



RATTE

Regensburger Analysetool für Texte
Dokumentation

29.09.2020

Johannes Wild
Markus Pissarek

Johannes.Wild@ur.de
Markus.Pissarek@ur.de

Inhalt

Wörter	3
Wörter zählen	3
Wortlänge: Buchstaben	3
Wortlänge: Silben	3
Wörter: Substantivierungen	4
Wörter: Lexikalische Dichte	4
Sätze	6
Sätze zählen	6
Satzlänge (Wörter)	6
Nebensätze	6
Infinitivkonstruktionen	6
Sätze im Passiv	7
Textgrammatische Aspekte	7
Personalpronomen	7
Abkürzungen	7
Type-Token	8
Lesbarkeitsindizes	8
Simple Measure of Gobbledygook – german (gSmog)	8
Lesbarkeitsindex (LIX)	9
Flesch (Amstad)	10
Vierte Wiener Sachtextformel (WSTF)	10
Die Skala ist als Schulstufe zu interpretieren. Sie beginnt bei 4 und endet bei 15.	10
Regensburger Index (RIX)	11
Lesedauer	11

Hinweis: Ratte setzt korrekte Interpunktion der analysierten Texte voraus. Für diejenigen Funktionen, die auf Wortlisten zurückgreifen vgl. Literaturverzeichnis. Das Programm ist online verfügbar unter: <http://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/ratte/index.html>

Wörter

Wörter zählen

Gallmann, Peter: Was ist ein Wort? In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 129-132 (129).

Ausgangspunkt ist das syntaktische Wort: Es werden die Wörter in der genauen Form wie sie im Text erscheinen, d.h. alle Wortformen, bei der Wörterzählung berücksichtigt. Die Worterkennung erfolgt pragmatisch: Alles, was zwischen zwei Leerzeichen bzw. Leerzeichen und Satzzeichen steht („Whitespace-Character“) wird als Wort aufgefasst.

Wortlänge: Buchstaben

Eisenberg, Peter: Phonem und Graphem. In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 19-94 (67f).

Die Wortlänge bezeichnet die durchschnittliche Zahl der Buchstaben in einem Wort (s.o.), d.h. keine Grapheme.

$$\text{Wortlänge} = \frac{\text{Summe Buchstaben aller Wörter}}{\text{Zahl der Wörter im Text}}$$

Abkürzungspunkte wie z.B. in *v.a.* werden ebenfalls als „Buchstaben“ erfasst. Bindestriche werden zuvor entfernt.

Wortlänge: Silben

Eisenberg, Peter: Phonem und Graphem. In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 19-94 (39ff).

Die Erfassung der Silben geht von der prototypischen dt. Silbenstruktur aus, d.h., dass jede Silbe einen Vokal als Silbenkern enthält, der von (höchstens drei) Konsonanten am Anfangs- bzw. Endrand umschlossen ist. Berechnet werden:

- Einsilber: genau eine Silbe.
- Zweisilber: genau zwei Silben.
- Drei+silber: drei oder mehr Silben.

Wörter: Substantivierungen

Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 1): Das Wort. 4. Auflage. Stuttgart, Weimar: Metzler 2013a. (267)

Substantivierungen werden anhand ihrer Großschreibung sowie den Suffixen {-ung}, {-in}, {-ler}, {-schaft}, {-chen}, {-heit}, {-keit}, {-ling}, {-lei} und {-erei} sowie deren Flexionsformen erkannt.

