muss keineswegs deckungsgleich ausfallen. Die *Times* ist am Monitor (nicht nur) nach meinen Erhebungen objektiv nicht schlechter lesbar als die *Verdana* – allerdings wird sie in dieser Qualität nicht wahrgenommen. Es muss vorläufig Spekulation bleiben, warum dies so ist – Tatsache ist aber: In meiner Erhebung bewerteten die meisten Teilnehmer „Attraktivität“ und „Lesbarkeit“ einer jeden der vier dargebotenen Schriftarten wechselseitig durchaus unterschiedlich, für jede einzelne Schriftart aber praktisch identisch in beiden Kategorien. Wer der *Verdana* eine Schulnote „2“ für ihre „Attraktivität“ verlieh, vergab in der Regel die selbe Note für deren „Lesbarkeit“. Dieser Zusammenhang ist auffällig und interessant. Ob die Nutzer freilich die Attraktivität einer Schriftart als Indiz für deren vermutete Lesbarkeit werteten, oder ob es sich gerade umgekehrt verhielt, ist eine Frage, die meine Studie aufgeworfen hat, aber nicht beantworten konnte. Folgestudien wären hier hilfreich.

Festzuhalten bleibt insgesamt: Die bloße Ästhetik einer Schriftart, oft als ein der Lesbarkeits-Qualität nachgeordneter Faktor bewertet, scheint in direkterem Verhältnis zur Rezipientenwirkung zu stehen als gelegentlich vermutet. Denn wenn auch eine Schriftart als gemessen „gut lesbar“ erkannt ist: Die Gefahr liegt nahe, dass Rezipienten sie eben nicht in dieser Güte wahrnehmen, sondern im Gegenteil als suboptimal. Dass diese Einschätzung den Rezeptionsprozess nicht unbedingt befördert, liegt als Vermutung nahe, harrt aber ebenfalls eines – bestenfalls empirischen – Nachweises.

Und der Zeitpunkt scheint nicht ungünstig für eine derartige Folgeuntersuchung. Immerhin werden mit der anstehenden neuen Version *Vista* des Betriebssystems *Microsoft Windows* auch diverse neue Schriftarten ihre absehbar weite Verbreitung finden auf den Computern weltweit; Schriftarten, die damit potenziell auch als Web-Schriftarten interessant werden – vorausgesetzt immer, *Windows* wird als Betriebssystem seine derzeitige Quasi-Monopolstellung auch mit der neuen Version behaupten. Woran zu zweifeln allerdings heute wenig Anlass besteht.

Die Bedeutung der Schriftwahl für Webseiten muss also vorläufig und nach meinen empirischen Erkenntnissen unter Vorbehalt betrachtet werden: Unter ästhetischem Gesichtspunkt

scheint sie durchaus relevant, unter dem Aspekt der quantitativ messbaren Lesbarkeit zumindest nachgeordnet. .

Dies gilt eindeutig nicht für jenen typografischen Faktor, den ich in meiner Studie als den einflussreichsten unter den vier getesteten überhaupt ermittelte unter dem Kriterium und der Leitfrage, wie die Lesegeschwindigkeit von Nutzern durch typografische Handgriffe zu optimieren sei: die Zeilenbreite. Insbesondere dieses typografische Phänomen wird in der einschlägigen Literatur geradezu atemberaubend widersprüchlich diskutiert: die Dogmen gehen von „maximal 35 Anschlägen pro Zeile“ bis „mindestens 60 Zeichen“, um nur zwei Extrempole beispielhaft anzuführen. Beide Vorgaben freilich sind nach meinen Erkenntnissen schlichtweg falsch. Wird es jemanden interessieren?

Wie dem auch sei. In einer praxisorientierten Auslegung interpretiere ich die von mir erhobenen Ergebnisse so: Auf Webseiten sollten Zeilen eingerichtet sein, die im Schnitt 40 bis 50 Schriftzeichen Platz bieten. Dieser Korridor ist dabei Optimum und Minimum zugleich: Schmalere Zeilen induzieren empfindliche, um bis zu 7 Prozent (und im Experiment signifikant) verlangsamte Lesezeiten. Breitere Zeilen mit 50 bis 65 Anschlägen reduzieren das Lesetempo insgesamt messbar, aber nicht merklich (4 Prozent Differenz zum Optimalwert), in noch höheren Dimensionen jedoch fielen die Messwert-Unterschiede durchaus ins Gewicht. Allerdings: Auch die breitesten im Experiment getesteten Zeilen (84 Anschläge im Schnitt) zeigten immer noch bessere Mess-Ergebnisse, als für Zeilen mit weniger als 30 Zeichen festgestellt wurden.

Spätestens an dieser Stelle muss allerdings ein Mess-Resultat ins Spiel und zur Erwähnung gebracht werden, das ich in seiner Bedeutung zumindest im letztlich festgestellten Ausmaß nicht erwartet hatte in der Konzeptionsphase meiner Studie: die Konstruktionsart des Monitors, an dem die über 1400 Probanden das in dieser Arbeit beschriebene Online-Experiment durchführten. Ich hatte die Abfrage dieses Phänomens eher routinemäßig als in Erwartung spektakulärer Ergebnisse in den Vorlauf meines Experiments integriert – zu Unrecht, wie sich erweisen sollte. Es wurden nämlich alle Texte in allen angebotenen typografischen Konstellationen an Flachbildschirmen und Laptops in der Gesamtschau um fast 4 Prozent schneller gelesen als an Röhrenbildschirmen.

Doch nicht nur dies. Die gemessenen Lesezeiten für einzelne typografische Konstellationen sowie für verschiedene Kombinationen typografischer Faktoren erwiesen sich an bestimmten Monitorgattungen als ausgeprägter oder erheblich weniger augenfällig als am Vergleichsprodukt. Mehr noch: Teilweise zeigten sich bestimmte typografische Mess-Auffälligkeiten sogar nur an einer Bildschirmart.

Die Ergebnisse für den Faktor Zeilenbreite zumindest fielen in der isolierten Betrachtung an Flachbildschirmen noch eindeutiger aus als in der Gesamtschau: Die oben erwähnte Marge von 40 bis 50 Zeichen pro Zeile als optimaler Zeilenlänge fand in den Ergebnissen der Nut-

zer an TFT-Monitoren eine noch deutlichere Bestätigung als in denen der Probanden, die das Experiment an einem Röhrenbildschirm durchführten.

Dass Zeilenabstand und Zeilenbreite einem Wechselverhältnis unterliegen, war den Ergebnissen meines Experiments hingegen nicht zu entnehmen. Wiewohl in der einschlägigen Literatur oft und gerne unterstellt, fordern breitere Zeilen offenkundig nicht analog vergrößerte Zeilenabstände ein. Es war dies zweifellos einer der auffälligsten „Interaktions“-Befunde meiner Erhebung, die ich ja gerade mit dem Anspruch konzipiert hatte, nicht nur vereinzelte typografische Faktoren auf ihre Wirkmacht zu testen, sondern auch auf ihr Zusammenspiel mit anderen Faktoren. Auch mit den beiden anderen überprüften typografischen Faktoren „Schriftart“ und „Schriftgröße“ interagierte der Zeilenabstand im Übrigen wenig bis gar nicht; lediglich die *Verdana* scheint mit wachsendem Zeilenabstand an Flachbildschirmen sukzessive ganz leicht bessere Lesezeiten hervorzubringen.

Nach meinen Erkenntnissen nimmt der Zeilenabstand insgesamt kaum Einfluss auf die Lesbarkeit von Schriftanordnungen; zumindest, wenn er, wie in meinem Experiment, bemessen und angesiedelt ist in einem Größenbereich des 2,5- bis 3-fachen der x-Höhe der jeweiligen Schriftart, auf die er angewendet wird. Und das ist heute in fast allen seriöseren Angeboten im Web der Fall. Auch zeigten sich keine Auffälligkeiten in der Detailbetrachtung der beiden Monitorgattungen: Der Zeilenabstand scheint als Faktor wenig relevant für die Lesbarkeit einer typografischen Anordnung – an Flach- wie Röhrenbildschirm gleichermaßen.

Durchaus Fehler sind hingegen augenscheinlich begehbar im Bereich der Schriftgröße. Ich bemaß selbige in meinem Experiment anhand der Mittellängen, der x-Höhen also der dargebotenen Schriftarten – denn was hätte es an Erkenntnisgewinn verschafft, etwa eine „offizielle“ 10-Punkt-*Verdana* mit einer 10-Punkt-*Times* zu vergleichen, also numerisch identisch eingerichtete Typografien, wo aufgrund der stark divergierenden Schriftgestalten beider Schriftarten ehrlicherweise ein Abgleich von 10-Punkt-*Verdana* mit 12-Punkt-*Times* angezeigt gewesen wäre? Es war und ist dies ein Hauptmanko der vorliegenden Studien aus angelsächsischen Ländern, dass hier durchweg numerische an Stelle tatsächlicher Buchstabengrößen verglichen wurden; ein Defizit, das ich in meinem Experiment, so meine ich, wirkungsvoll ausschaltete eben durch Abgleich der für die wahrgenommene Schriftgröße nach herrschender Ansicht bedeutsamen x-Höhen.

Zumindest ein Befund scheint dabei eindeutig (und erwies sich in der Varianzanalyse als nahezu ebenbürtig starker Einflussfaktor wie die Zeilenbreite): Schrift in Webbrowsern sollte in ihrer Größe so definiert sein, dass zur Darstellung der Kleinbuchstaben mindestens 7 Bildschirmpunkte (Pixel) in der Vertikalen zur Verfügung stehen. Kleinere Bemessungen führen nach meinen Ergebnissen zu signifikanter Verlangsamung der Lesegeschwindigkeiten. Die *Verdana* sollte mithin nie kleiner eingerichtet werden als in 12 bis 13 Pixeln (das entspricht in der *Windows*-Umgebung 9 bis 10 Punkt) Schriftgröße, die *Times* niemals kleiner als in 14 bis

15 Pixeln (10,5 bis 11 Punkt). Dabei profitiert die *Times* anscheinend und an beiden Bildschirmgattungen von einer weiteren Vergrößerung auf 8 Pixel x-Höhe (Schriftgröße 16 bis 17 Pixel oder 12,5 bis 13 Punkt), während sich bei der *Verdana* in dieser Dimensionierung keine Steigerungen der Lesegeschwindigkeiten mehr zeigen. Es zeigte sich mithin eine Interaktion der Faktoren „Schriftgröße“ und „Schriftart“; weitere Wechselwirkungen fand ich für den Faktor Schriftgröße nicht.

