

## Morphology and Corpora: References

- A. Albright and B. Hayes (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* **90**, 119-161.
- M. Aronoff (1976). *Word formation in generative grammar*. MIT Press, Cambridge MA.
- 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. Banko and E. Brill (2001) Scaling to very very large corpora for natural language disambiguation. *Proceedings of ACL-01*
- M. Baroni (2003). Distribution-driven morpheme discovery: A computational / experimental study. *Yearbook of Morphology 2003*, 213-248.
- M. Baroni (2005). Misurare la produttività: Esperimenti e riflessioni. Talk given at the MorBo Seminar, Dipartimento di Lingue, University of Bologna, Bologna, Italy.
- M. Baroni (to appear). Distributions in text. In A. Lüdeling and M. Kytö (eds.), *Corpus linguistics: An international handbook*, Mouton de Gruyter, Berlin.
- M. Baroni and A. Kilgariff (2006). Large linguistically-processed Web corpora for multiple languages. *Proceedings of EAACL 2006*.
- M. Baroni, J. Matiassek and H. Trost (2002a). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In Maxwell (2002), 48-57.
- M. Baroni and M. Ueyama (2006). Building general- and special-purpose corpora by Web crawling. *Proceedings of the 13th NIJL International Symposium*,

*Language Corpora: Their Compilation and Application*, 31-40.

M. Baroni and S. Vegnaduzzo (2003). Assessing morphological productivity via automated measures of semantic transparency. *DGFS Workshop on Explaining Productivity*.

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

G. Booij (1977). *Dutch morphology: A study of word formation in generative grammar*. De Ridder, Lisse.

M. Brent, S. Murthy and A. Lundberg (1995). Discovering morphemic suffixes: A case study in minimum description length induction. *Fifth International Workshop on AI and Statistics*.

M. Creutz and K. Lagus (2002). Unsupervised discovery of morphemes. In Maxwell (2002), 21-30.

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

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

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

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*.

J. Goldsmith (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics* **27**, 153-198.

J. Goldsmith (to appear). An algorithm for the unsupervised learning of morphology . To appear in *Natural Language Engineering*.

Z. Harris (1955). From phoneme to morpheme. *Language* **31**, 190-222.

- J. Hay (2003). *Causes and consequences of word structure*. Routledge, New York.
- J. Hay and H. Baayen (2002). Parsing and productivity. *Yearbook of Morphology 2001*, 203-235.
- J. Hay and H. Baayen (2004). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, **15**, 99-130.
- A. Kilgarriff (2005). Language is never ever ever random. *Corpus Linguistics and Linguistic Theory* **1**, 263-276.
- A. Kilgarriff, P. Rychly, P. Smrz and D. Tugwell (2004). The Sketch Engine. *Proceedings of EURALEX 2004*, 105-116
- R. Lieber (2004). *Morphology and lexical semantics*. Cambridge University Press, Cambridge.
- D. Lin (1998). Automatic retrieval and clustering of similar words. *Proceedings of COLING-ACL98*, 768-774.
- 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*, pages 351-370. Mouton de Gruyter.
- A. Lüdeling, S. Evert and U. Heid, Ulrich (2000). On measuring morphological productivity. *Proceedings of KONVENS 2000*. 57-61.
- C. Mair (2003) Tracking ongoing grammatical change and recent diversification in present-day standard English: The complementary role of small and large corpora. *ICAME Conference*.
- C. Manning and H. Schütze (1999). *Foundations of statistical natural language processing*. MIT Press, Cambridge MA.
- M. Maxwell (ed.) (2002). *Morphological and phonological learning: Proceedings of the sixth ACL-SIGPHON meeting*. ACL, Philadelphia.
- J. Meibauer, Jörg, A. Guttropf and C. Scherer (2004). Dynamic aspects of German *-er*-nominals: a probe into the interrelation of language change and language acquisition. *Linguistics*, **42**, 155-193.
- E. Nishimoto (2004). *A corpus-based delimitation of new words: cross-segment comparison and morphological productivity*. Ph.D. thesis, City University of New York.

- H. Paul (1920). *Deutsche Grammatik. Band V: Wortbildungslehre*. Max Niemeyer Verlag, Halle a.S.
- V. Pirrelli, B. Calderone, I. Herreros and M. Virgilio M (2003). Non-locality all the way through: Emergent global constraints in the Italian morphological lexicon. *Quaderni del Laboratorio di Linguistica* **4**, 227-251.
- I. Plag (1999). *Morphological productivity: Structural constraints in English derivation*. Mouton de Gruyter, Berlin.
- I. Plag, C. Dalton-Puffer and H. Baayen (1999). Morphological productivity across speech and writing. *English Language and Linguistics*, **3**, 209-228.
- C. Scherer (2005). *Wortbildungswandel und Produktivität: Eine empirische Studie zur nominalen -er-Derivation im Deutschen*. Niemeyer, Tübingen.
- P. Schone, D. Jurafsky (2001). Knowledge-free induction of inflectional morphologies. *Proceedings of NAACL-2001*.
- H. Schultink (1961). Produktiviteit als morphologisch fenomenen. *Forum der Letteren*, **2**, 110-125.
- M. Snover and M. Brent (2001). A Bayesian model for morpheme and paradigm identification. *Proceedings of ACL-2001*, 482-490.
- A. Stefanowitsch and S. Gries (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory* **1**, 1-43.
- M. Tomasello (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press, Harvard.
- 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.
- D. Yarowsky and R. Wicentowski (2000). Minimally supervised morphological analysis by multimodal alignment. *Proceedings of ACL-2000*, 207-216.