Cross-linguistic research into derivational networks

Alexandra Bagasheva
Sofia University "St. Kliment Ohridski"

Lívia Körtvélyessy
P.J. Šafárik University, Košice

Pavol Štekauer
P.J. Šafárik University, Košice

Salvador Valera
University of Granada

Jan Genci
TUKE, Technical University Kosice

Until relatively recently, derivational networks[1] have received little theoretical attention (exceptions of various theoretical frameworks include, e.g. Dokulil 1962, Pounder 2000, Beecher 2004, Ševčíková & Žabokrtský 2014, Bonami & Strnadová 2016). Largely as a result of this, no major empirical, let alone cross-linguistic, research has been implemented yet. The derivational network is defined as the intersection of the *paradigmatic capacity*, i.e. the system of all direct derivatives from a simple underived word-formation base, and the *order of derivation*, i.e. all the linear derivatives from the word-formation base. This does not necessarily assume linear derivation.

The poster presents the fundamental theoretical and methodological principles as well as selected results of a research project covering 40 European languages.[2] The point of departure is the choice of 10 basic words per three major word-classes: nouns, adjectives and verbs. Starting from a larger list, the informants compiled a list of 30 words identical for the sample languages, excluding any derived basic words. The initial longer lists of words were selected from Swadesh's core vocabulary. Since these words serve as a starting point for the development of cross-linguistically comparable derivational networks, all of them must be simple underived words in all sample languages.

The research introduces a semantic classification of all the possible derivatives (as attested in corpora and/or dictionaries, in this order as available for each language) in individual orders of derivation by means of a list of semantic categories that was compiled for this purpose. This means that each derivational step (order of derivation) away from the base is defined by means of a specific semantic category. This approach allows to identify:

i)   correlations between the occurrence of individual semantic categories and the order of derivation,

ii)   semantic categories that systematically block any further derivation, and

iii) the maximum number of derivatives per each semantic category within a given order of derivation.

We speak of the paradigmatic capacity of the word-formation base represented by the number of derivatives from the word-formation base. The paradigmatic capacity and the orders of derivation establish the *derivational network*, that is a network of derivatives derived from the same word-formation base (simple underived word) with the aim of

---

[1] Derivational is used here in the narrow sense of affixation. No other processes are taken into account. Conversion being excluded for methodological reasons, no pure transposition was recorded, so word-class change was always found to be associated with an intervening, additional semantic category.

[2] Basque, Bulgarian, Catalan, Chechen, Croatian, Czech, Danish, Dargwa, Dutch, English, Estonian, Finnish, French, Frisian, Galician, Georgian, German, Greek (Modern), Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Saami (North), Serbian, Slovak, Slovene, Spanish, Swedish, Tatar, Turkish, Ukrainian, Welsh.

formally representing specific semantic categories. Thus, the sum total of maximum derivations for all ten words of a given word-class identifies the *maximum derivation network*. This variable serves as a basis for the calculation of the *saturation value* for individual words of a given word-class as well as for the word-class as a whole. The saturation value is calculated as a proportion between the number of actual derivatives in a particular derivational network and the maximum derivational network. The saturation value identifies the degree to which a particular word makes use of the derivational capacity of the derivational network. If the saturation values for all ten words are consistent in a given order of derivation, i.e. if they do not deviate statistically significantly, the predictability of derivation within a given word-class is high. This has been evaluated by means of the parameter of *standard deviation*. In practice, the above allows to evaluate language genera by orders of derivation and word-classes.

Additional findings concern the identification of:
i)   the average number and the maximum number of orders of derivation within the examined sample of languages,
ii)   the maximum and the average paradigmatic capacity for three groups of sample words (nouns, verbs and adjectives),
iii)   the total number of derivatives per order of derivation for each basic word and the average number of derivatives per order of derivation per word-class, and
iv)   systematic combinations of semantic categories characteristic of the specific sample/language.

All in all, the poster presents a new perspective for studying word-formation combinability which has been restricted to the study of affix combinations so far. This is particularly relevant, considering semantic combinability in its diverse manifestations in word-formation has not been explored yet. The poster presents the following main conclusions:
i)   Languages differ considerably in their derivational capacity, i.e. in the number of derivatives in derivational networks.
ii)   The derivational potential of simple underived nouns and adjectives is very similar. Verbs have the highest maximum derivational network value in every order of derivation.
iii)   The richness of derivational networks is sensitive to the word-class of the basic word and to the order of derivation.
iv)   There is a tendency for languages to actualize 20%-30% of the derivational potential of a word-class.
v)   Some languages keep high saturation values across all three word-classes.
vi)   There is a tendency for saturation values to fall gradually with the rising order of derivation in all three word-classes.
vii)   There is no geographically homogeneous territory on which the languages of topmost saturation values are spoken.
viii)   Derivational networks are most predictable in the 1st order.
ix)   The maximum number of orders of derivation is five for all three word-classes.
x)   Semantic categories with blocking effects are reported in all the languages of the sample except in Welsh.
xi)   Typical combinations of semantic categories are reported for most languages, also regularly for each word-class.
xii)   Recursiveness is relatively frequent in the sample, but it is comparatively higher in Slavic languages.

xiii) Differences can be found between language types as regards multiple occurrence of semantic categories.

Bibliography

Beecher, Henry. 2004. Derivational paradigm in word formation. Available at: http://pdfcast.org/pdf/research-paperi-derivational-paradigm-in-word-formation.

Bonami, Olivier & Jana Strnadová. 2016. *Derivational paradigms: pushing the analogy*. Paper presented at the *49th Annual Meeting of the Societas Linguistica Europaea (SLE)*, Naples, Italy, 31 August–3 September 2016.

Dokulil, Miloš. 1962. *Tvoření slov v češtině I. Teorie odvozování slov*. [Word-formation in Czech, I. Theory of derivation.] Prague: Nakladatelství Československé akademie věd.

Pounder, Amanda. 2000. *Processes and Paradigms in Word-Formation Morphology*. Berlin/New York: Mouton de Gruyter.

Ševčíková, Magda & Zdeněk Žabokrtský. 2014. Word-formation network for Czech. In Nicoletta Calzolari et al. (eds.), *Proceedings of the 9th International Language Resources and Evaluation Conference (LREC 2014)*, 1087–1093. Paris: ELRA.