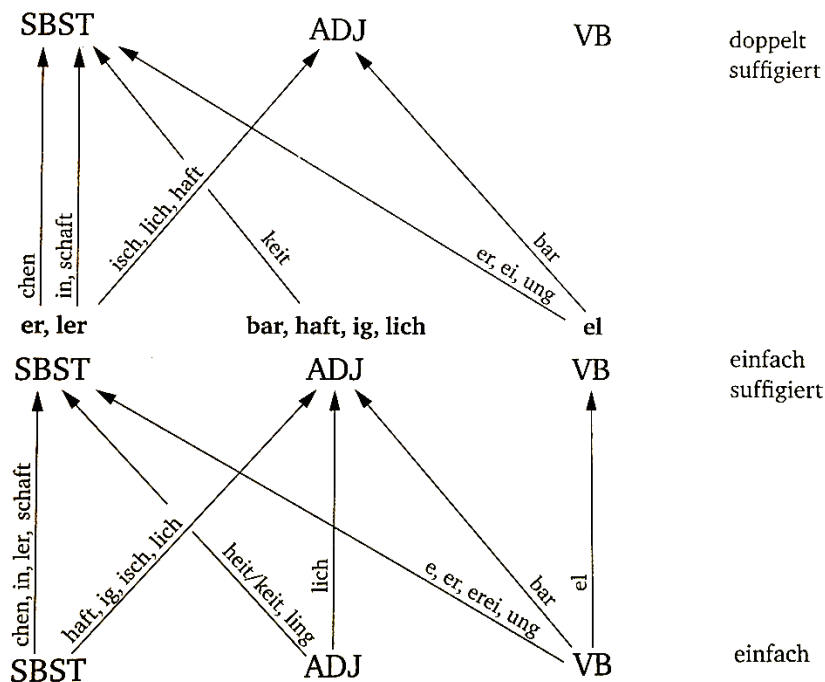


Abb.: Eisenberg 2013, 267.

Wörter: Lexikalische Dichte

Stubbs, Michael: Text and corpus analysis. Oxford: Blackwell 1996. (72ff)

Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 2): Der Satz. 4. Auflage. Stuttgart, Weimar: Metzler 2013b. (53, 150-205)

Die lexikalische Dichte beschreibt das Verhältnis von Inhaltswörtern („noun, adjective, adverb, main verb“ (Stubbs 1996, 72)) zur Gesamtwörterzahl:

$$\text{Lexikalische Dichte} = 100 \cdot \frac{L}{N}$$

wobei N= Zahl der Wörter im Text, L= Lexikalische Wörter (d.h. N-Funktionswörter) (vgl. ebd.)

Als Funktionswörter gibt Stubbs (1996, 72) „auxiliary verb, modal verb, pronoun, preposition, determiner, conjunction“ an sowie deren Flexionsformen. Der Wortliste von RATTE verwendet aus Geschwindigkeitsgründen eine vereinfachte Form der Bestimmung: Das Verhältnis von Substantiven und Verben zur Gesamtwortzahl.

Nach einer Studie von Ure haben geschriebene Texte eine lexikalische Dichte von mehr als 40%, gesprochene liegen in der Regel darunter. Stubbs gibt eine Spannweite von 40-65% bzw. 34-58% an. (Vgl. ebd., 72.) Die lexikalische Dichte kann als Indikator für semantisch verdichtete Texte fungieren. Stubbs nennt für verschiedene Textsorten folgende Richtwerte:

Kategorie	Kategorie bei Stubbs 1996	lex. Dichte (Bsp. dafür bei Stubbs)
Belletristik	fiction	40-54% (art letter 58%)
Sachtexte	non-fiction	40-65% (religious pamphlet 41%)
Monologe (mündlich)	spoken language without feedback	46-64% (horse racing commentary 58%) (political speech 48%)
Dialoge (mündlich)	spoken language where audience is physically present	34-44% (conversation 36%)

Vgl. Stubbs 1996, 73ff.

Sätze

Sätze zählen

Gallmann, Peter: Der Satz. In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 773-1066 (774, 902).

Die Zählung von Sätzen folgt der pragmatischen Duden-Definition: „Ein Satz ist eine abgeschlossene Einheit, die nach den Regeln der Syntax gebildet worden ist.“ (Gallmann 2005, 774) Erfasst werden somit sowohl einfache Sätze als auch Satzgefüge. Da ein Satz eine Texteinheit ist, „die von einem Satzschlusszeichen abgeschlossen wird“ (ebd., 1030), erfolgt die Satzerkennung über eben diese. (Alleinstehende Überschriften ohne Satzschlusszeichen werden daher beispielsweise nicht als eigenständiger Satz aufgefasst.) Als Satzschlusszeichen werden <.>, <!>, <?>, <...> sowie optional <;> und <:> aufgefasst (vgl. ebd., 1073f). Parenthesen durch Klammern oder Gedankenstriche werden aus technischen Gründen nicht erfasst (vgl. dazu auch Zählung der Nebensätze).

Satzlänge (Wörter)

Die durchschnittliche Satzlänge errechnet sich aus der Gesamtzahl der Wörter in einem Text (s.o.) im Verhältnis zur Zahl der Sätze (s.o.) im Text.