Nimmt man die Mittelwert-Schwelle von 7 Pixel als nicht nur für diese beiden gemessenen Schriftarten, sondern als allgemein gültig an, so sollten auch *Arial* und *Georgia* am Monitor nicht kleiner definiert werden als in 12 bis 13 Pixeln (9 bis 10 Punkt) Schriftgröße.

Bleibt abschließend zu erwähnen jener weitere Einflussfaktor auf die Lesegeschwindigkeit, den ich, vergleichbar der Bildschirmart, eher der demografischen Vollständigkeit halber abfragte im Experiment – um ihn in der Varianzanalyse als offenbar zentrale Wirkmacht zu erkennen: das Lebensalter der Probanden. So lasen Teilnehmende im Alter zwischen 19 und 35 Jahren die Experimentaltexte im Schnitt erheblich schneller als die minderjährigen und insbesondere die über 55-jährigen Probanden; die Differenzen in den Messezeiten lagen bei teilweise über 10 Prozent. Da das hier vorgestellte Experiment vornehmlich auf typografische Faktoren fokussierte, wurde dieser für sich genommen hochinteressante Aspekt nicht weiter verfolgt; eine genauere Untersuchung des Leseverhaltens am Bildschirm nach Altersklassen scheint aber in jedem Falle angezeigt.

Dies gilt nicht weniger für einige typografische Faktoren, die vor allem aus Praktikabilitätsgründen in der hier beschriebenen Studie nicht auf ihren wechselseitigen und individuellen Einfluss auf die Lesegeschwindigkeiten in Webbrowsern getestet wurden. Zu nennen ist hier vor allem der Faktor „Schriftfarbe“, dem in der Literatur eine große Bedeutung zugemessen wird. Nun muss die Prioritätensetzung der Literatur, wie angedeutet, nicht unbedingt korrespondieren mit tatsächlichen Bedeutungen – jedoch erscheint es vordringlich, zumindest jene These empirisch genauer zu prüfen, die sich ungewohnt einhellig durch die Fachbücher zieht, dabei aber kaum praktische Resonanz im Web findet: dass nämlich schwarze Schrift auf weißem Grunde am Bildschirm nicht als angenehm, sondern übermäßig kontrastreich empfunden wird. Sehr unisono fordert die Literatur daher abgeschwächte Schrift-Farbgrund-Kontraste; eine Vorgabe, die, wie gesagt, dringend weiterer Erforschung bedürfte.

Auch im Feld der Schriftarten existiert durchaus noch Forschungsbedarf. Zwar hat die in der vorliegenden Dissertation beschriebene Studie gezeigt, dass die Schriftgestalt offenkundig wenig Einfluss ausübt auf Lesegeschwindigkeiten, die mit typografischen Konstellationen erzielbar sind. Zu überprüfen wird aber sein, ob dies auch für andere Schriftarten als die hier verglichenen *Times* und *Verdana* gilt. Insbesondere jene Schriftarten, die mit *Windows Vista* absehbar die Rechner weltweit bereichern werden, sind lohnende Betrachtungsobjekte einer

möglichen Folgestudie, jedoch auch aktuell häufiger eingesetzte Typen wie *Georgia* oder *Arial*.

Vor allem jedoch lohnte eine gezieltere und breiter angelegte Hinterfragung der Sympathiewerte verschiedener Schriftarten. Schließlich liegt, wie dargelegt, durchaus der Verdacht nahe, dass die reine Schriftästhetik zumindest die Rezeptionsbereitschaft von Nutzern stärker beeinflusst als oft angenommen.

Für all die eben skizzierten Vorhaben scheint das Messinstrumentarium meiner eigenen Studie durchaus tauglich und einer weiteren Verwendung würdig. Obgleich das Stolperwort-Prinzip gelegentlich im Dienst zu versagen schien, wenn Nutzer mehr als 10 Klicks zur Identifizierung von gerade einmal zwei Stolperwörtern in einem rund 100 Wörter langen Text aufwandten, scheint sich die Methodik in der Breite doch bewährt zu haben. Sie scheint insbesondere so selbsterklärend konzipiert, dass sie vor allem in Online-Studien (die ja in aller Regel die Absenz von Versuchsleitern implizieren) gut einsetzbar ist.

Lediglich die Bewertung der Schriftarten zum Abschluss des Experiments bedarf wohl einer Detailkorrektur im Experimentaldesign – derjenigen nämlich, dass die arhythmisch-ordinal organisierte Schulnotenskala durch eine gleich-intervallige Nominalskala ersetzt werden sollte, die auch die Berechnung von seriösen Durchschnitts- und paarweisen Vergleichswerten gestattet.

Das Basis-Konzept „Online-Studie“ als spezifische Form des Feldversuchs habe ich in jedem Falle als tauglich erlebt im Verlaufe meiner Studiendurchführung. Natürlich birgt Sozialforschung via Internet ihre Gefahren und Fallstricke; bei sorgfältiger Versuchskonzeption freilich überragen die dem Konzept des Online-Experiments immanenten Chancen meines Erachtens viele der immer wieder geäußerten Bedenken: Es sind in Online-Experimenten in erstaunlich kurzer Zeit mit sehr geringem Kostenaufwand sehr viele Messdaten einholbar. Hätte ich das in dieser Arbeit beschriebene Experiment als Laborstudie durchgeführt mit dem Anspruch, vergleichbar viele Probanden zu akquirieren, wie es mir letztlich gelang, läge die Arbeit wohl heute noch nicht vor.

8 Quellen

- van Aaaken, Gerrit (2004): Der 72dpi-Mythos. <http://praegnanz.de/essays/72dpi> (Besuch 07.11.2006)
- van Aaaken, Gerrit (2005a): Arial geht, Segoe kommt. <http://praegnanz.de/weblog/Arial-geht-segoe-kommt> (Besuch 07.11.2006)
- van Aaaken, Gerrit (2005b): Georgia. Kuschelige Perfektion. <http://praegnanz.de/essays/georgia> (Besuch 07.11.2006)
- van Aaaken, Gerrit (2006): Spickzettel für Vista-Schriften. <http://praegnanz.de/weblog/spickzettel-fuer-vista-schriften> (Besuch 07.11.2006)
- Abbate, Janet (2000): *Inventing the internet*. Massachusetts
- Abdullah, Rayan/Hübner, Roger (2002): *Corporate Design. Kosten und Nutzen*. Mainz
- Adobe Systems Inc. (2005): Adobe Flash Player Version Penetration. http://www.adobe.com/products/player_census/flashplayer/version_penetration.html (Besuch 12.11.2005)
- Ahumada, Albert J./Scharff, Lauren F. V. (2003): Letter Identification: Contrast Polarity and Speed-Accuracy Trade-Off Strategies. In: Abstracts of the Psychonomic Society 44th. Annual Meeting 8/2003. S. 67ff.
- Aicher, Otl (1989): *typographie*. Lüdenscheid, Berlin
- Aicher, Otl (1992): *analog und digital*. Lüdenscheid, Berlin
- Alkan, Saim Rolf (2002): *Texten für das Internet. Ein Handbuch für Online-Redakteure und Webdesigner*. Bonn
- Altmeppen, Klaus-Dieter/Donges, Patrick/Engels, Kerstin (2002): Technisierung und organisatorischer Wandel. In: Neverla, Irene/Grittmann, Elke/Pater, Monika (Hrsg.) (2002): *Grundlagentexte zur Journalistik*. Konstanz. S. 350ff.
- Anderson, John R. (2001): *Kognitive Psychologie*. Übersetzt und herausgegeben von Ralf Graf und Joachim Grabowski. Heidelberg, Berlin
- Andreyev, Vlad/Martynov, Aleksey (2000): Effects of Splitting Text into Multiple Columns. <http://www.otl.umd.edu/SHORE2000/multicol/index.html> (Besuch 07.11.2006)
- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (2001): *Standards zur Qualitätssicherung für Online-Befragungen*. O.O.
- Atteslander, Peter (2003): *Methoden der empirischen Sozialforschung*. Berlin, New York (10)
- Baker, J. Ryan (2005): Is Multiple-Column Online Text Better? It Depends! <http://psychology.wichita.edu/surl/usabilitynews/72/columns.htm> (Besuch 07.11.2006)
- Ballstaedt, Steffen-Peter/Mandl, Heinz/Schnotz, Wolfgang/Tergan, Sigmar-Olaf (1981): *Texte verstehen, Texte gestalten*. München
- Balzert, Heide (2004): *Webdesign & Web-Ergonomie. Websites professionell gestalten*. Herdecke, Dortmund
- Bandilla, Wolfgang (1999): WWW-Umfragen – eine alternative Datenerhebungstechnik für die empirische Sozialforschung? In: Batinic, Bernad/Werner, Andreas/Gräf, Lorenz/Bandilla, Wolfgang (1999): *Online Research. Methoden, Anwendungen und Ergebnisse*. Göttingen. S. 12ff.
- Bartel, Stefanie (2003): *Farben im Webdesign. Symbolik, Farbpsychologie, Gestaltung*. Berlin, Heidelberg
- Barthel, Gustav (1972): *Konnte Adam schreiben? Weltgeschichte der Schrift*. Köln

