Lívia Körtvélyessy, Alexandra Bagasheva, Pavol Štekauer, Salvador Valera

1 Introduction

This monograph is aimed at the examination of derivational networks across European languages. The concept of a derivational network is not new. The first ideas of network regularities and the network organization of derivational morphology can be traced back to the 1960s in relation to the Dokulilean tradition in word-formation. Unfortunately, apart from an outline of general principles, very little has been done in the field since. In recent years, however, we have been witnessing a growing interest in derivational paradigms and larger derivational systems based on them. A brief overview of this direction of morphological research is presented in section 1.1.

In spite of, or better, precisely because of what is outlined in section 1.1, this volume is pioneering in terms of both the theory and its scope for a number of reasons:

(i) First and foremost, a new method of examination and comparison of derivational networks in various languages is introduced, including new criteria and parameters for their evaluation, including the maximum derivational network, the saturation value, the number of orders of derivation, the correlation between the paradigmatic capacity and the order of derivation, the typical combinability of semantic categories, and the blocking effects of semantic categories.

(ii) Research into word-formation paradigms is mostly exploratory even though the basic utility of paradigms is assumed to be explanatory, so in this sense the current research explores the applicability of word-formation paradigms in typological derivational research.

(iii) It introduces the idea of derivational networks relying on the concept of the derivational paradigm, extendable both vertically and horizontally. The *derivational network* is conceived as an intersection of *paradigmatic capacity* per *order of derivation*, and is evaluated in terms of the *structural richness* that is quantitatively represented by calculating the *saturation value*.

(iv) The vertical dimension operates with the narrow understanding of a paradigm as being applicable across the fluid boundary between inflection and derivation, while the horizontal incorporates specific features of word-formation families (understood narrowly as equal to series) as constituting one type of associative, paradigmatic relation.

Körtvélyessy, Lívia, Bagasheva, Alexandra, Štekauer, Pavol and Valera, Salvador. "1 Introduction". *Derivational Networks Across Languages*, edited by Lívia Körtvélyessy, Alexandra Bagasheva and Pavol Štekauer. Berlin, Boston: De Gruyter Mouton, 2020, pp. 1-26. https://doi.org/10.1515/9783110686630-001

(v) It introduces into the research of derivational networks a third dimension – a strong semantic perspective in the form of the classification of individual derivatives by means of *comparative semantic categories*. In other words, rather than representing the meaning of the derived word as a whole, a semantic category represents the derivational meaning of the affix attached to the word-formation base. Semantic categories constitute an open-ended set of theory-neutral, cross-linguistically applicable, comparative semantic concepts. For the purposes of this research, we utilize a list of 49 semantic categories that seem to exhaust the semantic specificity of the sample languages. As comparative concepts employed in typological research (Corbett 2010; Haspelmath 2010), these conceptualize prototypes abstracted from descriptive categories. An extensive investigation of the available non-decompositional models of semantic analysis of affixation phenomena and the calibration of posited onomasiological categories in comparative semantic concepts served as the basis for compiling the set applied here. (For a detailed presentation of the principles of the compilation of the list of comparative semantic categories employed in the research, see Bagasheva 2017.)

(vi) The research methodology rests upon a usage-based approach to language (see Barlow and Kemmer 2000; Bybee 2006), utilizing a bottom-up approach of analyzing actual data gathered in individual languages for drawing the respective derivational networks.

(vii) It is the first piece of *large-scale empirical research* into derivational networks. By implication, there has been no *cross-linguistic research* into derivational networks to date. This monograph presents and evaluates data from a sample of 40 languages from across Europe.

(viii) This makes it possible to draw generalizations and evaluate the role of the genetic factor, the morphological type,[1] the nature of a language's word-formation system, the word-class of the basic word, and the order of derivation in the construction, complexity and richness of derivational networks.

(ix) Last but not least, the data enables us to contribute to the discussion on the areal typology of European languages to determine a zonation according to the parameter of derivational network richness.

1 When referring to morphological typology, we rely on the traditional classification proposed by Sapir (1921) and Skalička (2004–2006).

1.1 Previous research

As indicated above, the idea of complex and systematic relations among derivatives organized around a simple underived base word is not new. Relevant discussions can be found, for example, in the works of Czech and Slovak linguists from the 1960s and the following decades, inspired (as for many other word-formation issues) by Dokulil's seminal work (1962), with the fundamental theory systemized in Horecký et al. (1989) and Furdík (2004). This line of research is based on the principle of word-formation motivation as a universal principle encompassing and influencing almost the whole word-stock. Furdík (2004: 74) even speaks of derivational 'cases'. Then, the *derivational paradigm* is conceived as an ordered system of motivated units grouped around a single motivating unit and constituting motivation pairs with it (Horecký et al. 1989: 28–29; see example 1):

(1) *škola* 'school' *škol-ák* 'schoolboy'
škol-ník 'school janitor'
škôl-ka 'kindergarten'
škol-stvo 'education system'
škol-ička 'small school' (Furdík 2004: 74)

A sequence of consecutive motivation pairs constitutes a derivational series (Dokulil 1962: 13) or chain (Zych 1999: 12). This is illusatrated in example (2), including seven orders of derivation:

(2) *hodný* 'worthy' >
hodnota 'value' >
hodnotiť 'evaluate' >
zhodnotiť 'evaluate.RESULTATIVE' >
zhodnocovať 'evaluate.DURATIVE' >
zhodnocovateľ 'evaluator' >
zhodnocovateľský 'evaluating' >
zhodnocovateľsky 'in an evaluating manner' (Furdík 2004: 74)

A system of derivational paradigms and series/chains organized around one basic underived (non-motivated) word constitutes a *derivational nest* (Horecký et al. 1989; Furdík 2004).