Nebensätze

Hoffmann, Ludger: Deutsche Grammatik. Grundlagen für Lehrerbildung, Schule, Deutsch als Zweitsprache und Deutsch als Fremdsprache. 2. Auflage. Berlin: Erich Schmidt 2014. (70)

Als „Sätze mit Nebensätzen“ werden Satzgefüge bezeichnet. Das Programm berechnet die Summe aller Sätze, die Satzgefüge sind. (D.h. es zählt nicht die Zahl der Nebensätze in einem Text.) Analysiert werden Sätze in Hinblick auf das Vorliegen eingeleiteter Nebensätze in Vor- bzw. Nachstellung. Es kann sich hierbei um Objektsätze, Attributsätze und Adverbialsätze handeln (vgl. Hoffmann 2014, 70). Zu den Subjunktionen vgl. zuvor: „Wörter: lexikalische Dichte“. Die Subjunktion *um* wird bei der Analyse ausgespart, da Infinitivkonstruktionen separat erfasst werden.

Infinitivkonstruktionen

Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 2): Der Satz. 4. Auflage. Stuttgart, Weimar: Metzler 2013b. (347f)

Es wird die Zahl der Sätze mit erweitertem Infinitiv berechnet, in denen diese mit Komma abgetrennt sind oder nicht. Erfasst werden sowohl der Infinitiv mit *zu* als „Verbpräfix“ (Eisenberg 2013b, 351) wie in

Er mag es, Sätze zu analysieren.

als auch Infinitive mit *zu* als morphologischer Bestandteil (im Wort) wie in

Er mag es, aufzubrechen.

Sätze im Passiv

Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 2): Der Satz. 4. Auflage. Stuttgart, Weimar: Metzler 2013b. (118)

Es wird die Zahl der Sätze erfasst, die Passivformen enthalten. Die Erkennung beschränkt sich auf das *werden*-Passiv, d.h. Formen von *werden*. Passivfähige Verben (vgl. Eisenberg 2013b, 121) und modaler Infinitiv (vgl. ebd., 124) werden nicht analysiert.

Textgrammatische Aspekte

Personalpronomen

Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 1): Das Wort. 4. Auflage. Stuttgart, Weimar: Metzler 2013a. (18)

Die Analyse von Pro-Formen beschränkt sich auf einen Teil der „echten“ Pronomen, d.h. diejenigen, die selbstständig stehen können (vgl. Eisenberg 2013a, 18), nämlich auf Personalpronomen (*ich, du, er/sie/es, wir, ihr, sie*), Possessivpronomen (*mein, dein, sein/ihr/sein, unser, euer, ihr*), Demonstrativpronomen (*diese(r/s), jene(r/s), der-/die-/dasjenige, der-/die-/dasselbe, solche, der, die, das*), Relativpronomen (*der, die, das, welche(r/s)*), Reflexivpronomen (*mich, dich, sich, uns, euch*) und Indefinitpronomen (*alle, einige, keine, jemand, manche, man, niemand, andere*) und ihre Flexionsformen. Probleme bereitet *sein*, da es sich hier auch um das Hilfsverb handeln kann (dies gilt nicht für die Flexionsformen des PP). Diese Grundform wird daher für die Analyse der PP ausgeklammert.

Abkürzungen

Erfasst werden Abkürzungen mit Großbuchstaben (z.B. *EU, FCKW*) und mit Punkt (z.B. *u.a., z.B.*). Abkürzungen, die aus mehreren Buchstaben bestehen und den Punkt nur am Ende haben (z.B. *usf.*) werden nicht erfasst.

Type-Token

Popescu, Ioan-Iovitz: Word Frequency Studies. Berlin u.a.: de Gruyter 2009. (online verfügbar: <http://evrika-braila.ro/wp-content/uploads/2014/10/Book1-Word-Frequency-Studies-June-2009.pdf>). (249f.)

Die Type-Token-Relation wird zur Messung der Beziehung zwischen der Anzahl der Types (= unterschiedliche Wörter) zu der Anzahl der Tokens (= Gesamtzahl der Wörter im Text) herangezogen. Grundsätzlich bestehen zwei Möglichkeiten, die Zahl der Types in einem Text zu bestimmen: (1) Die Betrachtung verschiedener Wortformen ohne Berücksichtigung des Konzeptes von Lexem/Lemma. Die Types geben hier eher die Formenvielfalt eines Textes an. (2) Die Betrachtung verschiedener Lexeme. Hierzu ist der Abgleich mit Lexemlisten nötig, da Präfixe und Suffixe abgetrennt werden müssen. Dies macht die Datenaufbereitung sehr fehleranfällig und aufwändig. Ratte folgt Konzept 1.

