Cross-linguistic research into derivational networks

Lívia Körtvélyessy
P.J.Šafárik University/Moyzesova 5, 04001Košice, Slovakia
livia.kortvelyessy@upjs.sk

1 Introduction

In the past decades, word-formation vindicated its position in the system of linguistic sciences. More and more attention has been paid to phenomena which are typical of this field. Frequently, morphologists are in search for analogies between inflection and derivation. A good example of this are derivational paradigms. Even though they are not in the centre of theoretical considerations, they have already been discussed within various theoretical frameworks, e.g. Dokulil (1962), Horecký et al. (1989), Pounder (2000), Beecher (2004), Furdík (2004), Ševčíková and Žabokrtský (2014), Bonami and Strnadová (2016). The paper presents a new contribution to the discussion. The basic notion is the derivational network which differs from derivational paradigms by its three-dimensional nature.

2 Theoretical background

Derivational paradigms can be treated as a system of complex words derived from a single word-formation base. This includes all direct derivatives from a single word-formation base (first dimension), for example:

(1)	(i)	dom	'house'
	(iii)	dom-ček	'little house'
	(iv)	dom-ík	'little house'
	(v)	dom-isko	'large house'
	(vi)	dom-ov (adverb of direction)	'towards one's home'

In this case, we speak of the **paradigmatic capacity** of the word-formation base represented by the number of derivatives from the word-formation base.

In addition, there is another (second) dimension that should be taken into consideration, in particular, all **linear** derivations from a single word-formation base, as in (2):

(2)	(a)	dom 'house'	dom-ov 'home'	dom-ov-ina 'homeland'	dom-ov-in-ový 'related to homeland'
	(b)	dom 'house'	dom-ček 'little house'	dom-ček-ový 'related to a little house'	
	(c)	dom 'house'	dom-ík 'little house'	dom-ík-ový 'related to a little house'	
	(d)	dom 'house'	dom-isko 'large house'	dom-isk-ový 'related to a large house'	

This dimension enables us to identify the number of affixation operations available for a given basic underived word. Each affixation operation represents one order of derivation. By implication, this dimension identifies the number of linear derivations. In example (2), (2a) shows three orders of derivation, while (2b) through (2d) permit two orders of derivation from the same simple underived word *dom* 'house'.

Proceedings of the 2nd Int. Workshop on Resources and Tools for Derivational Morphology (DeriMo 2019), pages 1–4,
Prague, Czechia, 19-20 September 2019.

Each derivational step introduces (and therefore expresses and represents) a particular semantic category (third dimension). In (2a), these are, respectively, Location, Location and Quality, in (2b) and (2c) Diminutive and Quality, and in (2d) Augmentative and Quality. By implication, a combination of derivatives from the same base identifies a combination of semantic categories realized in the process of consecutive affixal derivations. It follows from example (2) that one and the same basic word can give rise to several paths of consecutive derivations, each of which has its specific number of derivatives representing specific semantic categories.

The paradigmatic capacity and orders of derivation establish a derivational network, that is, a network of derivatives derived from the same word-formation base (simple underived word) with the aim of formally representing specific semantic categories.

3 What was compared?

Derivational networks may substantially differ from language to language in their complexity. However, no major empirical, the less so cross-linguistic research has been implemented yet. For this reason, a research project was designed that is aimed at comparison of derivational networks in 40 languages of Europe. Two criteria for the selection of languages were applied: (i) each language presented a language of Europe. The primary source was the languages covered in the HSK Word-Formation; (ii) the number of languages was reduced on the basis of their data availability, i.e., according to the possibility to verify the existence of derived words by means of representative dictionaries and/or corpora. An important reference guide in this respect was *Ethnologue*, in particular, its *Expanded Graded Intergenerational Disruption Scale*.

Parallel to inflectional paradigms, derivational networks rely on word-classes. Our derivational networks include nouns, verbs and adjectives as basic words. Each of these word-classes is represented by 10 simple underived words. Words were selected from Swadesh's core vocabulary list.

4 Points of comparison

The primary objective of the project was to compare derivational networks in 40 languages of Europe. For any typological analysis a tertium comparationis is needed. For this purpose, we introduced the concept of saturation value as quantitative representation of structural richness. Saturation value calculations are based on the concept of the Maximum Derivational Network (MDN). The maximum derivational network results from the intersection of all implemented (actual) derivations found for all basic words of an examined sample within a particular word-class. For its computation, it is necessary to identify the highest number of derivatives for a given semantic category from among all ten sample words (in our research) of a given word-class. The MDN values enable us to calculate the saturation value for individual words by means of the formula:

$$SV = \frac{D}{MDN} \times 100(\%) \tag{1}$$

Legend: SV - Saturation value, D - Number of derivatives, MDN - Maximum derivational network

Saturation value is computed for each derivational network in each language. As explained above, in each language 30 derivational networks were constructed (10 for nouns, 10 for verbs, 10 for adjectives). The gained data enable us to compare languages in terms of the number of derivatives, the number of orders of derivation, the maximum derivational network, and the saturation value. Obviously, we also search for various associations between these parameters. Furthermore, the dimension of semantic categories is evaluated. This makes it possible to identify correlations between semantic categories and orders of derivation and typical combinations of semantic categories; to identify semantic categories with blocking effects, their multiple occurrences, and the reversibility of semantic categories.