(3)

rezať	*rez-ač* 'cutter.AG'	*rezač-ka* 'cutter.AG.F'		
'to cut'	*rez-ačka* 'cutter.INSTR'			
	rez-ák 'incisor'			
	rez 'cut.N'			
	reza-nie 'cutting'			
	rez-ba 'carving'	*rezb-ár* 'carver'	*rezbár-ka* 'carver. F'	
			rezbárstvo 'woodcarving'	*rezbár*-sky 'concerning woodcarving'
	rez-ivo 'lumber'		*rezb-ársky* 'in the woodcarving manner'	
	rez-eň 'cutlet'	*rezn-ík* 'small cutlet'		
	rez-ina 'sawdust'			
	rez-ký 'brisk'			
	rez-ací 'cutting.ADJ'			
	reza-teľný 'cuttable'	*rezateľn-osť* 'cuttability'		
	rez-aný 'cut.PP'	*rezan-ka* 'noodle'		
		rezan-ec 'noodle'		
	od-rezať 'cut off.V'	*odrez-ok* 'shred'		
	v-rezať 'cut into'	*vrezať sa* 'cut into.REFLEXIVE'		
	nar-ezať 'slice'	*nárez* 'slice.N'		
		nareza-nie 'slicing'		
	etc.			(Horecký et al. 1989: 39–40)

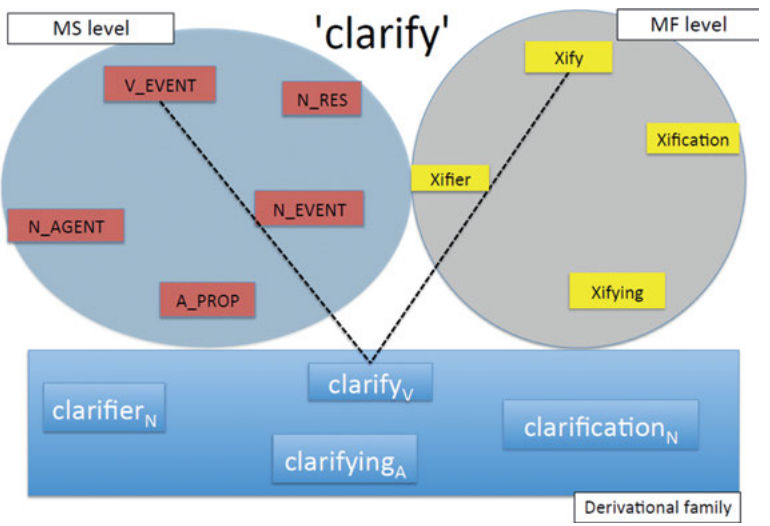
This means that derivational nests are constituted by a set of derivational series in the syntagmatic direction and by a set of derivational paradigms in the paradigmatic dimension. A derivational nest covers motivated words with identical onomasiological marks but different onomasiological bases (Horecký et al. 1989: 31). As noted by Kardela (2015: 294), "[t]he theoretical import of the lexical nest should be obvious: the nests form a network of interrelated items which help state the complex derivational relations between the various lexical items and derivatives thereof."

Word-formation research in the recent period has brought renewed interest in complex derivational systems from various perspectives, including a revived interest in derivational paradigms. The discussion of paradigms in word-formation, as Blevins (2013) remarks, continues a venerable tradition in word-based models of the architecture of grammar dating back to ancient Greece. Despite vagaries of disparate development, all such models, including contemporary ones, according to Blevins, "project morphological analysis primarily upwards from the word, and treat the association of words with paradigms or other sets of forms as the most fundamental morphological task" (2013: 375).

Hathout and Namer (2016, 2019) propose a multi-level paradigm-based model, relying on the concepts of derivational family, arrangement relations, and the derivational paradigm. The derivational family is defined as a network of derivationally-related lexemes (e.g. *clarify*, *clarifier*, *clarifying*, *clarification*);

the arrangement relations (that correspond to the alignment relations discussed by Bonami and Strnadová 2019) connect the lexemes formed by the same derivational process. Hathout and Namer (2016) distinguish between the morpho-semantic (MS) and the morpho-formal (MF) levels of description. These levels of description of a derivational family are related by pairing individual morpho-formal units (e.g. *-ify* in *clarify*) with the corresponding morpho-semantic category (labeled 'concept'); in this case, it is *V_Event*, i.e. an Event represented by a verb. The individual morpho-formal units as well as morpho-semantic categories are interconnected to constitute modules that, as assumed by Hathout and Namer, represent systems of interpredictability between words (derivational family), concepts (MS) and formal patterns (MF).

(4) An example of a multi-level paradigm-based model (Hathout and Namer 2016)



The concept of the derivational paradigm is also a point of departure for the discussion of more complex relations by Bonami and Strnadová (2016, 2019). They work with derivational (sub)families that exhibit key properties shared by inflection systems. Their understanding of the morphological subfamily is analogical to the definition of the derivational family by Hathout and Namer above, i.e. they define it as a set of morphologically related words. A paradigmatic system is then a collection of (partial) families that are aligned in terms of the content-based relations that their members entertain. The notion of alignment is purely content-based, and so it covers word pairs, such as *random, randomize*; *class*, *classify*; or

order$_N$, *order*$_V$ – all these pairs are aligned through the CAUSATIVE relation. In general, they consider the content-based contrast between words to be the fundamental feature of paradigm structure.