- Batinic, Bernad/Werner, Andreas/Gräf, Lorenz/Bandilla, Wolfgang (1999): Online Research. Methoden, Anwendungen und Ergebnisse. Göttingen
- Baumgardt, Michael (1998): Web Design kreativ! Mit CD-ROM und zahlreichen Abbildungen. Berlin, Heidelberg, New York, Barcelona, Budapest, Hongkong, London, Mailand, Paris, Santa Clara, Singapur, Tokio
- Bayer, Sonja Katrin (2003): Bildschirmtypografie. Technische und psychologische Determinanten der Gestaltung von Online-Dokumenten. Erlangen, Nürnberg
- Beier, Markus/von Gizycki, Vittoria (Hrsg.) (2002): Usability. Nutzerfreundliches Web-Design. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio
- Bentele, Günter/Hesse, Kurt R. (Hrsg.) (1994): Publizistik in der Gesellschaft. Festschrift für Manfred Rühl. Konstanz
- Bergner, Walter (1990): Grundlagen der Typografie. Leipzig
- Bernard, Michael/Chaparro, Barbara/Thomasson, Ray (2000): Finding Information on the Web: Does the Amount of Whitespace Matter? <http://psychology.wichita.edu/surl/usabilitynews/2W/whitespace.htm> (Besuch 07.11.2006)
- Bernard, Michael/Fernandez, Marissa/Hull, Spring (Bernard et al. 2002a): The Effects of Line Length on Children and Adults' Online Reading Performance. http://psychology.wichita.edu/surl/usabilitynews/42/text_length.htm (Besuch 07.11.2006)
- Bernard, Michael/Liao, Corrina/Mills, Melissa (Bernard et al. 2001a): Determining the Best Online Font for Older Adults. <http://psychology.wichita.edu/surl/usabilitynews/3W/fontSR.htm> (Besuch 07.11.2006)
- Bernard, Michael/Lida, Bonnie/Riley, Shannon/Hackler, Telia/Janzen, Karen (Bernard et al. 2002b): A Comparison of Popular Online Fonts: Which Size and Type ist Best? <http://psychology.wichita.edu/surl/usabilitynews/41/onlinetext.htm> (Besuch 07.11.2006)
- Bernard, Michael/Mills, Melissa (2000): So, What Size and Type of Font Should I Use on My Website? <http://psychology.wichita.edu/surl/usabilitynews/2S/font.htm> (Besuch 07.11.2006)
- Bernard, Michael/Mills, Melissa/Frank, Talissa/McKnown, Jan (Bernard et al. 2001b): Which Fonts Do Children Prefer to Read Online? <http://psychology.wichita.edu/surl/usabilitynews/3W/font-JR.htm> (Besuch 07.11.2006)
- Bernard, Michael/Mills, Melissa/Peterson, Michelle/Storrer, Kelsey (Bernard et al. 2001c): A Comparison of Popular Online Fonts: Which is Best and When? <http://psychology.wichita.edu/surl/usabilitynews/3S/font.htm> (Besuch 07.11.2006)
- Blankertz, Stefan/Doubrawa, Erhard (2005): Lexikon der Gestalttherapie. Köln
- Blöbaum, Bernd (2000): Zwischen Redaktion und Reflektion. Integration von Theorie und Praxis in der Journalistenausbildung. Münster
- Bloomer, Carolyn M. (1976): Principles of visual perception. New York, Cincinetti, Toronto, London, Melbourne
- Blum, Joachim/Bucher, Hans-Jürgen (1998): Die Zeitung: Ein Multimediu. Textdesign – ein Gestaltungskonzept für Text, Bild und Grafik. Konstanz
- Blumenstock, Leonhard (1997): Handbuch der Leseübungen. Vorschläge und Materialien zur Gestaltung des Erstleseunterrichts mit Schwerpunkt im sprachlich-akustischen Bereich. Weinheim, Basel
- Blythe, Mark A./Overbeeke, Kees/Monk, Andrew F./Wright, Peter C. (Hrsg.) (2003): Funology. From Usability to Enjoyment. Dordrecht

- Böhringer, Joachim/Bühler, Peter/Schlaich, Patrick/Ziegler, Hanns-Jürgen (2003): Kompendium der Mediengestaltung für Digital- und Printmedien. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio (2)
- Bolesch, Cornelia (1995): Mit vollem Schwung zurück an den Start. In: Süddeutsche Zeitung vom 06.03.1995. S. 3
- Bollwage, Max (2001): Typografie kompakt. Vom richtigen Umgang mit Schrift am Computer. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio
- Bonfadelli, Heinz/Jarren, Otfried/Siegert, Gabriele (Hrsg.) (2001): Einführung in die Publizistikwissenschaft. Bern, Stuttgart, Wien (2)
- Bonsiepe, Gui (1996): Interface. Design neu begreifen. Mannheim
- Bortz, Jürgen (2004): Statistik für Human- und Sozialwissenschaftler. Berlin (6)
- Boyarski, Dan/Neuwirth, Christine/Forlizzi, Jodi/Harkness Regli, Susan (1998): A Study of Fonts Designed for Screen Display. In: Proceedings of the SIGCHI conference on Human factors in computing systems 1998. S. 87ff.
- Braun, Gerhard (1987): Grundlagen der visuellen Kommunikation. München
- Braun, Katrin (2006): Klicken statt blättern. Deutsche Zeitungen im Internet. In: Rager, Günther/Graf-Szczuzka, Karola/Hassemer, Gregor/Süper, Stephanie (Hrsg.) (2006): Zeitungsjournalismus. Empirische Leserschaftsforschung. Konstanz. S. 68ff.
- Brekle, Herbert E. (1994): Typographie. In: Günther, Hartmut/Ludwig, Otto (Hrsg.) (1994): Schrift und Schriftlichkeit. Writing and its use. Ein interdisziplinäres Handbuch internationaler Forschung. 1. Halbband (Handbücher zur Sprach- und Kommunikationswissenschaft 10.1). Berlin, New York. S. 204ff.
- Brielmaier, Peter/Wolf, Eberhard (2000): Zeitungs- und Zeitschriftenlayout. Konstanz (2)
- F. A. Brockhaus GmbH (Hrsg.): Brockhaus! Was so nicht im Lexikon steht. Kurioses und Schlaues aus allen Wissensgebieten. Augsburg 2000
- Brosius, Hans-Bernd/Koschel, Friederike (2005): Methoden der empirischen Sozialforschung. Eine Einführung. Wiesbaden (3)
- Bürgel, Matthias/Neumann, Walter (2001): Screen Design und visuelle Kommunikation. Gestaltung interaktiver Oberflächen. Hüthig
- Cavanaugh, Sean (1997): Insiderbuch Type-Design. Zürich
- Champeon, Steven/Fox, David S. (1999): Building Dynamic HTML GUIs. Foster City (CA), Chicago (IL), Indianapolis (IN), New York (NY)
- Chaparro, Barbara S./Baker, J. Ryan/Shaiikh, A. Dawn/Hull, Spring/Brady, Laurie (2004): Reading Online Text: A Comparison of Four White Space Layouts. <http://psychology.wichita.edu/surl/usabilitynews/62/whitespace.htm> (Besuch 07.11.2006)
- Chaparro, Barbara S./Shaiikh, A. Dawn/Chaparro, Alex (2006): Examining the Legibility of Two New ClearType Fonts. <http://psychology.wichita.edu/surl/usabilitynews/81/legibility.htm> (Besuch 07.11.2006)
- Chen, Qijang (2002): Human Factors in Interface Design: An Analytical Survey and Perspective. In: Szewczak, Edward/Snodgrass, Coral (Hrsg.) (2002): Human Factors in Information Systems. Hershey, London, Melbourne, Singapore, Beijing. S. 49ff.
- Chung, I-Hsin/Kolatch, Erica/Sculimbrene, Sofia/Wen, Hui-Fang (2000): The Effect of Screen Size on Readability Using Three Different Portable Devices. <http://www.otal.umd.edu/SHORE2000/portadevs/> (Besuch 07.11.2006)

- Cloninger, Curt (2000): Usability Experts are from Mars, graphic designers are from Venus. <http://www.alistapart.com/articles/marsvenus/> (Besuch 07.11.2006)
- Constantine, Larry L./Lockwood, Lucy A. D. (1999): Software for use: a practical guide to the models and methods of usage-centered design. Reading (Massachusetts), Harlow (England), Menlo Park (California), Don Mills (Ontario), Sydney, Bonn, Amsterdam, Tokyo, Mexico City
- Couper, Mick P./Coutts, Elisabeth (2006): Probleme und Chancen verschiedener Arten von Online-Erhebungen. In: Diekmann, Andreas (Hrsg.) (2006): Methoden der Sozialforschung. Wiesbaden. S. 217ff.
- Crederholm, Dan (2006): Bulletproof WebDesign. Absolut flexibel und für alles gewappnet mit CSS und XHTML. München
- Creed, Arthur/Dennis, Ian/Newstead, Steven (1988): Effects of display format on proof-reading with VDUs. In: Behaviour & Information Technology 7/1988, S. 467ff.
- Dengler, Florian/Volland, Holger (2000): Webdesign professionell. Expertenstrategien und Tipps von Pixelpark, frogdesign und Echopool. Bonn
- Diekmann, Andreas (Hrsg.) (2006): Methoden der Sozialforschung. Wiesbaden
- Dillon, Andrew (1992): Reading from paper versus screens: A critical review of the empirical literature. In: Ergonomics, 35/1992. S. 1297ff.
- Dillon, Andrew/Kleinman, Lisa/Bias, Randolph/Choi, Gil Ok/Turnbull, Don (2004): Reading and Searching Digital Documents: An Experimental Analysis of the Effects of Image Quality on User Performance and Perceived Effort. Austin/Texas (unveröffentlichtes Skript)
- Dillon, Andrew/Kleinman, Lisa/Choi, Gil Ok/Bias, Randolph (2006): Visual Search and Reading Tasks Using ClearType and Regular Displays: Two Experiments. In: CHI 06 conference proceedings. Montreal 2006. S. 12ff.
- Dillon, Andrew/McKnight, Cliff/Richardson, John (1988): Reading from paper versus reading from screens. In: The Computer Journal 31/1988. S. 457ff.
- Dörmann, Jürgen/Pätzold, Ulrich (1998): Journalismus, neue Technik, Multimedia und Medienentwicklung. Ein Plädoyer für journalistische Produktion und Qualifikation in den Neuen Medien. In: journalist 7/1998. S. 59ff.
- Droste, Magdalena (1998): bauhaus. Berlin
- Duller, Christine (2006): Einführung in die Statistik mit EXCEL und SPSS. Ein anwendungsorientiertes Lehr- und Arbeitsbuch. Heidelberg
- Dyson, Mary C. (2004): How physical text layout affects reading from screen. In: Behaviour & Information Technology 23/2004. S. 377ff.
- Dyson, Mary C./Gregory, Judy (2002): Typographic cueing on screen. In: Visible Language. 36/2002. S. 326ff.
- Dyson, Mary C./Haselgrove, Mark (2000): The Effects of Reading Speed and Reading Patterns on Our Understanding of Text Read from Screen. In: Journal of Research in Reading 23/2000. S. 210ff.
- Dyson, Mary C./Haselgrove, Mark (2001): The influence of reading speed and line length on the effectiveness of reading from screen. In: International Journal of Human-Computer Studies 54/2001. S. 585ff.
- Dyson, Mary C./Kipping, Gary J. (1997): The legibility of screen formats: Are three columns better than one? In: Computers & Graphics 21/1997. S. 703ff.
- Dyson, Mary C./Kipping, Gary J. (1998): The Effects of Line Length and Method of Movement on Patterns of Reading from Screen. In: Visible Language 32/1998. S. 151ff.

