Appendix B
Bibliography

Abdi, Hervé, and Lynne J. Williams. 2010. Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4): 433–459. doi:10.1002/wics.101.

Agresti, Alan. 2002. *Categorical Data Analysis*. Wiley Series in Probability and Statistics, 2nd ed. New York: Wiley-Interscience.

Albert, Jim, and Maria Rizzo. 2012. *R by Example*. Use R! New York: Springer.

Allan, Lorraine G. 1980. A Note on Measurement of Contingency Between Two Binary Variables in Judgment Tasks. *Bulletin of the Psychonomic Society* 15 (3): 147–149.

Andor, József. 2004. The Master and His Performance: An Interview with Noam Chomsky. *Intercultural Pragmatics* 1 (1), 93–111. doi:10.1515/iprg.2004.009.

Arnold, Taylor, and Lauren Tilton. 2015. *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text*. Quantitative Methods in the Humanities and Social Sciences. New York: Springer.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, Rolf Harald. 1989. *A Corpus-Based Approach to Morphological Productivity. Statistical Analysis and Psycholinguistic Interpretation*. Amsterdam: Centrum Wiskunde en Informatica.

Baayen, Rolf Harald. 1993. On Frequency, Transparency and Productivity. In *Yearbook of Morphology 1992*, ed. Geert Booij and Jaap van Marle, 181–208. Dordrecht, London: Kluwer.

Baayen, Rolf Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers.

Baayen, Rolf Harald. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In *Corpus Linguistics. An International Handbook*, ed. Anke Lüdeling and Merja Kytö, 899–919. Berlin: Mouton de Gruyter.

Baayen, Rolf Harald. 2011. Corpus Linguistics and Naive Discriminative Learning. *Brazilian Journal of Applied Linguistics* 11: 295–328.

Baayen, Rolf Harald, and Rochelle Lieber. 1991. Productivity and English Derivation: A Corpus-Based Study. *Linguistics* 29: 801–843.

Baroni, Marco, et al. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43 (3): 209–226.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Benzécri, Jean-Paul. 1973. *L'analyse des données*. L'analyse des correspondances, Vol. 2. Paris, Bruxelles, Montréal: Dunod.

Benzécri, Jean-Paul. 1984. *Analyse des correspondances, exposé élémentaire*. Pratique de l'analyse des données, Vol. 1. Paris: Dunod.

Biber, Douglas. 1993. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8 (4): 241–257.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Boas, Hans C., and Sarah Schuchard. 2012. A Corpus-Based Analysis of Preterite Usage in Texas German. In *Proceedings of the 34th Annual Meeting of the Berkeley Linguistics Society*.

Bourdieu, Pierre. 1979. *La distinction: critique sociale du jugement*. Paris: Les Editions de Minuit.

Burns, Patrick. 2011. *The R Inferno*. `http://www.burns-stat.com/pages/Tutor/R_inferno.pdf` (visited on 08/26/2015).

Butts, Carter T. 2008. Network: A Package for Managing Relational Data in R. *Journal of Statistical Software* 24 (2). `http://www.jstatsoft.org/v24/i02/paper`.

Butts, Carter T. 2015. *Network: Classes for Relational Data*. R Package Version 1.13.0. `http://CRAN.R-project.org/package=network`.

Bybee, Joan. 2010. *Language Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan L. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82 (4): 711–733.

Camiz, Sergio, and Gastão Coelho Gomes. 2013. Joint Correspondence Analysis Versus Multiple Correspondence Analysis: A Solution to an Undetected Problem. In *Classification and Data Mining*, ed. Antonio Giusti, Gunter Ritter, and Maurizio Vichi, 11–18. Berlin, Heidelberg: Springer. doi:10.1007/9783642288944_2.

Cheshire, Jenny. 2007. Discourse Variation, Grammaticalisation and Stuff Like That. *Journal of Sociolinguistics* 11 (2): 155–193. ISSN: 1467-9841. doi:1.1111/j.1467-9841.2007.00317x. `http://dx.doi.org/10.1111/j.1467-9841.2007.00317.x`.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1962. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Church, Kenneth, and Patrick Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics* 16 (1): 22–29.