Die Type-Token-Ratio wird in Ratte mit der Köhler-Galle Methode berechnet. Diese versucht, die Textlänge und Textdynamik bei der Berechnung zu berücksichtigen.

$$TTR = \frac{TypesbisPosition100 + TypesGesamttext - \frac{100 * TypesGesamttext}{WörterGesamttext}}{WörterGesamttext}$$

Somit wird der Einfluss der Position eines Types im Text wird bei der Berechnung der TTR berücksichtigt. Ist der Text kürzer als 100 Wörter wird Types/Tokens gerechnet.

Lesbarkeitsindizes

Bamberger, Richard: Erfolgreiche Leseerziehung. Theorie und Praxis. München: Domino 2006. (285f)
Spiegel, Glenn/Campbell, John: Measuring Readability with a Computer: What We Can Learn. Los Angeles 1985 (Url: <http://files.eric.ed.gov/fulltext/ED262386.pdf>, zuletzt geprüft 27.07.2016.)

Lesbarkeitsindizes gehen auf die Idee zurück, dass sprachliche Schwierigkeiten eines Textes auch Indikator für inhaltliche Schwierigkeiten sind. Die Berechnung dieser Indizes beruht auf sprachstatistischen Verfahren und umfasst „alle Merkmale eines Textes, die es einer bestimmten Gruppe von Lesern erleichtern, den Sinn zu verstehen.“ (Bamberger 2006, 285.) Stilistische, semantische Kriterien oder strukturelle Aspekte eines Textes sowie des Layouts/Drucks werden i.d.R. nicht erfasst (vgl. Bamberger 2006, 285). Demnach können auch inhaltlich anspruchsvolle Texte den Lesbarkeitsindex eines einfachen Textes erhalten (z.B. Kafka). Die Lesbarkeit wird erschwert durch Erhöhen der: Zahl der schwierigen Wörter, Variation in der Lexik, durchschnittlichen Satzlänge oder Präpositionalphrasen (vgl. Spiegel/Campbell 1985, 4).

Simple Measure of Gobbledygook – german (gSmog)

Bamberger, Richard/Vanecek, Erich: Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache. Wien: Jugend und Volk 1984. (58f)

Bamberger passte die ursprüngliche Formel von McLaughlin für den deutschsprachigen Raum an. Die Formel setzt die Zahl der mehrsilbigen (drei oder mehr, s.o.) Wörter ins

Verhältnis zur Zahl der Sätze im ganzen Text. Da sich die ursprüngliche Formel auf eine Stichprobe von 30 Sätzen bezieht, muss sich diesbezüglich angepasst werden:

$$gSmog = \sqrt{\frac{\text{Wörter mit drei oder mehr Silben} * 30}{\text{Zahl der Sätze}}} - 2$$

Das Resultat ergibt näherungsweise das Lesealter (in Schulstufen), für das der Text geeignet ist. (Vgl. Bamberger 1984, 58f.)

Lesbarkeitsindex (LIX)

Anderson, Nathan: Analysing the Readability of English and Non-English Texts in the Classroom with Lix.

Darwin 1981. (Url: <http://files.eric.ed.gov/fulltext/ED207022.pdf>, zuletzt geprüft 27.07.2016.) (13)

Bamberger, Richard: Erfolgreiche Leseerziehung. Theorie und Praxis. München: Domino 2006. (286)

Der LIX (Lesbarkeitsindex) berechnet sich aus der durchschnittlichen Satzlänge und dem prozentualen Anteil langer Wörter. Er wurde ursprünglich von dem Schweden Björnsson entwickelt und nimmt Werte zwischen 15 und 80 an.

$$LIX = \frac{\text{Zahl der Wörter}}{\text{Zahl der Sätze}} + 100 * \frac{\text{Zahl der Wörter mit mehr als sechs Buchstaben}}{\text{Zahl der Wörter}}$$

Interpretation der Werte:

	LIX	Jgst. nach Bamberger	Jgst.
sehr leicht	15+	1-2	2
leicht	30+	3-8	3-4
mittel	40+		5-7
schwer	50+		8-9
sehr schwer	60+		10-12

Vgl. Anderson 1981, 13 und Bamberger 2006, 286.

Flesch (Amstad)

Vor der Brück, Tim/Leveling, Johannes: Parameter Learning for a Readability Checking Tool. [unpag.] (Url: http://www.vdb1.de/papers/brueck_leveling07.pdf, zuletzt geprüft 27.07.2016.) (4)

Flesch, Rudolf: How to Write Plain English. [unpag.] (Url:

http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml, zuletzt geprüft 27.07.2016.)