- Eberleh, Edmund/Oberquelle, Horst/Oppermann, Horst (Hrsg.) (1994): Einführung in die Software-Ergonomie. Berlin, New York
- Eco, Umberto (1972): Einführung in die Semiotik. München
- Eibl, Maximilian (2003): Visualisierung im Document Retrieval: theoretische und praktische Zusammenführung von Softwareergonomie und Graphik Design. Bonn (2)
- Eichhorn, Wolfgang (1993): Das Experiment. In: Knieper, Thomas (Hrsg.) (1993): Statistik. Eine Einführung für Kommunikationsberufe. München. S. 163ff.
- van Eimeren, Birgit/Frees, Beate (2005): Nach dem Boom: Größter Zuwachs in internetfernen Gruppen. ARD/ZDF-Online-Studie 2005. In: Media Perspektiven 8/2005. S. 362ff.
- van Eimeren, Birgit/Frees, Beate (2006): Schnelle Zugänge, neue Anwendungen, neue Nutzer? ARD/ZDF-Online-Studie 2006. In: Media Perspektiven 8/2006. S. 402ff.
- van Eimeren, Birgit/Gerhard, Heinz/Frees, Beate (2004): Internetverbreitung in Deutschland: Potenzial vorerst ausgeschöpft? ARD/ZDF-Online-Studie 2004. In: Media Perspektiven 8/2004. S. 350ff.
- EN ISO 9241 (Usability)
- Erben, Ben (1998): Schrift & Farbe am Bildschirm. Mainz
- Evans, Harold (1974): Editing and Design. A Five-Volume Manual of English, Typography and Layout. London
- Evans, Harold (1974): Handling Newspaper Text.. In: Evans, Harold (1974): Editing and Design. A Five-Volume Manual of English, Typography and Layout. London. Buch 5
- Faulkner, Kristine (2000): Usability Engineering. London
- Fisch, Martin/Gscheidle, Christoph (2006): Onliner 2006: Zwischen Breitband und Web 2.0 – Ausstattung und Nutzungsinnovation. Ergebnisse der ARD/ZDF-Online-Studien 1997 bis 2006. In: Media Perspektiven 8/2006. S. 431ff.
- Földes-Papp, Károly (1966): Vom Felsbild zum Alphabet. Die Geschichte der Schrift von ihren frühesten Vorstufen bis zur modernen lateinischen Schreibschrift. Stuttgart
- Forschungsgruppe Wahlen (O.J.): Milieuquotierte Stichproben. O.O. http://www.forschungsgruppe.de/Online-Forschung/Milieuquotierte_Stichproben/ (Besuch 07.11.2006)
- Friedl, Friedrich (1998): Die Univers von Adrian Frutiger. Frankfurt am Main
- Friedrich, Johannes (1966): Geschichte der Schrift. Unter besonderer Berücksichtigung ihrer geistigen Entwicklung. Heidelberg
- Frick, Richard/Graber, Christine (2001): Typografie am Bildschirm. Bern
- Frutiger, Adrian (2001): Der Mensch und seine Zeichen. Schriften, Symbole, Signete, Signale. Wiesbaden (8)
- Funke, Ralf (1992): Die dritte Abstraktion. In: Karow, Peter (1992): Schrifttechnologie. Methoden und Werkzeuge. Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest. S. 77ff.
- Garcia, Mario R. (1987): Contemporary Newspaper Design. Englewood Cliffs/Prentice Hall
- Garcia, Mario (1997): Schöner lesen! In: Die Zeit 6/1998. S. 15ff.
- Gerhards, Maria/Mende, Annette (2006): Offliner: Vorbehalte und Einstiegsbarrieren gegenüber dem Internet bleiben bestehen. ARD/ZDF-Offline-Studie 2006. In: Media Perspektiven 8/2006. S. 416ff.

- George, Helen (1995): *The Good Usability Handbook*. London, New York, St. Louis, San Francisco, Auckland, Bogotá, Caracas, Lisbon, Madrid, Mexico, Milan, Montreal, New Delhi, Panama, Paris, San Juan, Sao Paulo, Singapore, Sydney, Tokyo, Toronto
- Gerstner, Karl (1990): *Kompendium für Alphabeten. Eine Systematik der Schrift*. Heiden
- Gibson, James J. (1982): *Wahrnehmung und Umwelt. Der ökologische Ansatz in der visuellen Wahrnehmung*. München, Wien, Baltimore
- Glaser, Wilhelm R. (1998): *Wahrnehmung*. In: Straub, Jürgen/Kempf, Wilhelm/Werbik, Hans (Hrsg.) (1998): *Psychologie. Eine Einführung. Grundlagen, Methoden, Perspektiven*. München. S. 225ff.
- Goldstein, E. Bruce (2002): *Wahrnehmungspsychologie. Zweite deutsche Ausgabe herausgegeben von Manfred Ritter*. Heidelberg, Berlin (2)
- Gorbach, Rudolf Paulus (2001): *Typografie professionell*. Bonn
- Götz, Veruschka (1998): *Schrift & Farbe am Bildschirm*. Mainz
- Gould, James D./Grischkowsky, Neil (1984): *Doing the same work with hard copy and with cathode-ray tube (CRT) computer terminals*. In: *Human Factors*, 26/1984. S. 323ff.
- Gross, Sabine (1994): *Lese-Zeichen. Kognition, Medium und Materialität im Leseprozess*. Darmstadt
- Grube, Gernot/Kogge, Werner/Krämer, Sibylle (Hrsg.) (2005): *Schrift. Kulturtechnik zwischen Auge, Hand und Maschine*. München
- Gulbins, Jürgen/Kahrmann, Christine (1992): *Mut zur Typographie. Ein Kurs für DTP und Textverarbeitung*. Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest
- Günher, Hartmut/Ludwig, Otto (Hrsg.) (1994): *Schrift und Schriftlichkeit. Writing and its use. Ein interdisziplinäres Handbuch internationaler Forschung. 1. Halbband (Handbücher zur Sprach- und Kommunikationswissenschaft 10.1)*. Berlin, New York
- Guski, Rainer (1996): *Wahrnehmen. Ein Lehrbuch*. Stuttgart
- Hacke, Axel (1994): *Das Magazin bin ich! FOCUS – oder wie man dem Medienmarkt Feuer macht*. In: *Süddeutsche Zeitung* vom 23.08.1994. S. 3
- Hafner, Katie/Lyon, Mathew (2000): *ARPA KADABRA oder Die Geschichte des Internet*. Heidelberg
- Hagge, Kira (1994): *Informations-Design*. Heidelberg
- Hallnäs, Lars/Reström, Johan (2002): *From Use to Presence: On the Expressions and Aesthetics of Everyday Computational Things*. In: *ACM transactions on Computer-Human Interaction* 9/2002. S. 106ff.
- Hartley, James (1985): *Designing instructional text*. New York
- Hartley, James (1987): *Designing Electronic Text: The Role of Print-Based Research*. In: *ERIC/ECTJ Annual Review* 14/1987. S. 3ff.
- Hassemer, Gregor/Rager, Günther (2006): *Das Bessere als Feind des Guten – Qualität in der Tageszeitung*. In: Rager, Günther/Graf-Szczuzuka, Karola/Hassemer, Gregor/Süper, Stephanie (Hrsg.) (2006): *Zeitungsjournalismus. Empirische Leserschaftsforschung*. Konstanz. S. 19ff.
- Hassenzahl, Marc/Platz, Axel/Burmester, Michael/Lehner, Katrin (2000): *Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal*. In: *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems*. S. 201ff.

- Hauben, Michael/Hauben, Ronda (1997): Netizens: On the History and Impact of Usenet and the Internet. Los Alamitos (California)
- Hauptmanns, Peter: Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. In: Batinic, Bernad/Werner, Andreas/Gräf, Lorenz/Bandilla, Wolfgang (1999): Online Research. Methoden, Anwendungen und Ergebnisse. Göttingen. S. 24ff.
- Hauser, Tobias/Wenz, Christian (2002): Web Publishing. Design und Programmierung. München
- Heijnk, Stefan (2002): Texten fürs Web. Grundlagen und Praxiswissen für Online-Redakteure. Heidelberg
- Heinecke, Achim/Köpcke, Wolfgang (O.J.): JUMBO – Java-unterstützte Münsteraner Biometrie-Oberfläche. Glossar. <http://www.mh-hannover.de/institute/biometrie/JUMBO/bio/bio.html> (Besuch 07.11.2006)
- Heitz, Olivier/Täschler, Christof/Blum, Claudia (1998): WebDesign in der Praxis. Kilchberg/CH
- Hellbusch, Jan Eric (2005): Barrierefreies Webdesign. Praxishandbuch für Webgestaltung und grafische Programmoberflächen. Heidelberg
- Heller, Eva (2002): Wie Farben wirken. Farbpsychologie, Farbsymbolik, kreative Farbgestaltung. Reinbek
- Herrmann, Ralf (2003): Index Schrift. Schriften für DTP, Screen, Dekor. Schriftmuster und Beispielanwendungen. Mit zahlreichen Opentype-Fonts. Bonn
- Herrmann, Ralf (2005): Zeichen setzen. Satzwissen und Typoregeln für Textgestalter. Bonn
- Hill, Alyson L./Scharff, Lauren F. V. (1997): Readability of Websites With Various Foreground/Background Color Combinations, Font Types and Word Styles. <http://hubel.sfasu.edu/research/AHNCUR.html> (Besuch 07.11.2006)
- Hillstrom, Kevin (2005): Defining moments. The internet revolution. Detroit
- Hingst, Armin (1992): Weißraum weltweit. Trends und Perspektiven der Zeitungsgestaltung. In: Rager, Günther/Werner, Petra (Hrsg.) (1992): Die tägliche Neu-Erscheinung. Untersuchungen zur Zukunft der Zeitung. Münster, Hamburg. S. 95ff.
- Hingst, Armin (2006): Prägnante Bilder, passende Texte, und dann ...? Leerstellen der Rezeptionsforschung zum Zeitungslayout. In: Rager, Günther/Graf-Szczuzuka, Karola/Hassemer, Gregor/Süpper, Stephanie (Hrsg.) (2006): Zeitungsjournalismus. Empirische Leserschaftsforschung. Konstanz. S. 103ff.
- Hochuli, Jost (1987): Das Detail in der Typographie. Wilmington
- Hofstätter, Peter R./Wendt, Dirk (1974): Quantitative Methoden der Psychologie. Band 1: Deskriptive, Inferenz- und Korrelationsstatistik. Frankfurt
- Hoofacker, Gabriele (2004): Online-Journalismus. Würzburg
- Hüttner, Paul (2005): Voraussetzungen für t-Tests bei unabhängigen Stichproben. <http://www.psychologie.uni-freiburg.de/Members/leonhart/Skripte/SPSS/node104.html> (Besuch 07.11.2006)
- Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern (IVW) (2006): Online-Nutzungsdaten. <http://ivwonline.de/ausweisung2/suchen2.php> (Besuch 07.11.2006)
- Janetzko, Dietmar (1999): Statistische Anwendungen im Internet. Daten in Netzumgebungen erheben, auswerten und präsentieren. München
- Jansen, Angela/Scharfe, Wolfgang (1999): Handbuch der Infografik. Berlin, Heidelberg
- Jean, Georges (1991): Die Geschichte der Schrift. Ravensburg