Church, Kenneth, et al. 1991. Using Statistics in Lexical Analysis. In *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, ed. Uri Zernik, 115–164. Hillsdale: Lawrence Erlbaum.

Cohen, Ayala. 1980. On the Graphical Display of the Significant Components in a Two-Way Contingency Table. *Communications in Statistics - Theory and Methods* 9 (10): 1025–1041.

Crawley, Michael. 2012. *The R Book*, 2nd ed. Chichester: Wiley.

Cressie, Noel and Timothy R.C. Read. 1989. Pearson's χ^2 and the Loglikelihood Ratio Statistic G^2: A Comparative Review. *International Statistical Review* 57 (1), 19–43. ISSN: 03067734, 17515823. `http://www.jstor.org/stable/1403582`.

Crowdy, Steve. 1993. Spoken Corpus Design. *Literary and Linguistic Computing* 8 (4): 259–265.

Csárdi, Gábor, and Tamas Nepusz. 2006. The igraph Software Package for Complex Network Research. *InterJournal Complex Systems*: 1695. `http://igraph.org`.

Dalgaard, Peter. 2008. *Introductory Statistics with R*, 2nd ed. New York: Springer.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection*. London: J. Murray.

Davies, Mark. 1859. *Corpus of Canadian English: 50 Million Words, 1920s–2000s*. `http://corpus.byu.edu/can/`.

Davies, Mark. 2008–2012. *The Corpus of Contemporary American English: 450 Million Words, 1990–Present*. `http://corpus.byu.edu/coca/`.

Davies, Mark. 2010. *The Corpus of Historical American English: 400 Million Words, 1810–2009.* http://corpus.byu.edu/coca/.

de Leeuw, Jan, and Patrick Mair. 2009. Simple and Canonical Correspondence Analysis Using the R Package anacor. *Journal of Statistical Software* 31 (5): 1–18. http://www.jstatsoft.org/v31/i05/.

Desagulier, Guillaume. 2014. Visualizing Distances in a Set of Near Synonyms: *Rather, Quite, Fairly,* and *Pretty.* In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy,* ed. Dylan Glynn and Justyna Robinson, 145–178. Amsterdam: John Benjamins.

Desagulier, Guillaume. 2015. Forms and Meanings of Intensification: A Multifactorial Comparison of *Quite* and *Rather. Anglophonia* 20. doi:10.400/anglophonia558. http://anglophonia.revues.org/558.

Desagulier, Guillaume. 2016. A Lesson from Associative Learning: Asymmetry and Productivity in Multiple-Slot Constructions. *Corpus Linguistics and Linguistic Theory* 12 (1): 173–219.

Dik, Simon C. 1978. *Functional Grammar.* Amsterdam, Oxford: North-Holland Publishing Co.

Dik, Simon C. 1997. *The Theory of Functional Grammar,* ed. Kees Hengeveld, 2nd ed. Berlin, New York: Mouton de Gruyter.

Divjak, Dagmar, and Nick Fieller. 2014. Finding Structure in Linguistic Data. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy,* ed. Dylan Glynn and Justyna Robinson, 405–441. Amsterdam: John Benjamins.

Divjak, Dagmar, and Stefan Thomas Gries. 2006. Ways of Trying in Russian: Clustering Behavioral Profiles. *Corpus Linguistics and Linguistic Theory* 2 (1): 23–60.

Divjak, Dagmar, and Stefan Thomas Gries. 2008. Clusters in the Mind?: Converging Evidence from Near Synonymy in Russian. *The Mental Lexicon* 3 (2): 188–213. ISSN: 18711340. http://dx.doi.org/10.1075/ml.3.2.03div.

Dixon, Robert M.W., and Alexandra Y. Aikhenvald. 2004. *Adjective Classes: A Cross-Linguistic Typology.* Oxford: Oxford University Press.

Dunning, Ted. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19 (1): 61–74.

