

Measuring changes in the productivity of derivational affixes

Claire Cowie, Technische Universität Chemnitz

This paper is concerned with certain methodological issues and problems in the measurement of the productivity of, or the number of new words generated by, a word-formation process over time. These will be illustrated through a study of the productivity of Latin and native nominalizing affixes (the deadjectival nominalizing affixes *-ness* and *-ity* and the deverbal nominalizing affix *-tion*) in Early and Late Modern English in a historical corpus of English, namely the ARCHER corpus (Biber, Finegan and Atkinson 1994, Biber and Finegan 1997).

Productivity in morphological theory is typically construed as a value attached to a word-formation rule and is thus synchronic (see for instance Aronoff 1976, Bauer 1983, 1988). Diachronic productivity might then be expected to involve a comparison of those values at different points in time. Discussion of the optimal methods of measuring productivity over time has however been limited. Aronoff (1980) has suggested that diachronic productivity can be calculated by measuring the number of new types produced by a WFR from T1 to T2.

I contrast the use of a historical corpus with the use of a historical dictionary such as the OED in the measurement of productivity over time. The latter approach is represented by Cannon (1987), Anshen and Aronoff (1989), and Aronoff and Anshen (1998). I argue that it is preferable to use historical corpora, given the unreliability of historical dictionaries in terms of first citation dates, and adequate representation of different registers and time periods. I demonstrate how productivity has been measured over time using historical or diachronic corpora in Dalton-Puffer (1996), Baayen and Renouf (1996), and the present study.

Diachronic productivity can be measured by comparing the difference in type frequency over time, or over different subperiods of a corpus, as in Dalton-Puffer (1996). With a historical corpus it emerges in practice that the values at T1 and T2 (the "points" in time) are in fact determined by counting types over a time stretch, or a subperiod of the corpus. I assess whether type frequency is the most appropriate value for a comparison of two points in time or time periods, given other methods that have been developed for obtaining a synchronic value for the productivity of a WFR (Baayen 1992, 1993, Baayen and Lieber 1991).

It is also possible, instead of comparing the type frequency of subperiods, to plot only new types over time, i.e. types that have not occurred in the corpus previously. Baayen and Renouf (1996) plot new types over time for an 80 million word corpus of four years of contemporary newspaper English. To do this for historical corpora however is to risk certain distortions, given that they are so much smaller than their Present Day English counterparts. For small corpora the likelihood of capturing existing types in the first or first few subperiods of the corpus is reduced, making it more difficult to ascertain whether types in later periods are likely to be "new" in the language. I investigate ways of treating this problem by using a "starting lexicon". In this study I adopt the first two periods of the Early Modern section of the Helsinki corpus (1500-1640) as a starting lexicon, given that ARCHER runs from 1650.

I discuss two problems of normalization, necessary when subperiods are of different sizes, as they often are in historical corpora. Firstly, normalization involves counting types out of tokens, which is not counting like out of like. Secondly, statistically, a larger sample size (i.e. a larger subperiod, or subcorpus) may capture more rare types. Finally, I examine methods of comparing the productivity of a WFR over time for different subcorpora, in this study, the range of registers in ARCHER (and spoken and written registers in Plag, Dalton-Puffer and Baayen, in press). I suggest that it may be problematic to measure the increase in type frequency or new type frequency for each subcorpus independently, as this involves treating each subcorpus as having an independent lexicon, as opposed to measuring the new words generated against the lexicon as a whole, represented by the entire corpus.

References

- Anshen, F. and M. Aronoff (1989). Morphological Productivity, Word Frequency and the OED. *Language Change and Variation*. eds. R. Fasold and D. Schrifin. Amsterdam/Philadelphia: John Benjamins.
- Aronoff, M. (1980). The relevance of productivity in a synchronic theory of word formation. *Historical morphology*. ed. J. Fisiak. The Hague: Mouton. 71-83.
- Aronoff, M. (1976). *Word Formation and Generative Grammar*. Massachusetts: MIT Press.
- Aronoff, M. and F. Anshen (1998). Morphology and the Lexicon: Lexicalization and Productivity. *Handbook of Morphology*. eds. A. Spencer and A.M. Zwicky. Oxford: Blackwell. 237-248.
- Baayen, R.H. On frequency, transparency and productivity. *Yearbook of Morphology 1992*. eds. G. Booij and J. van Marle. Dordrecht: Kluwer. 181-208
- Baayen, R.H. and R. Lieber (1991). Productivity and English derivation: a corpus-based study. *Linguistics* 29: 801-843.
- Baayen, R.H. and A. Renouf (1996). Chronicling The Times: Productive Lexical Innovations in an English Newspaper. *Language* 72 (1): 69-76.
- Bauer, L. (1983). *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, L. (1988). *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.
- Biber, D. and E. Finegan (1997). Diachronic relations among Speech-based and Written Registers in English. *To Explain the Present: Studies in the changing English language in honour of Matti Rissanen*. eds. T. Nevalainen and L. Kahlas-Tarkka. Helsinki: Soci ete Neophilologique. 253-276.
- Biber, D., E. Finegan and D. Atkinson (1994). ARCHER and its challenges: Compiling and exploring a representative corpus of historical English registers. *Creating and Using English Language Corpora* eds. U. Fries., G. Tottie and P. Schneider. Amsterdam/Atlanta: Rodopi.
- Cannon, G. (1987) *Historical Change and English Word Formation*. New York and Bern: Lang.
- Dalton-Puffer, C. (1996) *The French Influence on Middle English Morphology: a Corpus-based study of Derivation*. Berlin: Mouton de Gruyter.
- Plag, I., C. Dalton-Puffer and R. H. Baayen (in press). Morphological Productivity across speech and writing. *English Language and Linguistics*.