- Jordan, Patrick W. (2002): *Designing Pleasurable Products. An introduction to the new human factors*. London, New York
- Kapr, Albert (1983): *Ästhetik der Schriftkunst*. Leipzig
- Karow, Peter (1992): *Schrifttechnologie. Methoden und Werkzeuge*. Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest
- Kent, Peter (1999): *Das ABC der eigenen Website. So sparen Sie Zeit und Geld*. München
- Kersken, Sascha (2003): *Kompendium der Informationstechnik. EDV-Grundlagen, Programmierung, Mediengestaltung*. Bonn
- Kesper, Björn/Matthai, Matthias/Lindner, Martin (2004): *HTML, XHTML & CSS*. Düsseldorf
- Kiehn, Anja/Titzmann, Ina (1998): *Typographie interaktiv! Ein Leitfaden für gelungenes Screen-Design*. Berlin, Heidelberg, New York, Barcelona, Budapest, London, Mailand, Paris, Santa Clara, Singapur, Tokio
- Klammer, Bernd (2005): *Empirische Sozialforschung. Eine Einführung für Kommunikationswissenschaftler und Journalisten*. Konstanz
- Klanten, Robert (Hrsg.) (2000): *72 dpi*. Berlin
- Klein, Jakob (2004): *Good Bye Arial?* In: *Süddeutsche Zeitung* vom 28.04.2004. S. 14
- Klein, Jakob (2005): *Der Bauch des „a“*. In: *Süddeutsche Zeitung* vom 21.12.2005. S. 13
- Knieper, Thomas (Hrsg.) (1993): *Statistik. Eine Einführung für Kommunikationsberufe*. München
- Kobert, Thomas (2004): *HTML 4. Das bhv Taschenbuch*. Landsberg
- Kol, Sara/Scholnik, Miriam (2000): *Enhancing Screen Reading Strategies*. In: *CALICO Journal* 18/2000. S. 67ff.
- Kolers, Paul A./Duchnicky, Robert L./Ferguson, Donald C. (1981): *Eye movement measurement of readability of CRT displays*. In: *Human Factors* 23/1981. S. 517ff.
- Kommer, Isolde/Mersin, Tilly (2002): *Typografie und Layout für digitale Medien*. München, Wien
- Köhler, Ralf (2002): *Typo & Design*. Bonn
- Kornelius, Stefan/Milz, Annette (1992): *Laßt mich den Löwen spielen*. In: *Medium Magazin* 6/1992. S. 4ff.
- Kösler, Bertram (1992): *Gebrauchsanleitungen richtig und sicher gestalten*. Wiesbaden
- Krämer, Walter (1997): *So lügt man mit Statistik*. Frankfurt, New York
- Krämer, Walter (2006): *Statistik: Vom Geburtshelfer zum Bremser der Erkenntnis in den Sozialwissenschaften?* In: *Diekmann, Andreas (Hrsg.) (2006): Methoden der Sozialforschung*. Wiesbaden. S. 54ff.
- Kreuzberger, Thomas (1997): *Internet. Geschichte und Begriffe eines neuen Mediums*. Wien, Köln, Weimar, Böhlau
- Kromrey, Helmut (2006): *Empirische Sozialforschung*. Stuttgart (11)
- Krone, Dieter (Hrsg.) (2003): *Woher kommt der Ausdruck ... Vergnügliches zum Ursprung der Ausdrücke*. Leichlingen
- Kübler, Magdalene/Struppek, Holger (1996): *Web-Design. Praxis der Seitengestaltung im World Wide Web*. Heidelberg 1996
- Kunz, Willi (1998): *Typographie – Makro- und Mikroästhetik*. Heiden

- Küpper, Norbert (1984): Die Gestaltung von Zeitungen aus der Sicht eines Grafikers. Teil I: Die Zeitungsköpfe müssen vom Ballast befreit werden. In: Deutscher Drucker 22/1984. S. 21ff.
- Küpper, Norbert (2004): Zeitungsgestaltung. In: Pürer, Heinz/Rahofer, Meinrad/Reitan, Claus (Hrsg.) (2004): Praktischer Journalismus. Presse, Radio, Fernsehen, Online. Konstanz
- von La Roche, Walther (2003): Einführung in den praktischen Journalismus. Mit genauer Beschreibung aller Ausbildungswege. München
- Laborenz, Kai (2005): CSS-Praxis. Grundlagen und Referenz, Praxisbeispiele, Tipps und Tricks, barrierefreies Webdesign mit CSS. Bonn
- Lackerbauer, Ingo (2003): Handbuch für Online-Texter und Online-Redakteure. Berlin
- Lam, Kevin/Lam, Yan/Liu, Joshua/Shin, Un-Gi (2000): Reading Comprehension and Rate: One Column vs. Three Columns. <http://www.otal.umd.edu/SHORE2000/columns/index.html> (Besuch 07.11.2006)
- Lamprecht, Stephan (2002): WebDesign-Handbuch. Websites programmieren und gestalten. München, Wien
- de Lange, Rudi W./Esterhuizen, Hendry L./Beatty, Derek (1993): Performance differences between Times and Helvetica in a reading task. In: Electronic Publishing 6/1993. S. 241ff.
- Lankau, Ralf (2000): Webdesign und -publishing. Grundlagen und Designtechniken. München, Wien (2)
- Larson, Kevin (2004): The Science of Word Recognition. Or how I learned to stop worrying and love the bouma. <http://www.microsoft.com/typography/ctfonts/WordRecognition.aspx#m1> (Besuch 07.11.2006)
- Lazar, Jonathan (2005): Web Usability. A user-centered design approach. Boston, San Francisco, New York, London, Toronto, Sydney, Tokyo, Singapore, Madrid
- Le Peuple, Jenny/Scane, Robert (2003): User Interface Design. Exeter (GB)
- Leeson, Heidi V. (2006): The Mode Effect: A Literature Review of Human and Technological Issues in Computerized Testing. In: International Journal of Testing 6/2006. S. 1ff.
- Lefrancois, Guy R. (1994): Psychologie des Lernens. Heidelberg, Berlin
- Leroi-Gourhan, André (1980): Hand und Wort. Frankfurt a. M.
- Liebig, Martin (1999): Die Infografik. Konstanz
- Lindsay, Peter H./Norman, Donald A. (1981): Psychologie. Informationsaufnahme und -verarbeitung beim Menschen. Übersetzt von H.-D. Dumpert, F. Schmidt, M. Schuster, M. Steeger. Berlin, Heidelberg, New York
- Lubkowitz, Mark (2005): Webseiten programmieren und gestalten. Bonn
- Luidl, Philipp (1989): Typografie. Herkunft, Aufbau, Anwendung. Hannover (2)
- Lynch, Patrick/Horton, Sarah (1999): Web style guide. Basic principles for creating Web sites. New Haven
- Maier, Michael (2004): Herausforderungen einer Online-Zeitung. In: Mast, Claudia (Hrsg.) (2004): ABC des Journalismus. Ein Handbuch. Konstanz. S. 681ff.
- Makowski, Dirk (2006): Windows 3.0. <http://www.winhistory.de/index.html> (Besuch 07.11.2006)
- Mast, Claudia (Hrsg.) (2004): ABC des Journalismus. Ein Handbuch. Konstanz

- Mathes, Harald (1993): Schätzen und Testen. In: Knieper, Thomas (Hrsg.) (1993): Statistik. Eine Einführung für Kommunikationsberufe. München. S.333ff.
- Maurice, Florence/Rex, Patricia (2006): CSS. Standardkonformes Webdesign mit Cascading Style Sheets. München
- Mayer, Horst O. (2000): Einführung in die Wahrnehmungs-, Lern- und Werbepsychologie. München
- McClelland, James L./Rumelhart, David E. (1980): An Interactive Model of the Effect of Context in Perception. San Diego/California
- McKelvey, Roy (2000): Hypergraphics. Design und Architektur von Websites. Reinbek
- McLean, Ruari (1992): The Thames and Hudson Manual of Typography. London
- McMullin, Jess/Varnhagen, Connie K./Heng, Pheng/Apedoe, Xornam (2002): Effects of Surrounding Information and Line Length on Text Comprehension from the Web. In: Canadian Journal of Learning and Technology 28/2002. S. 19ff.
- Meier, Cordula (Hrsg.) (2003): Design-Theorie. Beiträge zu einer Disziplin. Frankfurt
- Meier, Klaus (2002): Internet-Journalismus. Konstanz
- Meiert, Jens/Helmdach, Ingo (2006): Webdesign mit CSS. Designer-Techniken für kreative und moderne Webseiten. Beijing, Cambridge, Farnham, Köln, Paris, Sebastopol, Taipei, Tokyo
- Menzel, Wolfgang (1990): Lesen lernen – schreiben lernen. Braunschweig
- Merten, Klaus/Schmidt, Siegfried J./Weischenberg, Siegfried (Hrsg.) (1994): Die Wirklichkeit der Medien. Eine Einführung in die Kommunikationswissenschaft. Opladen
- Metze, Wilfried (O.J.): Kostenlose Lesetests. <http://www.wilfriedmetze.de/html/lesetest.html> (Besuch 07.11.2006)
- Metzger, Wolfgang (1966): Handbuch der Psychologie. Band 1. Göttingen
- Metzger, Wolfgang (1976): Gesetze des Sehens. Frankfurt
- Meyer, Eric K. (1997): Designing Infographics. Theory, creative techniques & practical solutions. Indianapolis
- Microsoft Corp. (1997): Channel Verdana. <http://www.microsoft.com/typography/web/fonts/Verdana/default.htm> (Besuch 07.11.2006)
- Microsoft Corp. (2004): Where Design meets Technology: Microsoft ClearType Font Collection. <http://www.microsoft.com/resources/design/ClearType.html> (Besuch 07.11.2006)
- Microsoft Corp. (2006): Georgia. <http://www.microsoft.com/typography/fonts/font.aspx?FID=4&FNAME=Georgia> (Besuch 07.11.2006)
- Mittenecker, Erich (1966): Planung und statistische Auswertung von Experimenten. Wien
- Mohr, Winfried (1984): Visuelle Wahrnehmung und Zeichenfunktion. Untersuchungen zur Grundlage des Kategorieneffekts bei der Wahrnehmung von Buchstaben und Ziffern. Regensburg
- Morrison, Sarah/Noyes, Jan (2003): A Comparison of Two Computer Fonts: Serif versus Ornate Sans Serif. http://psychology.wichita.edu/surl/usabilitynews/52/UK_font.htm (Besuch 07.11.2006)
- Müller, Marion G. (2003): Grundlagen der visuellen Kommunikation. Theorieansätze und Analysemethoden. Konstanz