Ellis, Nick. 2006. Language Acquisition as Rational Contingency Learning. *Applied Linguistics* 27 (1), 1–24.

Ellis, Nick, and Fernando Ferreira Junior. 2009. Constructions and Their Acquisition: Islands and the Distinctiveness of Their Occupancy. *Annual Review of Cognitive Linguistics* 7: 187–220.

Everitt, Brian S., et al. 2011. *Cluster Analysis.* Wiley Series in Probability and Statistics, Vol. 5, 330. Oxford: Wiley-Blackwell. ISBN: 9780470749913 (hbk.): 1755.00; 0470749911 (hbk.): 1755.00.

Evert, Stefan. 2005. The Statistics of Word Cooccurrences: Word Pairs and Collocations. PhD Dissertation. Universität Stuttgart. http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/pdf/Evert2005phd.pdf (Visited on 08/14/2015).

Evert, Stefan. 2009. Corpora and Collocations. In *Corpus Linguistics: An International Handbook,* ed. Anke Lüdeling and Merja Kytö, Vol. 2, 1212–1248. Berlin, New York: Mouton de Gruyter.

Evert, Stefan, and Marco Baroni. 2006. The zipfR Library: Words and Other Rare Events in R. Presentation at useR! 2006: The Second R User Conference, Vienna, Austria.

Evert, Stefan, and Marco Baroni. 2007. zipfR: Word Frequency Distributions in R. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics on Interactive Posters and Demonstration Sessions,* 29–32. R Package Version 0.6–6 of 2012-04-03), Prague, Czech Republic.

Ferraresi, Adriano. 2007. Building a Very Large Corpus of English Obtained by Web Crawling: ukWaC. Master's Thesis. University of Bologna.

Firth, John. 1957. A Synopsis of Linguistic Theory 1930–1955. In *Selected Papers of J.R. Firth 1952–1959*, ed. Frank Palmer, 168–205. London: Longman.

Fisher, Ronald Aylmer Sir. 1935. *The Design of Experiments*. Edinburgh, London: Oliver Boyd.

Fox, Robert Allen, and Ewa Jacewicz. 2009. Cross-Dialectal Variation in Formant Dynamics of American English Vowels. *The Journal of the Acoustical Society of America* 126 (5): 2603–2618. doi:10.1121/1.3212921.

Francis, W. Nelson, and Henry Kučera. 1979. *Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English, for Use with Digital Computers*. Department of Linguistics. Brown University. http://www.hit.uib.no/icame/brown/bcm.html (visited on 03/10/2015).

Friendly, Michael. 2000. *Visualizing Categorical Data*. Cary, NC: SAS Publications.

Garside, Roger. 1987. The CLAWS Word-Tagging System. In *The Computational Analysis of English: A Corpus-Based Approach*, ed. Roger Garside, Geoffrey Leech, and Geoffrey Sampson, 30–41. London: Longman.

Geeraerts, Dirk. 2010. The Doctor and the Semantician. In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, ed. Dylan Glynn and Kerstin Fischer, 63–78. Berlin, Boston: Mouton De Gruyter.

Gesmann, Markus, and Diego de Castillo. 2011. Using the Google Visualisation API with R. *The R Journal* 3 (2): 40–44. https://journal.r-project.org/archive/2011-2/RJournal_20112_Gesmann+de~Castillo.pdf.

Ghadessy, Mohsen, Alex Henry, and Robert L. Roseberry. 2001. *Small Corpus Studies and ELT*. Amsterdam: John Benjamins.

Givón, Talmy. 1995. *Functionalism and Grammar*. Amsterdam: John Benjamins.

Glynn, Dylan. 2010. Corpus-Driven Cognitive Semantics: Introduction to the Field. In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 1–42. Berlin: Mouton de Gruyter.

Glynn, Dylan. 2014. Correspondence Analysis: Exploring Data and Identifying Patterns. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Human Cognitive Processing, ed. Dylan Glynn and Justyna A. Robinson, 443–485. Amsterdam: John Benjamins.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2003. Constructions: A New Theoretical Approach to Language. *Trends in Cognitive Sciences* 7 (5): 219–224. http://dx.doi.org/10.1016/S1364-6613(03)00080-9.

