Counting Words: References

H. Baayen (1989). *A corpus based approach to morphological productivity.* Ph.D. thesis, Vrije Universiteit de Amsterdam.

H. Baayen (1992). Quantitative aspects of morphological productivity. *Yearbook of Morphology 1991*, 109-150.

H. Baayen (2001). *Word frequency distributions.* Kluwer Academic Publishers, Dordrecht.

H. Baayen (2003). Probabilistic approaches to morphology. In R. Bod, J. Hay and S. Jannedy (eds.) *Probabilistic linguistics.* MIT Press, Cambridge MA, 229-287.

H. Baayen and R. Lieber. Productivity and English derivation: A corpus-based study. *Linguistics*, **29**, 801-843.

M. Baroni (to appear). Distributions in text. In A. Lüdeling and M. Kytö (eds.) *Corpus linguistics: An international handbook.* Mouton de Gruyter, Berlin.

L. Bauer (2001). *Morphological productivity.* Cambridge University Press, Cambridge.

Creutz, M. (2003). Unsupervised segmentation of words using prior distributions of morph length and frequency. *Proceedings of ACL 03*, 280-287.

P. Dalgaard (2002). *Introductory statistics with R.* Springer, New York.

S. Evert (2004a). A simple LNRE model for random character sequences. *Proceedings of JADT 2004*, 411-422.

S. Evert (2004b). *The statistics of word cooccurrences: Word pairs and collocations.* Ph.D. thesis, Institut für Maschinelle Sprachverarbeitung, University of Stuttgart.

S. Evert and M. Baroni (2006a). Testing the extrapolation quality of word frequency models. *Proceedings of Corpus Linguistics 2005*.

S Evert and M. Baroni (2006b). ZipfR: Working with words and other rare events in R. *R User Conference 2006*.

S. Evert and A. Lüdeling (2001). Measuring morphological productivity: Is automatic preprocessing sufficient? *Proceedings of Corpus Linguistics 2001*, 167-175.

L. Gaeta and D. Ricca (2003). Italian prefixes and productivity: A quantitative approach. *Acta Linguistica Hungarica*, **50**, 89-108.

L. Gaeta and D. Ricca (to appear). Productivity in Italian word formation: A variable-corpus approach. *Linguistics*.

P. Garrard, L. Maloney, J. Hodges and K. Patterson (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, **128(2)**, 250-260.

I. Good (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40(3/4)**, 237-264.

L. Ha, E. Sicilia-Garcia, J. Ming and F. Smith (2002). Extension of Zipf's law to words and phrases. *Proceedings of COLING 2002*, 315-320.

L. Ha, D. Stewart, P. Hanna and F. Smith (2006). Zipf and type-token rules for the English, Spanish, Irish and Latin languages. *Web Journal of Formal, Computational & Cognitive Linguistics*, **8**.

K. Kageura (1998). A statistical analysis of morphemes in Japanese terminology. *Proceedings of COLING-ACL 98*, 638-645.

E. Khmaladze (1987). The statistical analysis of large number of rare events. Technical Report MS-R8804, Department of Mathematical Statistics, CWI, Amsterdam, Netherlands.

A. Kilgarriff (2005). Language is never ever ever random. *Corpus Linguistics and Linguistic Theory* **1**, 263-276.

R. Köhler, G. Altmann and R. Piotrowski (eds.) (2005). *Quantitative linguistics: An international handbook*. Mouton de Gruyter, Berlin.

A. Kornai (2002). How many words are there? *Glottometrics*, **4**, 61-86.

G. Landini and R. Zandbergen (1998). A well-kept secret of mediaeval science: The Voynich manuscript. *Aesculapius 1998*.

W. Li (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, **38**, 1842-1845.

W. Li (2002). Zipf's law everywhere. *Glottometrics*, **5**, 14-21.

A. Lüdeling and S. Evert (2003). Linguistic experience and productivity: Corpus evidence for fine-grained distinctions. *Proceedings of Corpus Linguistics*

2003, 475-483.

A. Lüdeling and S. Evert (2005). The emergence of productive non-medical *-itis*: Corpus evidence and qualitative analysis. In S. Kepser and M. Reis (eds.) *Linguistic evidence: Empirical, theoretical, and computational perspectives*. Mouton de Gruyter, Berlin, 351-370.

A. Lüdeling, S. Evert and U. Heid (2000). On measuring morphological productivity. *Proceedings of KONVENS 2000*, 57-61.

B. Mandelbrot (1953). An informational theory of the statistical structure of languages. In W. Jackson (ed.) *Communication theory*. Butterworth, London, 486-502.

C. Manning and H. Schütze (1999). *Foundations of statistical natural language processing*. MIT Press, Cambridge MA.

G. Miller (1957). Some effects of intermittent silence. *American Journal of Psychology*, **52**, 311-314.

B. Möbius (2003). Rare events and closed domains: Two delicate concepts in speech synthesis. *International Journal of Speech Technology*, **6**, 57-71.

Ch. Muller (1977). *Principes et méthodes de statistique lexicale*. Hachette, Paris.

S. Naranan and V. Balasubrahmanyan (1998). Models for power law relations in linguistics and information science. *Journal of Quantitative Linguistics*, **5**, 35-61.

I. Plag, C. Dalton-Puffer and H. Baayen (1999). Morphological productivity across speech and writing. *English Language and Linguistics*, **3**, 209-228.

A. Rouault (1978). Lois de Zipf et sources markoviennes. *Annales de l'Institut H. Poincaré (B)*, **14**, 169-188.

H. Schultink (1961). Produktiviteit als morphologish phenomeen. *Forum der Letteren*, **2**, 110-125.

H. Sichel (1971). On a family of discrete distributions particularly suited to represent long-tailed frequency data. *Proceedings of the Third Symposium on Mathematical Statistics*, 51-97.

H. Sichel (1975). On a distribution law for word frequencies. *Journal of the American Statistical Association*, **70**, 542-547.

H. Simon (1955), On a class of skew distribution functions. *Biometrika*, **42**, 425-440.

R. Thisted and B. Efron (1987). Did Shakespeare write a newly-discovered poem? *Biometrika*, **74(3)**, 445-455.

F. Tweedie and H. Baayen (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, **32**, 323-352.

G.K. Zipf (1949). *Human behavior and the principle of least effort.* Addison-Wesley, Cambridge MA.

G.K. Zipf (1965). *The psycho-biology of language.* MIT Press, Cambridge MA.