- Müller von Blumencron, Mathias (2004): Ein Haus auf Treibsand errichten. In: Mast, Claudia (Hrsg.) (2004): ABC des Journalismus. Ein Handbuch. Konstanz. S. 677ff.
- Münz, Stefan (2005): Ältere Elemente zur Schriftformatierung.
<http://de.selfhtml.org/html/text/schrift.htm> (Besuch 07.11.2006)
- Muter, Paul/Latremouille, Susane A./Treurniet, William C./Beam, Paul (1982): Extended Reading of Continuous Text on Television Screens. In: Human Factors 24/1982. S. 501ff.
- Muter, Paul/Maurutto, Paula (1991): Reading and skimming from computer screens and books: the paperless office revisited? In: Behaviour & Information Technology 10/1991. S. 257ff.
- Neel, Ann F. (1974): Handbuch der psychologischen Theorien. München
- Neisser, Ulric (1974): Kognitive Psychologie. Stuttgart 1974
- Netcraft Ltd. (2006): September 2006 Web Server Survey.
http://news.netcraft.com/archives/web_server_survey.html (Besuch 07.11.2006)
- Neuenschwander, Brody (2001): Letterwork. Creative letterforms in graphic design. London, New York
- Neutzling, Ulli (2002): Typo und Layout im Web. Reinbek
- Neverla, Irene/Grittmann, Elke/Pater, Monika (Hrsg.) (2002): Grundlagentexte zur Journalistik. Konstanz
- Niederst, Jennifer/Freedman, Edie (1996): Web-Design. Gestaltungsgrundlagen für ein neues Medium. Bonn
- Nielsen, Jakob (1993): Usability Engineering. London
- Nielsen, Jakob (1997a): Be succinct! (Writing for the Web).
<http://www.useit.com/alertbox/9703b.html> (Besuch 07.11.2006)
- Nielsen, Jakob (1997b): How Users Read on the Web.
<http://www.useit.com/alertbox/9710a.html> (Besuch 07.11.2006)
- Nielsen, Jakob (2000a): Erfolg des Einfachen. München
- Nielsen, Jakob (2000b): Flash: 99% Bad. <http://www.useit.com/alertbox/20001029.html> (Besuch 07.11.2006)
- Nielsen, Jakob/Tahir, Marie (2002): Homepage Usability. 50 websites deconstructed. Washington
- Norman, Donald E. (2002): Emotion & Attractive. In: Interactions: New Visions of Computer Interaction 9/2002. S. 36ff.
- Nöth, Winfried (1985): Handbuch der Semiotik. Stuttgart
- O.A. (2004): Spiegel online feiert zehnjähriges Bestehen. <http://shortnews.stern.de/show-news.cfm?id=525464&CFID=8690&CFTOKEN=58320325> (Besuch 07.11.2006)
- O'Hara, Kenton/Sellen, Abigail (1997): A Comparison of Reading Paper and On-Line Documents. Cambridge (unveröffentlichtes Manuskript)
- Osborne, David J./Holton, Doreen (1988): Reading from screen versus paper: There is no difference. In: International Journal of Man-Machine Studies 28/1988. S. 1ff.
- Oehmichen, Ekkehardt/Schröter, Christian (2006): Internet im Medienalltag: Verzögerte Aneignung des Angebots. Ergebnisse der OnlineNutzerTypologie in der ARD/ZDF-Online-Studie. In: Media Perspektiven 8/2006. S. 441ff.

- Paterson, Donald G./Tinker, Miles A. (1929): Studies of typographical factors influencing speed of reading: I. Influence of type form on speed of reading. In: Journal of Applied Psychology 12/1929. S. 359ff.
- Peck, Wendy (2000): Web Menus with Beauty and Brains. New York (NY), Cleveland (OH), Indianapolis (IN)
- Piolat, Annie/Roussey, Jean-Yves/Thunin, Olivier (1997): Effects of screen presentation on text reading and revising. In: International Journal of Human-Computer Studies 47/1997. S. 565ff.
- Pirhonen, Antti/Isomäki, Hannakaisa/Roast, Chris/Saariluoma, Pertti (Hrsg.) (2005): Future Interaction Design. London
- Plata, Werner (1977): A study of newspaper design. Nairobi (Kenia) 1974. Revised and reprinted in Germany 1977
- Poulton, Edward C. (1955): Letter differentiation and rate of comprehension of reading. In: Journal of Applied Psychology 49/1955, S. 358ff.
- Preidel, Michael (2004): Mythen, Monitoraufösungen und das Geheimnis der „72“. <http://www.qxm.de/gestaltung/20040802-084709> (Besuch 07.11.2006)
- Pring, Roger (2000): www.type : Optimale Typografie für das World Wide Web. München
- Pürer, Heinz/Rahofer, Meinrad/Reitan, Claus (Hrsg.) (2004): Praktischer Journalismus. Presse, Radio, Fernsehen, Online. Konstanz
- Puscher, Frank (2001): Das Usability-Prinzip. Wege zur benutzerfreundlichen Website. Heidelberg
- Rager, Günther: Dimensionen der Qualität. Weg aus den allseits offenen Richterskalen? In: Bentele, Günter/Hesse, Kurt R. (Hrsg.) (1994): Publizistik in der Gesellschaft. Festschrift für Manfred Rühl. Konstanz. S. 189ff.
- Rager, Günther (2000): Ethik – eine Dimension von Qualität? In: Schicha, Christian (Hrsg.) (2000): Medienethik zwischen Theorie und Praxis. Normen für die Kommunikationsgesellschaft. Münster. S. 76ff.
- Rager, Günther/Graf-Szczuzuka, Karola/Hassemmer, Gregor/Süper, Stephanie (Hrsg.) (2006): Zeitungsjournalismus. Empirische Leserschaftsforschung. Konstanz
- Rager, Günther/Müller-Gerbes, Sigrun/Haage, Anne (1994): Leserwünsche als Herausforderung. Neue Impulse für die Tageszeitung. Bonn
- Rager, Günther/Werner, Petra (Hrsg.) (1992): Die tägliche Neu-Erscheinung. Untersuchungen zur Zukunft der Zeitung. Münster, Hamburg
- Rahman, Tarjin/Muter, Paul (1999): Designing an interface to optimize reading with small display windows. In: Human Factors 1/1999. S. 106ff.
- Raible, Wolfgang (1991): Die Semiotik der Textgestalt. Erscheinungsformen und Folgen eines kulturellen Evolutionsprozesses. Heidelberg 1991
- Rayner, Keith/Pollatsek, Alexander (1989): The Psychology of Reading. Englewood Cliffs
- Redelius, Jürgen (1998): Der „digitale“ Gutenberg. Dissertation zur Erlangung des Grades eines Doktors der Erziehungswissenschaften der Pädagogischen Hochschule Ludwigsburg. Ludwigsburg
- Rediske, Michael (1994): Kommt TV-Spiegel? In: Die Tageszeitung vom 15.12.1994. S. 11
- Reeps, Inga E. (2004): State-of-the-Art Analyse. Usability, Design und Joy of Use. Konstanz
- Rehe, Rolf F. (1981): Typographie: Wege zur besseren Lesbarkeit. St. Gallen

- Rehe, Rolf F. (1986): Typografie und Design für Zeitungen. Darmstadt
- Reips, Ulf-Dietrich (1999): Theorie und Techniken des Web-Experimentierens. In: Batinic, Bernad/Werner, Andreas/Gräf, Lorenz/Bandilla, Wolfgang (1999): Online Research. Methoden, Anwendungen und Ergebnisse. Göttingen. S. 331ff.
- Rivers, Charlotte (2003): Corporate Design. München
- Rock, Irvin (1987): Wahrnehmung. Stuttgart
- Rommen, Josef (1989): Typographische Maße. In: Aicher, Otl (1989): typographie. Lüdenscheid, Berlin (2)
- Rottermund, Dieter (2002): Webdesign. Leicht, klar, sofort. München
- Ruß-Mohl, Stephan (2003): Journalismus. Das Hand- und Lehrbuch. Frankfurt am Main
- Schäffer, Florian (2001): Webdesign. Design, Navigation, Grafik, Animation. Düsseldorf
- Scharff, Lauren F. V./Ahumada, Albert J. Jr./Hill, Alyson L. (1999): Discriminability Measures for Predicting Readability. In: Proceedings of the SPIE Electronic Imaging Symposium 1999, paper 27
- Scharff, Lauren F. V./Ahumada, Albert J. (2002): Predicting the Readability of Transparent Text. In: Journal of Vision 2/2002. S. 653ff.
- Scharff, Lauren F. V./Ahumada, Albert J. Jr. (2003): Contrast measures for predicting text readability. In: Proceedings of the SPIE Electronic Imaging Symposium 2003. S. 463ff.
- Schicha, Christian (Hrsg.) (2000): Medienethik zwischen Theorie und Praxis. Normen für die Kommunikationsgesellschaft. Münster
- Schmider, Ekkehard (2003): Handbuch für Webtexter. So schreiben Sie fürs Internet. Berlin
- Schneider, Wolf/Raue, Paul-Josef (1996): Handbuch des Journalismus. Reinbek
- Schulz, Ingo A. (2001): Homepage. Kompakt, komplett, kompetent. München
- Schweizer, Peter (2003): Handbuch der Webgestaltung. Eine konzentrierte Einführung in professionelles Webdesign. Bonn
- Seibert, Björn/Hoffmann, Manuela (2006): Professionelles Webdesign mit (X)HTML und CSS. Standardkonformes Webdesign, Accessibility und Usability, Farbe, Grafik und Typografie. Bonn
- Seybold, J.W. (1987): The Desktop-Publishing Phenomenon. Personal Printing has come a long way in a short time. In: Byte 5/1987. S. 149ff.
- Shaikh, A. Dawn (2004): Paper or Pixels: What are People Reading Online? http://psychology.wichita.edu/surl/usabilitynews/62/online_reading_survey.htm (Besuch 07.11.2006)
- Shaikh, A. Dawn (2005): The Effects of Line Length on Reading Online News. <http://psychology.wichita.edu/surl/usabilitynews/72/LineLength.htm> (Besuch 07.11.2006)
- Shaikh, A. Dawn/Chaparro, Barbara S./Fox, Doug (2006): Perception of Fonts: Perceived Personality Traits and Uses. <http://psychology.wichita.edu/surl/usabilitynews/81/Personalityof-Fonts.htm> (Besuch 07.11.2006)
- Shaikh, A. Dawn/Slocum, Jeremy/Zaccagani, Zach (2005): Perceived Personality and Uses of Fonts – We Need You. http://psychology.wichita.edu/surl/usabilitynews/72/font-survey_shorter.htm (Besuch 07.11.2006)
- Siegel, Ralf (2002): Programmierhandbuch ActionScript. Leitfaden für die Programmierung von Flash MX-Anwendungen. München, Wien