Greenacre, Michael J. 2007. *Correspondence Analysis in Practice*. Interdisciplinary Statistics Series, Vol. 2. Boca Raton: Chapman Hall/CRC.

Greenacre, Michael J., and Jorg Blasius. 2006. *Multiple Correspondence Analysis and Related Methods*. Boca Raton: Chapman Hall/CRC.

Greenberg, Joseph H. 1963. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In *Universals of Human Language*, ed. Joseph H. Greenberg, 73–113. Cambridge: MIT Press.

Gries, Stefan Thomas. 2001. A Corpus-Linguistic Analysis of *-ic* and *-ical* Adjectives. *ICAME Journal* 25: 65–108. http://clu.uni.no/icame/ij25/gries.pdf.

Gries, Stefan Thomas. 2009a. *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York, NY: Routledge.

Gries, Stefan Thomas. 2009b. What is Corpus Linguistics? *Language and Linguistics Compass* 3: 1225–1241. doi:10.1111/j.1749818X.2009.00149.x.

Gries, Stefan Thomas. 2010. *Statistics for Linguistics with R: A Practical Introduction*. Berlin, New York: MoutonDe Gruyter. ISBN: 9783110205657.

Gries, Stefan Thomas. 2013. 50-Something Years of Work on Collocations: What is or Should be Next ...*International Journal of Corpus Linguistics* 18 (1), 137–166.

Gries, Stefan Thomas. 2014a. *Coll. Analysis 3.5. A Script for R to Compute Perform Collostructional Analyses*. University of California, Santa Barbara. http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/coll.analysis.r (visited on 07/09/2016).

Gries, Stefan Thomas. 2014b. Frequency Tables: Tests, Effect Sizes, and Explorations. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, ed. Dylan Glynn and Justyna Robinson, 365–389. Amsterdam: John Benjamins.

Gries, Stefan Thomas, and Anatol Stefanowitsch. 2004a. Co-Varying Collexemes in the *Into*-causative. In *Language Culture and Mind*, ed. Michel Achard and Suzanne Kemmer, 225–236. Stanford: CSLI.

Gries, Stefan Thomas, and Anatol Stefanowitsch. 2004b. Extending Collostructional Analysis: A Corpus-Based Perspective on 'Alternations'. *International Journal of Corpus Linguistics* 9(1): 97–129.

Gries, Stefan Thomas, and Anatol Stefanowitsch. 2010. Cluster Analysis and the Identification of Collexeme Classes. In *Empirical and Experimental Methods in Cognitive/Functional Research*, ed. John Newman and Sally Rice, 73–90. Stanford, CA: CSLI.

Haspelmath, Martin. 2011. The Indeterminacy of Word Segmentation and the Nature of Morphology and Syntax. *Folia Linguistica* 45 (1): 31–80.

Hilpert, Martin. 2011. Dynamic Visualizations of Language Change: Motion Charts on the Basis of Bivariate and Multivariate Data from Diachronic Corpora. *International Journal of Corpus Linguistics* 16 (4): 435–461. http://dx.doi.org/10.1075/ijcl.16.4.01hil. http://www.jbe-platform.com/content/journals/10.1075/ijcl.16.4.01hil.

Hilpert, Martin. 2014. Collostructional Analysis: Measuring Associations Between Constructions and Lexical Elements. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, ed. Dylan Glynn and Justyna Robinson, 391–404. Amsterdam: John Benjamins.

Hofland, Knut, and Stig Johansson. 1982. *Word frequencies in British and American English*. Bergen: Norwegian Computing Centre for the Humanities.

Hollmann, Willem B., and Anna Siewierska. 2007. A Construction Grammar Account of Possessive Constructions in Lancashire Dialect : Some Advantages and Challenges. *English Language and Linguistics* 11 (2): 407–424. doi:10.1017/S1360674307002304.

Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Husson, Francois, Sébastien Lê, and Jérôme Pagès. 2011. *Exploratory Multivariate Analysis by Example Using R*. Chapman Hall/CRC Computer Science and Data Analysis. Boca Raton: CRC Press.

Husson, François, Julie Josse, Sébastien Lê, and Jéméry Mazet. 2015. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. R Package Version 1.31.4. http://CRAN.Rproject.org/package=FactoMineR.

Isel, Frédéric, Thomas C. Gunter, and Angela D. Friederici. 2003. "Prosody-Assisted Head-Driven Access to Spoken German Compounds. *Journal of Experimental Psychology* 29 (2): 277–288. doi:10.1037/02787393.29.2.277.

Jockers, Matthew. 2014. *Text Analysis with R for Students of Literature*. New York: Springer.

Johansson, Stig, Geoffrey Leech, and Helen Goodluck. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English. University of Oslo. http://clu.uni.no/icame/manuals/LOB/INDEX.HTM (visited on 03/10/2015).

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. Harlow: Longman.

Kennedy, Graeme. 2003. Amplifier Collocations in the British National Corpus: Implications for English Language Teaching. *TESOL Quarterly* 37 (3): 467–487.

Kilgarriff, Adam. 2001. Comparing Corpora. *International Journal of Corpus Linguistics* 6 (1): 97–133.

Kilgarriff, Adam. 2005. Language is Never Ever Ever Random. *Corpus Linguistics and Linguistic Theory* 1 (2): 263–276. http://dx.doi.org/10.1515/cllt.2005.1.2.263.

Kolaczyk, Eric D., and Gábor Csárdi. 2014. *Statistical Analysis of Network Data with R*. Use R!, xii, 386 New York, London: Springer.

Kotze, Theunis J. van W., and D.V. Gokhale. 1980. A Comparison of the Pearson χ^2 and Log-Likelihood-Ratio Statistics for Small Samples by Means of Probability Ordering. *Journal of Statistical Computation and Simulation* 12 (1), 1–13. doi:10.1080/00949658008810422. eprint: http://dxdoiorg/10.108/00949658008810422.

Labov, William. 1975. Empirical Foundations of Linguistic Theory. In *The Scope of American Linguistics: Papers of the First Golden Anniversary Symposium of the Linguistic Society of America*, ed. Robert Austerlitz, 77–133. Lisse: Peter de Ridder.

Lacheret, Anne, Sylvain Kahane, and Paola Pietrandrea. 2017. *Rhapsodie. A Prosodic-Syntactic Tree-bank for Spoken French*. Amsterdam, Philadelphia: John Benjamins (to appear).

Lafon, Pierre. 1980. Sur la variabilité de la fréquence des formes dans un corpus. *Mots* 1: 127–165.

Lafon, Pierre. 1981. Analyse lexicométrique et recherche des cooccurrences. *Mots*: 95–148. ISSN: 0243-6450. http://wwwperseefr/web/revues/home/prescript/article/mots_0243-6450_1981_num_3_1_1041.

Lafon, Pierre. 1984. *Dépouillements et statistiques en lexicométrie*. Travaux de linguistique quantitative. Genève, Paris: Slatkine, Champion.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*, Vol. 1. Stanford: Stanford University Press.

Langacker, Ronald W. 1988. A Usage-Based Model. In *Topics in Cognitive Linguistics*, ed. Brygida Rudzka-Ostyn, 127–161. Amsterdam, Philadelphia: John Benjamins.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar: Descriptive Application*, Vol. 2. Stanford: Stanford University Press.

Langacker, Ronald W. 2008. *Cognitive Grammar : A Basic Introduction*. Oxford: Oxford University Press.

Le Roux, Brigitte. 2010. *Multiple Correspondence Analysis*. London: SAGE.

Lebart, Ludovic, André Salem, and Lisette Berry. 1998. *Exploring Textual Data*. Text, Speech and Language Technology. Dordrecht, Boston, London: Kluwer Academic Publishers.

