Need and competition in word formation and where to find data to study them

Anke Lüdeling, Marco Baroni, Stefan Evert

Humboldt-University Berlin, University of Bologna, University of Osnabrück

Anke.luedeling@rz.hu-berlin.de, baroni@sslmit.unibo.it, stefan.evert@uos.de

Morphological theory is concerned not only with the structure of existing complex words but also with the statement of the 'availability' (Bauer 2001) of morphological processes to form new complex words. While in generative approaches a rule is either available or not (qualitative productivity; Dressler 2003), there are many researchers in morphology that focus on the quantitative potential of a morphological process (Bauer calls this the profitability of a rule; we will speak of quantitative productivity). There are different aspects of this potential, e.g. how many complex words have been produced by a given process P, how likely it is that P will produce more words in the future, how the potential of a rule changes over time etc. A number of different measures and procedures to calculate these have been proposed (Baayen 2001, Nishimoto 2004, Lüdeling/Evert, in press, Meibauer/Guttropf/Scherer 2004). All of these measures use frequency data from corpora, mostly focussing on the distribution of the words formed by a given process, and using especially the low frequency data.

One important aspect of quantitative productivity is the situational or communicative need to form a new word. This has, of course, long been noticed and formulated time and again. Compare Hermann Paul

> Die Möglichkeit zur Bildung von Zuss. aus zwei Substantiven ist unbegrenzt. Ob solche aber wirklich gebildet werden, hängt natürlich vom Bedürfnis ab. (Paul 1920, 15; our translation: The possibility to form noun-noun compounds is unlimited. Whether they are actually formed, however, depends on the need.),

or

> Words are only formed as and when there is a need for them (Bauer 2001, 143).

The need can often be satisfied in several ways. For example, consider the need to express the thought that someone is doing too much of something (note that a "need" in the sense used here is not a specific concept, bur rather a template with variables in it – Y in this case). This need could be expressed as 'someone does too much of Y', 'someone has Y-hysteria', 'someone suffers from Yitis' etc. Qualitative pro-

ductivity is concerned with the nature of such possibilities, but it is not interested in their frequency of use. However, from the point of view of quantitative productivity, all morphological and syntactic ways of expressing the need compete with each other. While this is often stated for morphological means (less so for syntactic ones, although these are, of course, also players in the competition), calculations of quantitative productivity typically look at single unrelated processes – if measures of productivity of different processes are compared, need or competition are not taken into account.

As stated above, evidence used to calculate quantitative productivity stems from corpus counts. These corpus counts can be affected by

(a) variations in the need (changes over time or between genres, for example)

(b) the inherent potential of each process that could satisfy the need and

(c) the potentials of other competing processes.

For example, the rising productivity rates of non-medical *-itis* (a reading of the formative *-itis* used in non-medical contexts with the meaning of 'doing too much Y', as in *Telefonitis*) in the German of the 1990s reported in Lüdeling/Evert (in press) could be due to a changing need, a change of the inherent potential of the *-itis*-derivation itself, or a change in the potentials of its competitors (or a combination of these factors).

It can easily be seen that it is extremely difficult to find data that allow us to study the role of need and competition in productivity, for at least two reasons: (i) The need must have very specific semantics, so that some of the (more likely) means of expression can be enumerated. (ii) The corpus should be of sufficient size and variety, so that each alternative process appears frequently enough for statistical measures to be computed (see Lüdeling/Evert, in press, on data sparseness problems in studying diachronic productivity even with a very large diachronic corpus).

In this poster, we present a model for the calculation of the productivity of a need and of the competing processes that can satisfy it. This model represents the frequency distribution of the "need" as a mixture of the distributions of the respective processes, similar to the mixture models for word frequency distributions sketched in Baayen (2001).

References

Baayen, R.H. (2001). Word Frequency Distributions. Kluwer, Dordrecht.

Bauer, L. (2001). Morphological Productivity. Cambridge University Press, Cambridge

Dressler, W. (2003). Degrees of grammatical productivity in inflectional morphology. Rivista di Linguistica, **15**: 31-62.

Lüdeling, A. and Evert, S. (in press). The emergence of productive non-medical -itis. Corpus evidence and qualitative analysis. In S. Kepser and M. Reis, eds., Linguist-

ic Evidence. Empirical, Theoretical, and Computational Perspectives. Mouton de Gruyter, Berlin.

Meibauer J., Guttropf, A., and Scherer, C. (2004). Dynamic aspects of German -er-nominals: a probe into the interrelation of language change and language acquisition. Linguistics, **42**: 155-193.

Nishimoto, E. (2004). A corpus-based delimitation of new words: cross-segment comparison and morphological productivity. PhD Thesis. City University of New York.

Paul, H. (1920). Deutsche Grammatik. Teil V: Wortbildungslehre. Max Niemeyer Verlag, Tübingen (Unveränderter Nachdruck der ersten Auflage, 1968).