- Spona, Helma (2006): CSS-Praxishandbuch. Der Weg zum pflegeleichten Webdesign. München
- Spool, Jared M./Scanlon, Tara/Schroeder, Will/Snyder, Carolyn/DeAngelo, Terri (1999): Web Site Usability. A Designer's Guide. San Francisco
- Sprissler, Hanno (1999): Infografiken gestalten. Berlin, Heidelberg
- SPSS Inc. (2003): SPSS Advanced Models 12.0. Dublin
- Stöckl, Hartmut (2004): Typographie: Gewand und Körper des Textes – Linguistische Überlegungen zu typographischer Gestaltung. In: Zeitschrift für Angewandte Linguistik 41/2004. S. 5–48
- Stocksmeier, Thorsten (2002): Business-Webdesign. Benutzerfreundlichkeit, Konzeptionierung, Technik, Wartung. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio
- Stoessel, Sabine (2002): Methoden des Testings im Usability Engineering. In: Beier, Markus/von Gizycki, Vittoria (Hrsg.) (2002): Usability. Nutzerfreundliches Web-Design. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio. S. 77ff.
- Storm, Rainer (1974): Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle. Leipzig
- Straub, Jürgen/Kempf, Wilhelm/Werbik, Hans (Hrsg.) (1998): Psychologie. Eine Einführung. Grundlagen, Methoden, Perspektiven. München
- Szewczak, Edward/Snodgrass, Coral (Hrsg.) (2002): Human Factors in Information Systems. Hershey, London, Melbourne, Singapore, Beijing
- Tschichold, Jan (1961): Geschichte der Schrift in Bildern. Hamburg (4)
- Thissen, Frank (2001): Screen-Design-Handbuch. Effektiv informieren und kommunizieren mit Multimedia. Berlin, Heidelberg (2)
- Thissen, Frank (2003): Kompendium Screen-Design. Effektiv informieren und kommunizieren mit Multimedia. Berlin, Heidelberg (3)
- Tinker, Miles A. (1963): Legibility of print. Ames
- Tinker, Miles A. (1965): Bases for effective reading. Minneapolis
- Tinker, Miles A./Paterson, Donald G. (1928): Studies of Typographical Factors Influencing Speed of Reading. I. Influence of Type Form on Speed of Reading. In: Journal of Applied Psychology 12/1928. S. 359ff.
- Tinker, Miles A./Paterson, Donald G. (1936): Studies of Typographical Factors Influencing Speed of Reading. XII. Methodological Considerations. In: Journal of Applied Psychology 20/1936. S. 132ff.
- Tolksdorf, Robert (2003): HTML & XHTML – die Sprachen des Web. Informationen aufbereiten und präsentieren im Web. Heidelberg (5)
- Tullis, Thomas S./Boynton, Jennifer L./Hersh, Harry (1995): Readability of Fonts in the Windows Environment. http://sigchi.org/chi95/Electronic/documnts/intpost/tst_bdy.htm (Besuch 07.11.2006)
- Turtschi, Ralf (2000): Mediendesign. Sulgen, Zürich
- Veen, Jeffrey (2001): Webdesign. Konzept, Gestalt, Vision. München
- W3 Schools (2006): Browser Statistics. http://www.w3schools.com/browsers/browsers_stats.asp (Besuch 11.11.2006)
- Wagner, Thomas/Müller, Thomas (2005): Professionelle Webseiten mit web to date 4. Düsseldorf

- Walter, Peter (2000): Verfahren der sequentiellen Merkmalsanalyse für die Mustererkennung. Aachen
- Wandmacher, Jens (1993): Software-Ergonomie. Berlin, New York
- Warkentin, Karl H. (1992): Klassifikation der Schriften nach DIN. In: Karow, Peter (1992): Schrifttechnologie. Methoden und Werkzeuge. Berlin, Heidelberg, New York, London, Paris, Tokyo, Hong Kong, Barcelona, Budapest. S. 343ff.
- Webhits Internet Design GmbH (2006): Web-Barometer. <http://www.webhits.de/webhits/browser.htm> (Besuch 07.11.2006)
- Wehde, Susanne (2000): Typographische Kultur. Eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung. Tübingen
- Weidemann, Kurt (1996): Das Nachbild auf der Netzhaut. Göttingen
- Weidenmann, Bernd (1994): Wissenserwerb mit Bildern. Instruktionale Bilder in Print-Medien, Film-Video und Computerprogrammen. Bern, Göttingen, Toronto, Seattle
- Weiland, Sascha/von Gizycki, Vittoria (2002): Wahrnehmungspsychologische Erkenntnisse im Web-Design. In: Beier, Markus/von Gizycki, Vittoria (Hrsg.) (2002): Usability. Nutzerfreundliches Web-Design. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Tokio
- Weinman, Lynda (2000): WebDesign.2. Tips & Tricks für die Gestaltung professioneller Web-Pages. Zürich
- Weinman, Lynda/Lentz, John Warren (1998): Webdesign der Profis. Webdesign entschlüsselt in Fallstudien. Haar
- Weisenmiller, Eric Michael (1999): A Study of the Readability of On-Screen Text. Unveröffentlichte Dissertation, Virginia Polytechnic Institute and State University. Blackburg (Virginia)
- Welker, Martin/Werner, Andreas, Scholz, Joachim (2005): Online-Research. Markt- und Sozialforschung mit dem Internet. Heidelberg
- Wendt, Dirk (1992): Lesbarkeit von Druckschriften. In: Karow, Peter (1992): Schrifttechnologie. Methoden und Werkzeuge. Berlin, Heidelberg
- Wenz, Christian (2000): My own Site.de. Homepages bauen. München
- Wenz, Christian/Hauser, Tobias (2004): Jetzt lerne ich Webseiten programmieren und gestalten. München
- Wertheimer, Max (1991): Zur Gestaltpsychologie menschlicher Werte: Aufsätze 1934–1940. Hrsg. Von Hans-Jürgen Walter. Opladen
- Wertheimer, Michael (1991): Max Wertheimer – Gestaltprophet. In: Wertheimer, Max (1991): Zur Gestaltpsychologie menschlicher Werte: Aufsätze 1934–1940. Hrsg. Von Hans-Jürgen Walter. Opladen. S. 123f.
- Wetzel, Michael (1991): Die Enden des Buches oder die Wiederkehr der Schrift. Weinheim
- Wieland, Melanie/Spielkamp, Matthias (2003): Schreiben fürs Web. Konstanz
- Wilkinson, Roger T./Robinshaw, Robert M. (1987): Proof-reading: VDU and papertext compared for speed, accuracy and fatigue. In: Behaviour and Information Technology 6/1987. S. 125ff.
- Willberg, Hans Peter/Forssmann, Friedrich (2005): Lesetypografie. Mainz
- Wirth, Thomas (2004): Missing Links. Über gutes Webdesign. München, Wien 2004 (2)

- Wirth, Thomas (2005): Fallstudien über das Textlayout. <http://www.kommdesign.de/texte/aufmerk6.htm> (Besuch 07.11.2006)
- World Wide Web Consortium (W3C) (1999): Web Content Accessibility Guidelines 1.0. W3C Recommendation 5-May-1999. <http://www.w3.org/TR/WAI-WEBCONTENT/> (Besuch 07.11.2006)
- World Wide Web Consortium (W3C) (2006): Cascading Style Sheets, level 2 revision 1. CSS 2.1 Specification. W3C Working 06 November 2006. <http://www.w3.org/TR/CSS21/> (Besuch 07.11.2006)
- Wuttke, Heinrich (1872): Die Entstehung der Schrift, die verschiedenen Schriftsysteme und das Schrifttum der nicht alfabetaisch schreibenden Völker. Leipzig
- Wyss, Vinzenz (2002): Redaktionelles Qualitätsmanagement. Ziele, Normen, Ressourcen. Konstanz
- Youngman, Melissa/Scharff, Lauren (1998): Text Width and Margin Width Influences on Readability of GUI. <http://hubel.sfasu.edu/research/textmargin.html> (Besuch 07.11.2006)
- Zakon, Robert (2005): Hobbes' Internet Timeline v8.2. <http://www.zakon.org/robert/internet/timeline> (Besuch 07.11.2006)
- Zehnder, Matthias W. (1998): Geschichte und Geschichten des Internets. St. Gallen
- Ziefle, Martina (1998): Effects of display resolution on visual performance. In: Human Factors 40/1998. S. 554ff.
- Ziefle, Martina (2002): Lesen am Bildschirm. Eine Analyse visueller Faktoren. Münster
- Zimbardo, Philip G./Gerrig, Richard J. (2004): Psychologie. Bearbeitet und herausgegeben von Ralf Graf, Markus Nagler und Brigitte Ricker. München, Boston, San Francisco, Harlow (England), Don Mills (Ontario), Sydney, Mexiko City, Madrid, Amsterdam (16)
- Zuffo, Dario (1993): Die Grundlagen der visuellen Gestaltung. Hognau

9 Abbildungsverzeichnis

- Abb. 0.1** M.L.
Abb. 1.1 Aus: Frutiger 2001, S. 114
Abb. 1.2 Aus: Zimbardo/Gerrig 2004, S. 176
Abb. 1.3 Aus: Lefrancois 1994, S. 101
Abb. 1.4 Aus: Frutiger 2001, S. 199
Abb. 1.5 Aus: Lindsay/Norman 1981, S. 208
Abb. 1.6 Aus: Neisser 1974, S. 140
Abb. 3.1 M.L.
Abb. 3.2 M.L.
Abb. 3.3 M.L.
Abb. 3.4 M.L.
Abb. 3.5 M.L.
Abb. 3.6 M.L.
Abb. 3.7 M.L.
Abb. 3.8 M.L.
Abb. 3.9 M.L.
Abb. 3.10 M.L.
Abb. 3.11 M.L.
Abb. 3.12 M.L.
Abb. 3.13 M.L.
Abb. 3.14 M.L.
Abb. 3.15 M.L.
Abb. 3.16 M.L.
Abb. 3.17 <http://www.daimlerchrysler.de>
Screenshot 16.08.2006; Bearb.: M.L.
Abb. 3.18 Aus: van Aaken 2006
Abb. 3.19 Aus: Luidl 1989, S. 38
Abb. 3.20 M.L.
Abb. 3.21 M.L.
Abb. 3.22 M.L.
Abb. 3.23 M.L.
Abb. 3.24 M.L.
Abb. 3.25 M.L.
Abb. 3.26 M.L.
Abb. 4.1 M.L.
Abb. 4.2 Screenshot des Experiments
26.01.2006
Abb. 5.1 M.L.
Abb. 5.2 M.L.
Abb. 5.3 M.L.
Abb. 5.4 M.L.
Abb. 5.5 M.L.
Abb. 5.6 M.L.
Abb. 5.7 M.L.
Abb. 5.8 M.L.
Abb. 5.9 M.L.
Abb. 6.1 M.L.
Abb. 6.2 M.L.
Abb. 6.3 M.L.
Abb. 6.4 M.L.
Abb. 6.5 M.L.
Abb. 6.6 M.L.
Abb. 6.7 M.L.
Abb. 6.8 M.L.
Abb. 6.9 M.L.
Abb. 6.10 M.L.
Abb. 6.11 M.L.
Abb. 6.12 M.L.
Abb. 6.13 M.L.
Abb. 6.14 M.L.
Abb. 6.15 M.L.
Abb. 6.16 M.L.
Abb. 6.17 M.L.
Abb. 6.18 M.L.
Abb. 6.19 Screenshot des Experiments
26.01.2006
Abb. 6.20 M.L.
Abb. 6.21 M.L.
Abb. 6.22 M.L.
Abb. 6.23 M.L.
Abb. 6.24 M.L.
Abb. 6.25 M.L.