Leech, Geoffrey. 2007. New Resources, or Just Better Old Ones? The Holy Grail of Representativeness. In *Corpus Linguistics and the Web*, ed. Marianne Hundt, Nadja Nesselhauf, and Carolin Biewer, 133–149. Amsterdam: Rodopi.

Leech, Goeffrey, and Roger Fallon. 1992. Computer Corpora – What do They Tell Us About Culture? In *ICAME*, Vol. 16, 29–50.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, London: University of Chicago Press.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics : Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge, New York: Cambridge University Press.

McEnery, Tony, and Richard Xiao. 2005. Character Encoding in Corpus Construction. In *Developing Linguistic Corpora: A Guide to Good Practice*, ed. Martin Wynne, 47–58. Oxford: Oxbow Books.

Meyer, Charles F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Mikolov Tomas, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. In *Proceedings of NAACL-HLT*, 746–751. http://www.aclweb.org/anthology/N/N13/N13-1090.pdf.

Mikolov, Tomas et al. 2013a. Distributed Representations of Words and Phrases and Their Compositionality. CoRR abs/1310.4546. http://arxiv.org/abs/1310.4546.

Mikolov, Tomas et al. 2013b. Efficient Estimation of Word Representations in Vector Space. CoRR abs/1301.3781. http://arxiv.org/abs/1301.3781.

Muller, Charles. 1964. *Essai de statistique lexicale. L'Illusion Comique de Pierre Corneille*. Paris: Klincksieck.

Muller Charles. 1973. *Initiation aux méthodes de la statistique linguistique*. Paris: Champion.

Muller Charles. 1977. *Principes et méthodes de statistique lexicale*. Paris: Hachette.

Nenadic, Oleg, and Michael Greenacre. 2007. Correspondence Analysis in R, with Two- and Three-Dimensional Graphics: The ca Package. *Journal of Statistical Software* 20 (3): 1–13. http://www.jstatsoft.org.

Oakes, Michael P. 2009. Corpus Linguistics and Language Variation. In *Contemporary Corpus Linguistics*, ed. Paul Baker, 161–185. London: Bloomsbury Continuum.

Paradis, Carita. 1997. *Degree Modifiers of Adjectives in Spoken British English*. Lund: Lund University Press.

Pavlov, Ivan Petrovich. 1927. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. London: Oxford University Press/Humphrey Milford.

Pearson, Karl. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2 (11), 559–572. doi:10.1080/14786440109462720.

Pecina, Pavel. 2010. Lexical Association Measures and Collocation Extraction. *Language Resources and Evaluation* 44 (1): 137–158.

Pedersen, Ted. 1996. Fishing for Exactness. In *Proceedings of the South-Central SAS Users Group Conference*, 188–200. San Antonio, TX: SAS Users Group.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 1532–1543. http://www.aclweb.org/anthology/D14-1162.

Plag, Ingo. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.

Read, Timothy R.C., and Noel A.C. Cressie. 1988. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer Series in Statistics. New York: Springer-Verlag.

Rescorla, Robert A. 1968. Probability of Shock in the Presence and Absence of CS in Fear Conditioning. *Journal of Comparative and Physiological Psychology* 66: 1–5.

Robinson, David. 2016. *gutenbergr: Download and Process Public Domain Works from Project Gutenberg*. R Package Version 0.1.2. https://CRAN.Rproject.org/package=gutenbergr

Sampson, Geoffrey. 2001. *Empirical Linguistics*. London: Continuum.

Schmid, Hans-Jörg. 2010. Does Frequency in Text Really Instantiate Entrenchment in the Cognitive System? In *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, ed. Dylan Glynn and Kerstin Fischer, 101–133. Berlin, New York: Mouton de Gruyter.

Schmid, Helmut. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *International Conference on New Methods in Language Processing*, Manchester, UK, 44–49. http://www.cis.unimuenchen.de/~schmid/tools/TreeTagger/data/treetagger1.pdf.

Sinclair, John. 1966. Beginning the Study of Lexis. In *In Memory of J.R. Firth*, ed. C.E. Bazell, et al., 410–431. London: Longman.