Der Flesch-Index geht von dem Satz als kritische Einheit aus. Längere Sätze erfordern es, mehr Informationen im Gedächtnis zu halten, gleiches gilt für die Decodierung langer Wörter. (Vgl. Flesch 2016.) Der Flesch-Index berechnet sich daher aus der durchschnittlichen Satz- und Wortlänge (vgl. von der Brück/Leveling, 4).

$$\text{Flesch} = 180 - \frac{\text{Zahl der Wörter}}{\text{Zahl der Sätze}} - 58,5 * \frac{\text{Silbenzahl im Text}}{\text{Zahl der Wörter}}$$

Interpretation der Werte:

Flesch	Jgst.	
90.0–100.0	5.	Very easy
80.0– 90.0	6.	Easy
70.0– 80.0	7.	Fairly easy
60.0– 70.0	8.-9.	Plain English
50.0– 60.0	10.-12.	Fairly difficult
30.0– 50.0	Abiturienten	Difficult
0.0– 30.0	Hochschule	Very difficult

Vgl. Flesch 2016.

Vierte Wiener Sachtextformel (WSTF)

Bamberger, Richard: Erfolgreiche Leseerziehung. Theorie und Praxis. München: Domino 2006. (286, 291)

Die (vierte) Wiener Sachtextformel ist ein Index, der durch Regressionsanalysen von „einigen hundert Jugendbüchern“ (Bamberger 2006, 285) gewonnen wurde.

$$\text{WSTF} = 0,2656 * \frac{\text{Zahl der Wörter}}{\text{Zahl der Sätze}} + 0,2744 * \frac{\text{Zahl der Wörter mit mehr als drei Silben}}{\text{Zahl der Wörter}} * 100 - 1,693$$

Die Skala ist als Schulstufe zu interpretieren. Sie beginnt bei 4 und endet bei 15.

Regensburger Index (RIX)

Der Regensburger Index ist ein Metaindex, der über die üblichen Kennwerte hinaus auch Schwierigkeitsparameter wie Passiv, Satzkomplexität und Prädikationen in Form von Substantivierungen miteinbezieht.

$$\text{RIX} = \sqrt{\frac{\text{Zahl der Wörter}}{\text{Zahl der Sätze}} + \frac{\text{Zahl der Wörter mit mehr als sechs Buchstaben}}{\text{Zahl der Wörter}}} + \sqrt{\text{Anteil Sätze im Passiv} + \text{Anteil Sätze mit NS} + \text{Anteil Substantivierungen}} - 0,260$$

Das Resultat ergibt näherungsweise das Lesealter (in Schulstufen), für das der Text geeignet ist.

Lesedauer

Hasbrouk, Jan/Tindal, Gerald: Oral Reading Fluency: 90 Years of Measurement. In: Behavioral research & testing (2005). [unpag.] (Url: <http://files.eric.ed.gov/fulltext/ED531458.pdf>)

Rosebrock, Cornelia: Leseförderung aus systematischer Sicht: Dimensionen von Lesekompetenz und adaptive Förderverfahren. In: Sprechen, Lesen und Schreibenlernen. Hrsg. von Frank Hellmich und Katja Siekmann. Berlin: DGLS 2013. S. 112-134. (131)

Rosebrock, Cornelia/Nix, Daniel: Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung. Baltmannsweiler: Schneider 2014.

Zu einer flüssigen Lektüre gehört das ausreichend schnelle, fehlerfreie Dekodieren von Wörtern. Erst dann kann einem Text Sinn entnommen werden. Die Anforderung an die Lesegeschwindigkeit liegt zwischen 150 Wörtern pro Minute (WpM) bei schwachen Lesern und kann bis zu 300-350 WpM bei geübten Lesern erreichen (vgl. Rosebrock/Nix 2014, 38). Folgende Tabelle basiert auf einer US-amerikanischen Studie (Hasbrouck/Tindal 2005) zur Lesegeschwindigkeit (Lautlesen) von über 15.000 Kindern und zeichnet ein differenziertes Bild.