10 Anhang

10.1 Messergebnisse der 90 Schriftkonstellationen

Die unten stehende Tabelle listet die durchschnittlichen Lesezeiten auf, die im Experiment nach der Datenbereinigung für die 90 getesteten Schriftkonstellationen festgestellt wurden. (\emptyset = mittlere Lesezeit; S = Standardabweichung; n = Anzahl Messwerte; Sa = Schriftart [Ti = *Times*, Ve = *Verdana*]; SG = Schriftgröße; xH = x-Höhe; Za = Zeilenabstand; Br = Zeilenbreite [durchschn. Anschläge pro Zeile])

	\emptyset	S	n	Sa	SG	xH	Za	Br
1.	23,49	5,62	40	Ti	17px	8px	19px	42
2.	23,82	5,87	40	Ve	13px	7px	20px	42
3.	24,37	7,02	40	Ti	17px	8px	21px	42
4.	24,94	5,21	40	Ti	17px	8px	23px	70
5.	25,02	7,31	40	Ve	14px	8px	20px	42
6.	25,03	5,51	40	Ti	15px	7px	17px	42
7.	25,21	6,43	40	Ti	17px	8px	21px	56
8.	25,43	7,05	40	Ti	15px	7px	21px	56
9.	25,49	5,71	40	Ve	13px	7px	20px	56
10.	25,66	5,61	40	Ti	17px	8px	23px	56
11.	25,72	6,34	40	Ti	17px	8px	21px	70
12.	25,73	7,03	40	Ve	14px	8px	22px	42
13.	25,81	6,01	40	Ve	14px	8px	18px	42
14.	25,91	6,46	40	Ti	15px	7px	19px	56
15.	25,95	6,82	40	Ve	13px	7px	20px	84
16.	26,00	8,65	40	Ve	13px	7px	18px	42
17.	26,06	7,13	40	Ti	15px	7px	17px	84
18.	26,14	5,38	40	Ve	13px	7px	16px	56
19.	26,15	6,82	40	Ti	15px	7px	21px	28
20.	26,27	6,23	40	Ve	13px	7px	20px	28
21.	26,31	7,61	40	Ve	14px	8px	18px	56
22.	26,33	7,47	40	Ve	11px	6px	16px	42
23.	26,39	5,52	40	Ti	17px	8px	19px	56
24.	26,39	7,04	40	Ve	13px	7px	20px	70
25.	26,40	7,52	40	Ve	14px	8px	20px	70
26.	26,45	8,64	40	Ve	13px	7px	16px	28
27.	26,48	7,23	40	Ve	14px	8px	22px	84
28.	26,56	7,22	40	Ve	13px	7px	18px	70
29.	26,69	6,71	40	Ti	17px	8px	23px	28
30.	26,70	7,90	40	Ti	13px	6px	19px	42
31.	26,71	6,67	40	Ve	11px	6px	16px	84
32.	26,80	5,26	40	Ve	11px	6px	14px	56
33.	26,80	6,43	40	Ve	13px	7px	16px	84
34.	26,84	5,07	40	Ti	17px	8px	23px	84
35.	26,85	6,49	40	Ve	11px	6px	16px	56
36.	26,87	7,72	40	Ve	14px	8px	18px	70
37.	26,93	7,54	40	Ti	17px	8px	19px	84
38.	27,00	7,14	40	Ve	14px	8px	18px	28
39.	27,03	6,10	40	Ti	13px	6px	19px	70
40.	27,03	7,92	40	Ve	11px	6px	14px	84
41.	27,06	8,52	40	Ve	14px	8px	22px	70
42.	27,09	5,89	40	Ve	13px	7px	18px	28
43.	27,10	6,67	40	Ve	11px	6px	18px	56
44.	27,14	6,33	40	Ti	17px	8px	21px	84
45.	27,15	7,99	40	Ve	11px	6px	18px	42

	\emptyset	S	n	Sa	SG	xH	Za	Br
46.	27,18	7,27	40	Ve	13px	7px	16px	42
47.	27,18	7,46	40	Ve	14px	8px	20px	84
48.	27,20	7,57	40	Ti	17px	8px	19px	28
49.	27,23	6,96	40	Ve	11px	6px	14px	42
50.	27,30	6,71	40	Ti	15px	7px	19px	42
51.	27,31	7,22	40	Ti	13px	6px	17px	70
52.	27,32	6,40	40	Ti	15px	7px	21px	70
53.	27,34	7,40	40	Ve	11px	6px	16px	70
54.	27,43	8,18	40	Ti	15px	7px	17px	70
55.	27,45	7,21	40	Ti	17px	8px	19px	70
56.	27,55	6,51	40	Ti	13px	6px	17px	84
57.	27,58	5,41	40	Ve	13px	7px	18px	56
58.	27,58	6,30	40	Ti	17px	8px	21px	28
59.	27,63	5,02	40	Ti	17px	8px	23px	42
60.	27,65	6,11	40	Ve	11px	6px	16px	28
61.	27,74	7,63	40	Ve	14px	8px	20px	56
62.	27,77	8,97	40	Ve	14px	8px	22px	56
63.	27,82	8,10	40	Ti	13px	6px	17px	42
64.	27,86	6,67	40	Ti	13px	6px	19px	84
65.	28,00	6,99	40	Ve	13px	7px	18px	84
66.	28,16	6,05	40	Ti	13px	6px	15px	28
67.	28,19	7,37	40	Ti	15px	7px	19px	70
68.	28,26	8,50	40	Ti	15px	7px	19px	84
69.	28,31	7,07	40	Ti	13px	6px	17px	56
70.	28,34	7,35	40	Ve	11px	6px	18px	28
71.	28,40	8,61	40	Ve	14px	8px	18px	84
72.	28,48	6,57	40	Ti	15px	7px	19px	28
73.	28,60	6,96	40	Ti	15px	7px	21px	42
74.	28,76	7,45	40	Ti	13px	6px	19px	56
75.	28,91	9,54	40	Ve	11px	6px	18px	84
76.	29,05	7,25	40	Ti	13px	6px	15px	84
77.	29,30	6,78	40	Ti	15px	7px	17px	28
78.	29,31	8,25	40	Ve	14px	8px	22px	28
79.	29,39	7,07	40	Ti	13px	6px	15px	70
80.	29,39	7,80	40	Ve	11px	6px	14px	70
81.	29,55	6,90	40	Ve	11px	6px	14px	28
82.	29,58	8,03	40	Ve	11px	6px	18px	70
83.	29,69	9,08	40	Ve	14px	8px	20px	28
84.	29,78	7,70	40	Ti	13px	6px	17px	28
85.	30,09	6,40	40	Ve	13px	7px	16px	70
86.	30,30	6,03	40	Ti	13px	6px	15px	42
87.	30,57	8,13	40	Ti	15px	7px	17px	56
88.	30,61	6,21	40	Ti	13px	6px	15px	56
89.	30,90	8,09	40	Ti	13px	6px	19px	28
90.	31,23	9,48	40	Ti	15px	7px	21px	84

10.2 Digitale Dokumente

10.2.1 Das Experiment zum Ausprobieren im Web

Das in dieser Arbeit vorgestellte Experiment ist im Web abgelegt und kann nach wie vor durchgeführt werden. Die URL lautet:

http://www.j-pr.de/experiment/experiment_nfv.html

10.2.2 *Flash*- und PHP-Dateien auf der CD-ROM

Auf der beiliegenden CD-ROM findet sich im Ordner „Experiment-Dateien“ die Original-*Flash*-Datei, die Herz und Rückgrat meines Online-Experiments bildete (**experiment_nfv fla**).

Ferner sind in diesem Ordner abgelegt die drei PHP-Dateien, mittels denen ich die Kommunikation zwischen *Flash*-Film und Datenbank herstellte (**p_hol_ff.php**, **p_geb_ff.php** sowie **p_not_ff.php**). Die Serverzugangsdaten sind in diesen Dateien neutralisiert und müssen gegebenenfalls neu platziert werden.

10.2.3 SQL-Dateien auf der CD-ROM

Ebenfalls auf der beiliegenden CD-ROM finden sich SQL-Beschreibungen der beiden Datenbank-Tabellen, in denen die im Experiment erhobenen Messwerte abgebildet wurden. Die wichtigere der beiden Dateien beschreibt die Tabelle mit sämtlichen erhobenen Zeitmess-Daten (**ergebnisse.sql**), die andere Tabelle unterstützte die PHP-Datei **p_hol_ff.php** bei der gleichmäßigen Zuteilung der 90 gemessenen typografischen Konstellationen auf die Probanden (**aufrufkontrolle.sql**).

10.2.4 Diese Arbeit als PDF auf der CD-ROM

Auch die vorliegende Arbeit selbst ist auf der beiliegenden CD-ROM als PDF-Dokument abgelegt (**dissertation_liebig.pdf**).

10.2.5 Bevölkerungsstatistik auf CD-ROM

Die offizielle Bevölkerungsstatistik der Bundesrepublik Deutschland für das Jahr 2005, die mir das Bundesamt für Statistik zur Verfügung stellte, findet sich auf der beiliegenden CD-ROM als *Excel*-Datei (**statistik_BRD.xls**) im Ordner „Statistik BRD“

Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig und ohne unerlaubte Hilfe verfasst habe, dass ich keine anderen als die angegebenen Hilfsmittel und Quellen in Anspruch genommen und Zitate kenntlich gemacht habe. Ferner versichere ich, dass die vorliegende Arbeit nicht bereits auszugsweise oder vollständig als Prüfungsleistung vorgelegen hat.

Hannover, den 13. November 2006

Martin Liebig