Sinclair, John. 1987. Collocation: A Progress Report. In *Language Topics: Essays in Honour of Michael Halliday*, ed. R. Steele and T. Threadgold, Vol. 2, 19–331. Amsterdam: John Benjamins.

Sinclair, John. 1991. *Corpus, Concordance Collocation*. Oxford: Oxford University Press.

Sinclair, John, and Ronald Carter. 2004. *Trust the Text: Language Corpus and Discourse*. London: Routledge.

Smith, Nicholas. 1997. Improving a Tagger. In *Corpus Annotation: Linguistic Information from Computer Text Corpora*, ed. Roger Garside, Geoffrey N. Leech, and Tony McEnery, 137–150. London: Longman.

Stefanowitsch, Anatol, and Stefan Thomas Gries. 2003. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics* 8 (2): 209–243.

Stefanowitsch, Anatol, and Stefan Thomas Gries. 2005. Covarying Collexemes. *Corpus Linguistics and Linguistic Theory* 1 (1), 1–46.

Stubbs, Michael. 2001. Words and Phrases: Corpus Studies of Lexical Semantics. Oxford: Wiley–Blackwell.

Tagliamonte, Sali, and Rachel Hudson. 1999. *Be like* et al. Beyond America: The Quotative System in British and Canadian Youth. *Journal of Sociolinguistics* 3 (2): 147–172. doi:10.1111/1467-9481.00070

Tang, Naping, et al. 2009. Coffee Consumption and Risk of Lung Cancer: A Meta-Analysis. *Lung Cancer* 67 (1), 17–22. doi:1.1016/j.lungcan.2009.03.012. http://dx.doi.org/10.1016/j.lungcan.2009.03.012

Taylor, John R. 2012. *The Mental Corpus: How Language is Represented in the Mind*. Oxford: Oxford University Press.

Teubert, Wolfgang. 2005. My Version of Corpus Linguistics. *International Journal of Corpus Linguistics* 10 (1), 1–13. doi:10.1075/ijcl.10.1.01teu.

The Bank of English (1980s–). http://www.titania.bham.ac.uk/. Jointly owned by Harper-Collins Publishers and the University of Birmingham.

The British National Corpus. 2007. *BNC XML Edition*. Version 3. http://www.natcorp.ox.ac.uk/. Distributed by Oxford University Computing Services on behalf of the BNC Consortium.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Toutanova, Kristina, et al. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLTNAACL 2003*, 252–259.

van der Maaten, Laurens, and Geoffrey Hinton. 2008. Visualizing Data Using t-SNE. *Journal of Machine Learning Research* 9: 2579–2605.

Wagner, Allan R., and Robert A. Rescorla. 1972. A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In *Classical Conditioning II*, ed. Abraham H. Black, and William F. Prokasy, 64–99. New York: Appleton-Century-Crofts.

Ward, Joe H. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58 (301), 236. ISSN:01621459. http://dx.doi.org/10.2307/2282967.

Wasow, Thomas, and Jennifer Arnold. 2005. Intuitions in Linguistic Argumentation. *Lingua* 115 (11), 1481–1496. http://dx.doi.org/10.1016/j.lingua.2004.07.001.

Wiechmann, Daniel. 2008. On the Computation of Collostruction Strength: Testing Measures of Association as Expressions of Lexical Bias. *Corpus Linguistics and Linguistic Theory* 4 (2): 253.

Yates, Frank. 1984. Tests of Significance for 2×2 Contingency Tables. *Journal of the Royal Statistical Society Series A (General)*, 147 (3): 426–463.

Zeileis, Achim, David Meyer, and Kurt Hornik. 2007. Residual-Based Shadings for Visualizing (Conditional) Independence. *Journal of Computational and Graphical Statistics* 16 (3), 507–525. http://statmathwuacat/~zeileis/papers/Zeileis+Meyer+Hornik-2007.pdf.

Zeldes, Amir. 2012. *Productivity in Argument Selection: From Morphology to Syntax*. Berlin, New York: Mouton de Gruyter.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley.