

*Lexical Productivity: theoretical issues and quantitative measures*

Isabella Chiari

Dipartimento di Studi Filologici, Linguistici e Letterari

Università “La Sapienza” di Roma

[isabella.chiari@uniroma1.it](mailto:isabella.chiari@uniroma1.it)

A long and productive debate has been alive from the end of the Seventies up to now on theoretical and quantitative aspects of *morphological productivity* (e.g. Aronoff 1976, 1983; Booij 1977; Baayen 1992, 1993; Baayen, H., & A. Renouf. 1996; and, particularly on Italian language, Gaeta, L., & D. Ricca 2002, 2004), pointing out methodological and speculative issues related to affix usage in large corpora. On the other hand, while observing lexical roots, loan words usage and trends, we might wonder if it might be useful to introduce what analogously can be called *lexical productivity*? When we try to describe and measure neologism’s behaviour in language, we cannot avoid observing, besides textual usage and diachronic trends of the single lexeme, the fact that some new roots tend occur often in other grammatical functions, as basis for affixation, as elements of compounds, and in specific collocations. For example, when dealing with loan words in Italian, such as *internet*, we detect also: *free internet*, *internet banking*, *internet broker*, *internet café*, *internettiano*, *internettista*, *internettizzare*, *internettizzato*, *internavigatore*, *internauta*, *internavigatrice*. The simple presence of such related usages make lexemes more integrated in the lexical system and in the speaker’s linguistic conscience.

The idea of monitoring and defining lexical productivity is far less common than it’s morphological counterpart. We find traces of qualitative and quantitative definitions in Kocourek (1996), Nakagawa (2000), Diaz Vera (2003), Kageura (2004) where simple reports of number of types, tokens and compounding aspects are accounted for. What is needed is a new measure able to report synchronic and diachronic productivity. *Synchronic lexical productivity* can be viewed as an index of textual usage in a given corpus, able to include as its parameters: a) number and frequency of different grammatical usages of the lexeme (e.g. *download* as a noun and as a verb), number and frequency of derivational lexemes (e.g. *jeans* > *jeanseria*; *hobby* > *hobbista*); c) number and frequency of compound words formed with the lexeme as an element (and in this case weighting roles as head of the compound and position) (e.g. *mail* > *web-mail*, *voice-mail*, *junk-mail*), including recursive classes of type a, b and c; d) number and frequency of collocations. *Diachronic lexical productivity* should be seen as indicating positive, negative or multimodal trends in lexical productivity observed at regular intervals (a slight modification of a measure presented in Bolasco 2005 will be proposed). It is possible to suggest the substitution of the mere frequency with usage index, only if segmentation for the determination of dispersion can be homogeneously applied to each diachronic interval.

Quantitative measures presented will be tested on a Italian corpus of five years of the newspaper “La Repubblica” (1996-2000, about 22 million word token each year) with approximately 110 million word occurrences. A number of recently attested loan words will be observed from the point of view of synchronic and diachronic productivity as well as in correlation with textual usage of the root forms analyzed. A further classification of loan words according to usage trends and productivity evidence will be proposed.

Theoretical questions related to the delicate phase of weighting different parameters that need to be included in the formulas, as well as questions about what kind of lexemes can profitably be described in term of lexical productivity (neologisms, loan words, etc.) will be addressed. Theoretical and methodological similarities and differences with morphological productivity issues will also be focused.

## References

- Aronoff, M. 1976. *Word formation in Generative Grammar*. Cambridge, MA: MIT Press.
- Aronoff, M. 1983. Potential Words, Actual Words, Productivity and Frequency. In: S. e Inoue K. Hattori (eds.), *Proceedings of the XIII International Congress of Linguists*. Tokyo: Permanent International Committee on Linguistics, 163-71.
- Baayen, H. 1992. Quantitative aspects of morphological productivity. In: G. Booij, e J. van Marle (eds.), *Yearbook of Morphology 1991*. Kluwer: Dordrecht, 109-49.
- Baayen, H. 1993. On frequency, transparency and productivity. In: G. Booij, e J. van Marle (eds.), *Yearbook of Morphology 1992*, 181-208. Dordrecht: Kluwer.
- Baayen, H., and Renouf. 1996. Chronicling the Times: Productive lexical innovations in an English newspaper. *Language* 72: 69-96.
- Bolasco, S. (2005), “La reperibilità statistica di tendenze diacroniche nell’uso delle parole” in De Mauro T. e I. Chiari (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Roma, Aracne, 2005, pp. 335-354.
- Booij, G. 1977. *Dutch morphology*. Lisse: de Ridder.
- Diaz Vera, J. (2003), Lexical and Non Lexical linguistic Variation: in the Vocabulary of Old English, *Atlantis*, 21(1), 29-30.
- Gaeta, L., e D. Ricca. 2002. *Corpora testuali e produttività morfologica: i nomi d’azione italiani in due annate della Stampa (1996-1997)*. In: R. Bauer, e H. Goebel (eds.), *Parallela IX. Testo variazione informatica / Text Variation Informatik. Atti del IX Incontro italo-austriaco dei linguisti..* Wilhelmsfeld: Egert, 223-49
- Gaeta, L., e D. Ricca. 2005. “Aspetti quantitativi della produttività morfologica”. In T. De Mauro e I. Chiari (eds.), *Parole e numeri. Analisi quantitative dei fatti linguistici*, Roma: Aracne, pp. 107-124.
- Kocourek, R. 1996. The prefix post- in contemporary English terminology. *Terminology*, 3:1, pp. 85-110
- Nakagawa, H. 2000. Automatic term recognition based on statistics of compound nouns. *Terminology*, 6(2): 195–210.
- Kageura, K. 2004. Quantitative Portraits of Lexical Elements. In S. Ananadiou & P. Zweigenbaum (eds.), *COLING 2004 CompuTerm 2004: 3rd International Workshop on Computational Terminology*, Geneva: COLING, pp. 75-8.