5 Conclusions (selected)

i. There are considerable differences among languages in their derivational capacity, which is reflected in the number of derivatives in derivational networks.

ii. If we compare the average MDN values by word-classes and by orders of derivation it is obvious that the derivational potential of simple underived nouns and adjectives is very similar, in some orders almost identical. Verbs have clearly the highest MDN value in every order of derivation, significantly higher than the other two word-classes. This is especially due to an extreme derivational potential of those languages which employ prefixes for the expression of the category of Aktionsart.

iii. The richness of derivational networks is sensitive to the word-class of the basic word. This means that for the majority of languages the richness of derivational networks varies depending on the word-class of the basic words. A high consistency across all three orders in all three word-classes is rare but does occur in Bulgarian and Serbian. If restricted to order 1, highly consistent networks in all three word-classes have been identified for Croatian, Turkish and Basque, and, partly, Bulgarian, Polish and Welsh.

iv. The richness of derivational networks is sensitive to the order of derivation.

v. There is a tendency for languages to actualize 20-29.99 % of the derivational potential of a word-class. This tendency is almost identical for all three word-classes and is represented by 67.5 % of languages for nouns and 62.5 % of languages for both verbs and adjectives.

vi. There is a core group of languages that keep high saturation values across all three word-classes. They include Greek, Dutch, North Saami and Dargwa. They might be completed with German, Turkish and Lithuanian which have high values in two word-classes and a medium SV in the third word-class.

vii. There is an unambiguous tendency for saturation values to fall gradually with the rising order of derivation in all three word-classes.

viii. The saturation values do not vary for the examined genera in a significant way in any of the word-classes which indicates that it is possible to predict the level of richness of derivational networks for language genera.

ix. A medium saturation value (20-30 %) can be considered the most typical saturation value for all word-classes and the first three orders of derivation.

x. There is no geographically homogeneous territory on which the languages of topmost saturation values are spoken. These languages are of various genetical origins and are scattered across Europe. What, however, can be considered as a general tendency is the use of low- saturation values languages at geographically peripheral areas of Europe.

xi. The correlation between saturation value and paradigmatic capacity may significantly differ for the same language in different word-classes and different orders of derivation.

xii. The maximum number of orders of derivation, i.e., the maximum number of affixes attached to a simple underived word is five for all three word-classes. There are six languages reaching five orders of derivation in all three word-classes, none of them belonging to the Romance or Germanic genus. The average number of affixation steps is very similar for verb-based and adjective-based derivation (2.78 and 2.76, respectively). It is lower for nouns (2.46).

xiii. Also in terms of the total number of derivatives, the most prolific base is verb. The average of verb-based derivatives clearly outnumbers the figures for adjectives and nouns. This word-formation feature is dominated by Slavic and Uralic languages. The values for the adjective-based derivation are slightly higher than those for the noun-based derivation.

xiv. Inflectional and agglutinating languages tend to have a high number of derivational orders. However, the genetic factor might be influential, too. Romance inflectional languages have smaller number of derivational orders than Slavic. While Nakh-Daghestanian languages, classified as agglutinating, tend to have a very low number of derivation orders Uralic languages, also agglutinating, feature high numbers. Analytic languages are not consistent in their behaviour. Generally, they tend to have lower number of derivation orders, especially in the case of nouns.

xv. The most clearly correlated with the first order of derivation are the following semantic categories:

- Nominal bases: Diminutive, Quality, Privative, Relational and Action
- Verbal bases: Action, Agent, Resultative and Ability
- Adjectival bases: Manner and Stative

xvi. In the second order of derivation they are:

- Nominal bases: Action and Stative
- Verbal bases: Action and Agent.

xvii. No recurrent patterns could be established in terms of the combinability of semantic categories or the blocking capacity of semantic categories from a cross-linguistic point of view (with the exception of the blocking effect of Diminutive). This suggests that there is no universal cognitively founded succession of derivational operations.

xviii. Three semantic categories stand out in terms of their capacity to reoccur in successive orders of derivation: Quality, Action and Diminutive.

References

Henry Beecher. 2004. Derivational paradigm in word formation. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.9071&rep=rep1&type=pdf.

Olivier Bonami and Jana Strnadová. 2016. Derivational paradigms: pushing the analogy. A paper presented at the SLE conference, Naples, Italy, August- 3 September 2016.

Miloš Dokulil. 1962. *Tvoření slov v čestine. Teorie odvozování slov*. Academia, Prague.

Magda Ševčíková and 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)*. ELRA, Paris, pages 1087–1093.

Juraj Furdík. 2004. *Slovenská slovotvorba*. Edited by Martin Ološtiak. Náuka, Prešov.

Ján Horecký, Klára Buzássyová, and Ján Bosák et al. 1989. *Dynamika slovnej zásoby súčasnej slovenčiny*. Veda, Bratislava.

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