Lesegeschwindigkeit in WpM für	schwach (25%-Perzentil)	durchschnittlich (50%-Perzentil)	stark (75%-Perzentil)
Ende 1. Klasse	28	53	82
Beginn 2. Klasse	25	51	79
Beginn 3. Klasse	44	71	99
Beginn 4. Klasse	68	94	119
Beginn 5. Klasse	85	110	139
Beginn 6. Klasse	98	127	153
Beginn 7. Klasse	102	128	156
Beginn 8. Klasse	106	133	161
Ende 8. Klasse	124	151	177

Vgl. Rosebrock 2013, 131.

Im Schnitt nimmt die Lesegeschwindigkeit (ohne Berücksichtigung der Leseknicks, rote Markierung) pro Jahrgangsstufe um ca. 20 WpM zu, sodass die durchschnittliche Leseleistung näherungsweise aus der Jahrgangsstufe kalkuliert werden kann. Für das stille Lesen sind im Vergleich zu diesen Werten etwas höhere Werte anzunehmen. Im Umkehrschluss kann mit diesen Werten und der Zahl der Wörter eines Textes die benötigte Lesezeit berechnet werden:

$$\text{Lesezeit} = \frac{\text{Zahl der Wörter}}{47,612 * \ln(\text{Jahrgangsstufe}) + 45,011}$$

Literatur

- Anderson, Nathan: Analysing the Readability of English and Non-English Texts in the Classroom with Lix. Darwin 1981. (Url: <http://files.eric.ed.gov/fulltext/ED207022.pdf>, zuletzt geprüft 27.07.2016.)
- Bamberger, Richard: Erfolgreiche Leseerziehung. Theorie und Praxis. München: Domino 2006.
- Bamberger, Richard/Vanecek, Erich: Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache. Wien: Jugend und Volk 1984.
- Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 1): Das Wort. 4. Auflage. Stuttgart, Weimar: Metzler 2013a.
- Eisenberg, Peter: Grundriss der deutschen Grammatik (Band 2): Der Satz. 4. Auflage. Stuttgart, Weimar: Metzler 2013b.
- Eisenberg, Peter: Phonem und Graphem. In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 19-94.
- Flesch, Rudolf: How to Write Plain English (Url: http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml, zuletzt geprüft 27.07.2016.)
- Gallmann, Peter: Der Satz. In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 773-1066.
- Gallmann, Peter: Was ist ein Wort? In: Duden. Die Grammatik. 7. Auflage. Hrsg. Von der Dudenredaktion. Mannheim, Leipzig, Wien, Zürich: Dudenverlag 2005. S. 129-132.
- Hasbrouk, Jan/Tindal, Gerald: Oral Reading Fluency: 90 Years of Measurement. In: Behavioral research & testing (2005). [unpag.] (Url: <http://files.eric.ed.gov/fulltext/ED531458.pdf>)
- Hoffmann, Ludger: Deutsche Grammatik. Grundlagen für Lehrerbildung, Schule, Deutsch als Zweitsprache und Deutsch als Fremdsprache. 2. Auflage. Berlin: Erich Schmidt 2014.
- Popescu, Ioan-Iovitz: Word Frequency Studies. Berlin u.a.: de Gruyter 2009. (online verfügbar: <http://evrika-braila.ro/wp-content/uploads/2014/10/Book1-Word-Frequency-Studies-June-2009.pdf>).
- Rosebrock, Cornelia: Leseförderung aus systematischer Sicht: Dimensionen von Lesekompetenz und adaptive Förderverfahren. In: Sprechen, Lesen und Schreibenlernen. Hrsg. von Frank Hellmich und Katja Siekmann. Berlin: DGLS 2013. S. 112-134.
- Rosebrock, Cornelia/Nix, Daniel: Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung. Baltmannsweiler: Schneider 2014.
- Spiegel, Glenn/Campbell, John: Measuring Readability with a Computer: What We Can Learn. Los Angeles 1985 (Url: <http://files.eric.ed.gov/fulltext/ED262386.pdf>, zuletzt geprüft 27.07.2016.)
- Stubbs, Michael: Text and corpus analysis. Oxford: Blackwell 1996.
- Vor der Brück, Tim/Leveling, Johannes: Parameter Learning for a Readability Checking Tool. (Url: http://www.vdb1.de/papers/brueck_leveling07.pdf, zuletzt geprüft 27.07.2016.)

Wortlisten

- Institut für Deutsche Sprache (2014): Korpusbasierte Wortformenliste DeReWo, DeReKo-2014-II-MainArchive-STT.100000, <http://www.ids-mannheim.de/derewo>, Institut für Deutsche Sprache, Programmbereich Korpuslinguistik, Mannheim, Deutschland.