Bonami and Strnadová emphasize the considerable similarity between inflectional paradigms and morphological families, which is projected onto their fairly liberal understanding of the morphological family: it permits inclusion in a single family of both inflectionally and derivationally-related words (e.g. *sing*, *sang*, *singer*). A specific feature of their approach is that the paradigmatic system does not allow for gaps (defectivity) or synonymy within a paradigm (overabundance) – they are purposefully ignored. Given this theoretical background and a number of analogies between inflectional and derivational paradigms, the authors give evidence that the method of computation of the predictability within inflectional paradigms is also applicable to that within derivational paradigms.

Rodrigues and Rodrigues (2017) speak of cross-paradigms conceived as "mental patterns dynamically organized around more than one axis." They distinguish between two main paradigmatic organizations, in particular, the lexeme-based and the affix-based paradigms. The former is illustrated with deverbal nouns, including various affixes adding the same Processual meaning to the ACTION represented by the basic verb, as in the Portuguese examples *avaliar* 'to evaluate' > *avaliação* 'evaluation', *matar* 'to kill' > *matança* 'slaughter', or *aterrar* 'to land' > *aterragem* 'landing'. In other words, this type of paradigmatic organization relies on various affixes adding the same meaning to word bases belonging to the same class of derivational base. The second type of paradigmatic organization employs one and the same affix, for instance, the Portuguese suffix *-ism*(*o*) in the series *medievalismo* 'medievalism', *espiritualismo* 'spiritualism', *luteranismo* 'Lutheranism', *newtonianismo* 'Newtonianism' and *figurativismo* 'figurativism'.

The central claim of Rodrigues and Rodrigues is that these two types of paradigmatic organization can interact to establish cross-paradigms. The ability of an affix to operate on derivational bases of different word-classes is semantically grounded and accounted for by what Libben (2014) labels as morphological superstates. An important condition for the formation of this kind of cross-paradigms is the size of the morphological family: the formation of cross-paradigms is, as suggested by Rodrigues and Rodrigues (2017), restricted to rich morphological families.

A different line of research into complex derivational relations is represented by tools and models employed by computer linguistics, for example, 'neural' models serving the completion of derivational paradigms, inspired by well-established models of inflectional paradigm completion (Cotterell et al.

2017) and computerized systems employed for the establishment of derivational nests (networks) in the Czech language (e.g. Pala and Hlaváčková 2007; Ševčíková and Žabokrtský 2014; Pala and Šmerk 2015).

1.2 Theoretical principles

The point of departure in our approach is the concept of *paradigm*. It has traditionally been discussed exclusively within the field of inflectional morphology. The idea of derivational paradigms has, for a long time, been called into question.[2] Nevertheless, there is significant parallelism between inflectional and derivational paradigms. The main points of *correspondence* are reviewed in what follows.

(i) Both of these types of paradigm operate within *word-classes*. Thus, there are, among others, substantival, verbal, adjectival and adverbial paradigms in inflection, which means that (inflectional) affixes are attached, respectively, to nominal, verbal, adjectival and adverbial bases. In derivation, in an analogical manner, paradigms are also based on nouns, verbs, adjectives and adverbs. The fact that the word-class can change in the process of derivation is not important (word-class changing affixation is a typical feature of derivational processes) in this respect because it does not affect the categorial foundation of either inflectional or derivational paradigms. The class-changing capacity of derivation bears on the two-dimensional system of derivational nests. The inflectional and derivational paradigms are, however, unidimensional systems that rely on the identity of the word-class of the basic word.

(ii) The inflectional paradigm is based on expressing certain (grammatical) *categories* by affixes (among other possible means), for example, CASE, NUMBER and GENDER in nouns. Analogically, derivational paradigms are also based on expressing certain (semantic) categories, for example, AGENT, PATIENT, INSTRUMENT, LOCATION, ABSTRACTION, ITERATIVITY, CAUSE, RESULT OF ACTION, DIMINUTIVENESS, AUGMENTATIVENESS, etc. *By implication, both types of paradigms are organized around the concept of category*.

(iii) Each of the grammatical categories can be realized, depending on the morphological type of a language, by *one or more form-meaning units*, including various affixes. Thus, for example, the nominative plural slot of the

2 For an overview of various approaches to derivational paradigms see Štekauer (2014) and Hathout and Namer (2019).

substantival paradigm in Slovak can be represented by *-i*, *-y*, *-ovia*, *-e*, *-á*, *-ia*, or *-tá* (their use depends on the formally determined gender of the particular noun and the nature of the word-final phoneme), and this establishes various substantival paradigms for the category of nouns (twelve in total). By the same token, one can analogically speak of *one or more form-meaning units* constituting derivational paradigms, for instance, within the semantic category of AGENT in English:

(5) (i) verbal base + the suffix *-er* (*teacher*)
(ii) nominal base + the suffix *-ist* (*pianist*)
(iii) nominal base + the suffix *-ian* (*librarian*)
(iv) nominal base + the semisuffix *-man* (*milkman*)
(v) verbal base + the suffix -ist (*typist*)
(vi) verbal base + the suffix *-ee* (*escapee*)
(vii) nominal base + the suffix *-eer* (*profiteer*)
etc.

(iv) Both inflectional and derivational paradigms function as a *pattern* for new lexical items entering the system of a language. This means that both inflectional and derivational paradigms are controlled by the principles of *productivity*, *regularity* and *predictability*.

These analogies between inflectional and derivational paradigms can be completed with those proposed by Bonami and Strnadová (2016, 2019) using examples from Czech and French. These include points (v) through (vii) below.

(v) In a paradigmatic system, the formally unmarked cell (if any) need not be the same for all inflectional or derivational paradigms. In example (6a) from Czech, the same morphosyntactic feature is realized formally by different exponents: *X ~ Xů* vs. *Xa ~ X* vs. *Xa ~ Xů* vs. *X ~ X*. The same kind of differential exponence can be illustrated for a derivational paradigm (6b), where we see the alternations *X ~ Xɛ* vs. *Xi ~ X* vs. *Xi ~ Xɛ* vs. *X ~ X*:

(6) (a) Differential exponence in inflectional paradigms (example from Czech)

NOM. SG	GEN.PL	
hrad	hradů	‘castle’
žena	žen	‘woman’
táta	tátů	‘dad’
stavení	stavení	‘building’

(b) Differential exponence in derivational paradigms (example from French)

COUNTRY		INHABITANT	
France	'France'	Français	'French'
Russie	'Russia'	Russe	'Russian'
Albanie	'Albania'	Albanais	'Albanian'
Corse	'Corsica'	Corse	'Corsican'

(Bonami and Strnadová 2019: 180)

(vi) Both inflectional and derivational paradigms may use an exponence strategy that is a hybrid of two others (heteroclisis), as illustrated in (7). The contrast between nom.sg. and gen.pl. for *táta* is marked by a hybrid combination of the nom.sg. *žena* and the gen.pl. *hradů* (Xa ~ Xů). The same can be found in the derivational paradigm of (6b), where the contrast between *Albanie* and *Albanais* combines the exponents found in *Russie* (name of the country) and *Français* (inhabitant), i.e. Xi ~ Xɛ (Bonami and Strnadová 2019: 181):

(7)

NOM.SG	GEN.PL	COUNTRY	INHABITANT
hrad 'castle'	hradů	France 'France'	Français 'French'
táta 'dad'	tátů	Albanie 'Albania'	Albanais 'Albanian'

(vii) Some inflectional and derivational paradigms fail to mark the semantic difference with a corresponding form (syncretism) – this is a well-known violation of the principle of constructional iconicity proposed within the Natural Morphology theory (e.g. Dressler 2005):

(8)

nom.sg	gen.pl	country	Inhabitant
hrad 'castle'	hradů	**France** 'France'	Français 'French'
stavení 'building'	**stavení**	**Corse** 'Corsica'	**Corse** 'Corsican'

(Bonami and Strnadová 2016: 9)

This means for us that there seems to be only one substantial difference between the two types of paradigm: while the membership in inflectional paradigms is prototypically, due to the absence of competition, obligatory and automatic, the membership in derivational paradigms is prototypically, due to competition, facultative. As a result, while there are minimum gaps (but they do occur!) in the paradigms within the inflectional system, there are quite a lot of them in the derivational paradigmatic systems of natural languages. Related to this, while inflectional paradigms represent a closed system, derivational paradigms are an open system. This fact, however, does not project itself into a

chaotic and vague nature of derivational paradigms. In fact, the opposite is true. In spite of numerous gaps, derivational paradigms are highly regular and predictable, which is guaranteed by the possibility to fill any empty slot with a potential word that fits the *paradigmatic system*. From this it follows that the only major difference between inflectional and derivational paradigms concerns the fact that, while the former is based on *actual units*, the latter relies on a combination of *actual and potential units* (see also Bauer 1997). In other words, as also pointed out by, among others, Boyé and Schalchli (2019) and Gaeta and Angster (2019), the difference between inflectional and derivational paradigms is basically of a *quantitative nature*.

In summary, while an inflectional paradigm is conceived as a system of forms of a single word, 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 (vertical dimension), as illustrated in the following example with the Slovak word *dom* 'house':

(9)	(i)	*dom*	'house'
	(ii)	*dom-ov*	'home'
	(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 direct derivatives from the word-formation base (basic underived word).

In addition, there is another, syntagmatic dimension that should be taken into consideration, in particular, all linear derivations from a single word-formation base, as in (10):

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

This dimension enables us to identify the number of affixation operations available for a given basic underived word. Each such affixation operation represents one order of derivation. By implication, this dimension identifies *orders of derivation*. In example (10), (a) shows three orders of derivation, while (b) through (d) permit two orders of derivation from the same simple underived word *dom* 'house'.

The concept of derivational orders makes it possible to extend the scope of paradigmatic capacity beyond the immediate, direct derivatives from basic words to all orders of derivation. In that case, we can speak of the *derivational capacity* of the word-formation base (basic, underived word). The derivational capacity can be examined for each order of derivation separately, or it can cover all orders of derivation.

Finally, each derivational step introduces (and therefore expresses and represents) a particular *semantic category*. In (10a) these are, respectively, LOCATION, LOCATION and QUALITY, in (10b) and (10c) DIMINUTIVE and QUALITY, and in (10d) AUGMENTATIVE and QUALITY. By implication, a combination of derivatives from the same base simultaneously identifies a combination of semantic categories realized in the process of consecutive derivations. Semantics thus functions as an indispensable third dimension of our model. Any order of derivation can include more than one semantic category, and one and the same category can be formally represented by more than one affix as, for example, in (10b) and (10c), where the 1st order of derivation from *dom* 'house' includes two different affixal representations of the semantic category of DIMINUTIVE. From this it follows 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 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 formally representing specific semantic categories.

Derivational networks may substantially differ from language to language in their complexity in terms of both the number of orders of derivation and the number of derivatives in each order. This is illustrated by a comparison of derivational networks for equivalent basic words: the Icelandic word *drekka* and the Bulgarian word *pie*, both meaning 'to drink'[3]:

3 Due to the complexity of the derivational network for the Bulgarian word *pie* 'to drink', its derivational network is divided into three parts (Figures 1.1a, 1.2a and 1.3), each of which represents one order of derivation.

Figures 1.1–1.3 illustrate considerable differences between these two derivational networks in terms of both the number of derivatives and the number of orders.

The *maximum derivational network* results from the intersection ('horizontal' and 'vertical' derivations) of all implemented (actual) derivations found for all basic words of an examined sample within a particular word-class (see Figure 1.4).

In our example, the maximum derivational network for Bulgarian adjectives in the 1st order of derivation is 27 derivatives (the highlighted numbers). By adding up the maximum numbers for all orders of derivation, we get the maximum derivational network for the class of adjectives. In the case of Bulgarian adjectives, this is 88.

The concept of the *structural richness* of a derivational network of a single word-formation base is quantitatively represented by the *saturation value* calculated as a proportion between the number of actual derivatives in a particular derivational network and the maximum derivational network (cf. section 1.3.5).

1.3 Research project methodology and objectives

The research project from which this volume originates is aimed at the evaluation of a range of parameters defining derivational networks across languages: the paradigmatic capacity, the (maximum) derivational capacity, the order of derivation, the saturation value of derivational networks in individual languages and in language genera, typical combinations of semantic categories, and their potential blocking effects for a uniform sample of 30 words from 40 languages of Europe that yielded 1,200 derivational networks in total, an extremely rich source of data.

1.3.1 Sample of words

The point of departure is three word-classes, including nouns, verbs and adjectives. Each of these word-classes is represented by 10 simple underived words. Since each of the 30 basic words must be a simple underived word in each of the 40 sample languages, we chose Swadesh's core vocabulary counting 200 words because the chances of finding simple underived equivalents for core vocabulary words across the sample languages are relatively high. The first

1st order											
Quality	**Agent**	**Location**	**Entitiy**	**Finitive**	**Directional**	**Saturative**	**Inceptive**	**Augmentative**	**Singulative**	**Diminutive**	**Desiderative**
1A pivâk											
1B piteen											
1C pijan											
	1D pijač										
		1E pivnica									
			1F pivo								
			1G pitie								
			1H pijavica								
				1I dopie (si)							
				1J izpie							
					1K vpie (se)						
						1L napie (se)					
						1M opie (se)					
							1N zapie (se)				
							1O razpie (se)				
							1P propie (se)				
								1P nadpie			
								1Q prepie			
									1R otpie		
										1S popie	
										1T pijne	
											1U pie mu se

Figure 1.1: (a) Derivational network, 1st order, Bulgarian verb pie ‘to drink’.

1st order							
Quality	**Resultative**	**Agent**	**Instrument**	**Process**	**State**	**Action**	**Manner**
1A drekkanlegur							
				1B drekking			
				1C drekkandi			

Figure 1.1: (b) Derivational network, 1st order, Icelandic verb drekka 'to drink'.

selection identified 74 nouns, 54 verbs and 31 adjectives in total. Their equivalents in the sample languages were subsequently marked as simple or derived. This left us with 37 simple underived nouns, 12 simple underived verbs and 10 simple underived adjectives. Consequently, the sample of adjectives was 'naturally' identified. The samples of nouns and verbs were reduced to 10 each by eliminating those words that were excessively represented in Swadesh's semantic groups (Swadesh 1955). All in all, the resulting sample of 3x10 words includes only words that are simple, underived and, from a synchronic point of view, actively used in all 40 languages.

(11)

Nouns	**Verbs**	**Adjectives**
bone	cut	bad
eye	dig	new
tooth	pull	black
day	throw	straight
dog	give	warm
louse	hold	old
fire	sew	long
stone	burn	thin
water	drink	thick
name	know	narrow

Importantly, each word was assessed and confirmed as an inherent part of the present-day wordstock of a particular language by an expert morphologist(s) of that language. By implication, the derivational networks based on these 30 simple underived words rely on synchronically productive affixation rules in each of the 40 languages covered in this research. One of the fundamental principles in developing individual derivational networks was the exclusion from the network of any archaic, obsolete, regional, or slang words.

2nd order													
State	**Female**	**Manner**	**Collective + (Entity)**	**Agent**	**Relational**	**Diminutive**	**Action**	**Saturative**	**Entity**	**Instrument**	**Inceptive**	**Abstraction**	**Quality**
2A1 pivkost													
2C1 pijanstvo		2C2 pijano		2C3 pijanica	2C5 pijanski	2C6 pijaničák	2C7 pijanstva				2C8 vpijanči se		
	2D1 pijačka		2D2 pijačka	2C4 pijandur							2C9 opijanči se		
					2H1 pijavičen								
						2J1 poizpie		2J2 doizpie					
						2L1 ponapie (se)			2L3 napitka				
						2M1 poopie se							
						2N1 pozapie se						2N2 zapoj	
						2Q1 poprepie							
						2R1 pootpie							
										2S1 popivka			
						2T1 popijne				2S2 popivatelna	2U1 pripie mu se		2S3 popivatelen
											2U2 dopie mu se		

Figure 1.2: (a) Derivational network, 2nd order, Bulgarian verb pie ‘to drink’.

2nd order	
Quality	**Action**
1A1 ódrekkanlegur	
1A2 ódrekkandi	
	1C1 sídrekkandi

Figure 1.2: (b) Derivational network, 2nd order, Icelandic verb drekka 'to drink'.

3rdorder			
Agent	**Saturative**	**Diminutive**	**Pluriactionality**
	3C7a napijanstva (se)	3C6b popijanstva	
3C4a pijandurnik			
			3J1a izpoizpie
			3L1a izponapie

Figure 1.3: Derivational network, 3rd order, Bulgarian verb pie 'to drink'.

1.3.2 Sample of languages

The sample of 40 European languages was established in two steps. The primary source was the languages covered in Müller et al. (2015–2016). Their number was reduced on the basis of their data availability, i.e. according to the possibility of verifying 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* that includes 12 levels. Only levels 0–4 were taken into consideration because only languages falling within any of these five levels met the abovementioned criteria of representativeness (Table 1.1).

A list of the languages selected is given in Table 1.2.

1st order of derivation, Bulgarian, adjectives	Quality	Entity	State	Process	Inceptive	Resultative	Diminutive	Patient	Instrument	Manner	Temporal	Directional	Privative	Causative	Abstraction	Augmentative
	1	1	2				1	1		1				2		1
	1		1				1	3		1	1					2
			1				1			1		1	1	1		
			1				2							1	1	
	3		1	1			2	2		1				1		2
		1	3	1			1			1						
			2				1	2		1				3		2
	1		2		1	2	1	1		1						
	2		2				1			1				2		
		1	1				1		1							
Maximum derivational network	3	1	3	1	1	2	2	3	1	1	1	1	1	3	1	2

Figure 1.4: Maximum numbers of 1st order derivatives per semantic category – Bulgarian adjectives.

Table 1.1: Expanded Graded Intergenerational Disruption Scale (Ethnologue).

Level	Status	Description
0	International	The language is widely used between nations in trade, knowledge exchange, and international policy.
1	National	The language is used in education, work, mass media, and government at the national level.
2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
3	Wider Communication	The language is used in work and mass media without official status to transcend language differences across a region.
4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.

Table 1.2: Sample languages by language families and by genera (based on WALS).

Indo-European (29)	**Slavic** (9): Bulgarian, Croatian, Czech, Polish, Russian, Serbian, Slovak, Slovene, Ukrainian
	Germanic (8): Danish, Dutch, English, Frisian, German, Icelandic, Norwegian, Swedish
	Romance (7): Catalan, French, Galician, Italian, Portuguese, Romanian, Spanish
	Celtic (2): Irish, Welsh
	Baltic (2): Latvian, Lithuanian
	Greek
Uralic (4)	Estonian, Finnish, Hungarian, North Saami
Altaic (2)	Tatar, Turkish
Nakh-Daghestanian (2)	Chechen, Dargwa
Kartvelian (1)	Georgian
Afro-Asiatic (1)	Maltese
Isolate (1)	Basque

1.3.3 Semantic categories

For the sake of the semantic classification of each derived word, a provisional list of semantic categories was proposed and completed/modified in the course of the project's implementation, taking fine-grained nuances in different languages into consideration. The objective was to preserve a desired level of generalization without losing relevant distinctions (see Appendix 1). The overall semantic theory employed for devising a set of semantic categories is cast in constructionist (Booij 2010) and cognitive linguistic terms, at least in that it recognizes subsymbolic processes in networks as constitutive, since construction relations are obtained at all levels of linguistic patterning. The compiled set contains theory-neutral, cross-linguistically applicable, comparative semantic concepts (in Haspelmath's (2010) sense of the term 'comparative concept'). The semantic categories have been posited regardless of the formal means of their expression in different languages. In keeping with Croft's (2003) recommendations for enhancing cross-linguistic comparability, the semantic comparative categories allow for the examination of "the construction(s) or *strategies* used to *encode*" them in separate languages (Croft

2003: 14). The constructions can be read off the individual networks in each language, while the strategies can be detected in any alternative process employed in languages with poor derivational networks, where compensatory mechanisms, also known as "strategies", are identified.

Since comparative concepts are abstracted from descriptive categories as prototypes, the categories used in the individual chapters have been conceived on the basis of general typological considerations within the limits of typical meanings of affixation patterns in the European languages for which the categories have been posited. Bearing the underdetermination of lexical concepts (Evans 2009; Ludlow 2014) in mind and accepting that language is a complex adaptive system (Beckner et al. 2009) and that meaning in language is a synergetic, emergent phenomenon (Köhler 2011), the generation of meaning in derivational word-formation is a multifactorial process with particular meaning features not attributable to any single specific factor or constituent in a recoverable causal manner. With these preliminaries in mind, the set of semantic categories employed in building the networks are characterized by the following:

(i) they function as canonical points "from which the phenomena actually found can be calibrated" (Corbett 2010: 141) and, in that sense, they target common cross-linguistic specificities, not the peculiarities of individual languages;
(ii) they constitute "a special set of comparative concepts that are specifically created by typologists for the purposes of comparison" (Haspelmath 2010: 663) and are, in that sense, constructs, not part and parcel of the competence of speakers;
(iii) they are heterogeneous in terms of a number of criteria: a) the degree of granularity of the notional categories (in the sense that they combine different numbers of the ontological types discussed below); b) the number of cross-linguistic instantiations; and c) the typicality of individual semantic categories for a specific language;
(iv) their heterogeneity tries to avoid the association of comparative concepts with any specified word-class in any language, as well as distinctions between types of affixes (infixes, superfixes, prefixes, suffixes, etc.) and the associated problems of categorial headedness;
(v) the set of comparative concepts has been extracted from descriptive categories of individual languages. The language-specific categories were used as the lower limit of granularity, while the upper limit was determined by the *ontological types* defined by Cruse in dealing with lexical semantics, i.e. the "fundamental modes of conception that the human mind is presumably innately predisposed to adopt" (Cruse 2000: 49);

(vi) the degree of granularity can at best be illustrated, not explicated – e.g. RESULT OF ACTION is a subcategory of the basic ontological types, combining features of ACTION and STATE. Thus the employed semantic category RESULTATIVE is one level of generality removed from the ontological types. It is at that level that the comparative semantic concepts have been postulated;
(vii) each sense associated with a specific affix is accommodated under a separate semantic category from the set, and thus systematic polysemy in derivation can be captured;
(viii) when two inseparably linked meaning elements are associated with a single affix, two semantic categories are used to classify the derivative within a particular derivational step, e.g. Spanish *ojo* 'eye' > *oj-oso* 'a person having big eyes' labeled with PATIENT +AUGMENTATIVE; and
(ix) the semantic categories are associated with the last derivational step in a series, i.e. with the specific affix attached within this step, and derivatives are discretely arranged in orders of derivation so that the categories do not take into account the resultant lexical meaning of a specific derivative.

The set has been compiled on the basis of both semasiological (extensive reading of analyses of affixation phenomena on the basis of existing, actual words in various European languages) and onomasiological considerations (the onomasiological stance underlies the very cogitation of these concepts designed to incorporate possible words), even though the networks for the individual languages in the volume are based exclusively on existing, attested words.

1.3.4 Construction of derivational networks

Each contributor identified a derivational network for each of the 30 sample words for their language and calculated the saturation values of a derivational network:
a) for each sample word,
b) for the word-class of sample nouns, verbs and adjectives, and
c) for the whole sample of 30 words.

The development of derivational networks faces a number of theoretical problems primarily related to the fuzzy boundary between derivation and inflection (cf., for example, Scalise 1988; Dressler 1989; van Marle 1995; Booij 2006; ten Hacken 2014; Štekauer 2015). For illustration, the fuzzy boundary between past participles and their adjectival homonyms is one of many theoretical problems of

this kind that have not found a unanimous solution either among theoretical morphologists or from a cross-linguistic perspective. Therefore, past participles and the words derived from them have not been included in the derivational networks within this project. Gradation, treated as an inflectional category in traditional grammars, has not been included in the derivational networks either, even if some authors consider it to be a derivational phenomenon. Similar problems bear on the status of combining forms, affixoids, polysemy, and semantic shift. Owing to unequal theoretical approaches to their derivational relevance, it was decided that combining forms be excluded from derivational networks. A list of them was compiled in order to keep to a unified approach across all languages. The same is true of affixoids unless a representative grammar or a reference book explicitly identifies a particular unit as an affix. If a language does not have any basic form, such as English infinitive, and if there are several inflected forms that can serve as the basis for the 1st order of derivation, all of these forms were taken as a single zero-degree base. Any transflexion, transposition, conversion, etc. were excluded from the scope of derivational networks. This fact, certainly, cannot but be reflected in the richness of the derivational networks. So, for example, in Basque, all verb formation is based on conversion. In addition, conversion is also highly productive for nouns and adjectives (cf. Chapter 46, this volume). The networks contain lexical items that are exclusively constructed by affixation processes. Only basic meanings, directly derived in the process of derivation, count. Last but not least, it was found that individual language genera face their own specific problems. These are discussed in brief introductions to the individual language groups.

1.3.5 Evaluation of derivational networks

Saturation value calculations are based on the concept of the *Maximum Derivational Network* (MDN) (see Figure 1.4 above and the relevant text). 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 adjectives by means of the formula in (12):

(12)

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

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

For illustration, the Bulgarian adjective *topâl* 'warm' has 27 derivatives. Its saturation value is obtained as 25: 88 × 100 = 28.41%.

In the 1st order, it produces 7 derivatives. These are related to the 1st order MDN, which is 27. Therefore, its 1st order saturation value is 7:27 × 100 = 25.93%.

This procedure makes it possible to calculate average saturation values for each word-class by orders of derivation (13):

(13) Bulgarian adjectives: Average values of saturation by orders of derivation:
1st order saturation 30.74%
2nd order saturation 18.79%
3rd order saturation 20.00%
4th order saturation 11.67%
5th order saturation 10.00%

As a result, each sample language is characterized in terms of the complexity of its derivational networks. Based on this data, the authors of language-specific chapters comment on the results according to a unified structure defined for all 40 language-specific chapters. That means that each language is evaluated and discussed within a separate chapter. Language genera that are represented by more than one language of this sample are introduced by chapters reflecting problems related to the construction of derivational networks in all the languages of the genus/family.

1.4 Structure of language-specific chapters

The structure of each language-specific chapter is as follows:

(i) A brief description of the word-formation system of a language in order to determine the role of affixation processes in the word-formation system of that language.
(ii) Computation of maximum derivational networks.
(iii) Computation of saturation values for each sample word in each of the three word-classes, i.e. nouns, verbs and adjectives, and for each of the word-classes as a whole.

(iv) Identification of the average number and the maximum number of orders of derivation for nouns, verbs and adjectives.
(v) Computation of the derivational capacity for each word-class in each order of derivation.
(vi) Examination of the correlation between semantic categories and orders of derivation.
(vii) Identification of those semantic categories that are typical of individual orders of derivation. In other words, an answer will be provided for the question 'Are any semantic categories characteristic of a particular order of derivation within a particular word-class of ten basic words?'
(viii) Identification of semantic categories that systematically block any further derivation at individual orders of derivation. While there has been extensive research into the combinability of derivational affixes, the data on the combinability of semantic categories without regard to the specific affixes that realize them may reveal additional explanations for the theory of affix combinations and the blocking effects of affixes.
(ix) Identification of typical combinations of semantic categories as an answer to the question 'Are any combinations of semantic categories typical of derivational networks of a given language?'
(x) Identification of multiple occurrences of semantic categories in a series of derivations from a single basic word.
(xi) Identification of the reversibility of semantic categories as an answer to the question 'Are there typical combinations of semantic categories of the sort AB/BA, meaning that two semantic categories can occur in a reversed order?'
(xii) For languages with small derivational networks, a brief explanation of what this means for the word-formation system of that language and how this paucity of derivatives is compensated for is given.
(xiii) Conclusions.

Any deviations from this structure are due to the absence of a particular phenomenon in a given language. Each of the proposed evaluation criteria that constitute the language-specific structure is designed to provide us with a picture of the derivational potential and its actualization and the nature of the derivational system of individual languages, with the emphasis on the role, function and combinability of semantic categories in the formation of new complex words. In addition, these criteria enable us to identify preferred word-formation strategies in various languages and genera.

The last chapter of this volume compares and evaluates the data for all 40 sample languages by individual parameters and draws conclusions from this analysis.

References

Bagasheva, Alexandra. 2017. Comparative semantic concepts in affixation. In J. Santana-Lario & S. Valera (eds.), *Competing patterns in English affixation*, 33–65. Bern: Peter Lang.

Barlow, Michale & Suzanne Kemmer (eds.). 2000. *Usage-based models of language*. Stanford: CSLI.

Bauer, Laurie. 1997. Derivational paradigms. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 243–256. Dordrecht: Kluwer.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language as a complex adaptive system. In Nick C. Ellis & Diane Larsen-Freeman (eds.), Language as a complex adaptive system. [Special issue]. *Language Learning* 59 (1). 1–26.

Blevins, James, P. 2013. Word-based Morphology from Aristotle to Modern WP (Word and Paradigm Models). In K. Alled (Ed.), *The Oxford Handbook of the History of Linguistics*, 375–395. Oxford: Oxford University Press.

Bonami, Olivier & Jana Strnadová. 2016. Derivational paradigms: pushing the analogy. Paper presented at the SLE conference, Naples, 31 August–3 September.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29 (2). 167–197.

Booij, Gert. 2006. Inflection and derivation. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn., 654–661. Oxford: Elsevier.

Booij, Gert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Boyé, Gilles & Gauvin Schalchli. 2019. Realistic data and paradigms. *Morphology* 29 (2). 199–248.

Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* 82 (4): 711–733.

Corbett, Greville. 2010. Canonical Derivational Morphology. *Word Structure* 3 (2). 141–155.

Cotterell, R., E. Vylomova, H. Khayrallah, C. Kirov & D. Yarowsky. 2017. Paradigm completion for derivational morphology. arXiv:170809151.

Croft, William. 2003 [1990]. *Typology and Universals*, 2nd edn. Cambridge: Cambridge University Press.

Cruse, Alan. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Dokulil, M. 1962. *Tvoření slov v češtině. Teorie odvozování slov*. [Word-formation in Czech. A theory of derivation]. Praha: Nakladatelství československé akademie věd.

Dressler, Wolfgang U. 1989. Prototypical differences between inflection and derivation. *Zeitschrift für Sprachwissenschaft und Kommunikationsforschung* 42. 3–10.

Dressler, Wolfgang U. 2005. Word formation in natural morphology. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word formation*, 267–284. Dordrecht: Springer.

Evans, Vyvyan. 2009. *How Words Mean*. Oxford: Oxford University Press.

Furdík, Juraj. 2004. *Slovenská slovotvorba*. Edited by Martin Ološtiak. Prešov: Náuka.
Gaeta, Livio & Marco Angster. 2019. Stripping paradigmatic relations out of the syntax. *Morphology* 29 (2). 249–270.
Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86 (3). 663–687.
Hathout, Nabil & Fiameta Namer. 2016. Multi-level Paradigm-based Model of Competition in Word Formation. Paper presented at the 17th International Morphology Meeting, Vienna, 18–21 February.
Hathout, Nabil & Fiameta Namer. 2019. Paradigms in word-formation: what are we up to? *Morphology* 29 (2). 153–165.
Horecký, Ján, Klára Buzássyová & Ján Bosák. 1989. *Dynamika slovnej zásoby súčasnej slovenčiny*. Bratislava: SAV.
Kardela, Henryk. 2015. Lexical nests revisited: a cognitive grammar account. In R. Gregová, L. Körtvélyessy, S. Tomaščíková and S. Valera (eds.), A Festschrift for Pavol Štekauer. [Special issue]. *SKASE Journal of Theoretical Linguistics* 12 (3). 292–312.
Köhler, Reinhard. 2011. Laws of language. In Patrick Colm Hogan (ed.), *The Cambridge Encyclopedia of the Language Sciences*, 424–426). Cambridge: Cambridge University Press.
Libben, Gary. 2014. The nature of compounds: a psychocentric perspective. *Cognitive Neuropsychology* 31 (1–2). 8–25.
Ludlow, Peter. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford: Oxford University Press.
Müller, Peter O., Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.). 2015–2016. *Word-Formation. An International Handbook of the Languages of Europe*. Berlin: Mouton.
Pala, Karel & Dana Hlaváčková. 2007. Derivational Relations in Czech WordNet. In J. Piskorski and H. Tanev (eds.), *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*, 75–81. Madison: Omniopress.
Pala, Karel & Pavel Šmerk. 2015. Derivancze – *Deriv*ational *An*alyzer of *Cze*ch. In. P. Král and V. Matoušek (eds.), *Text, Speech and Dialogue. TSD 2015. Lecture Notes in Computer* Science, *vol. 9302*, 515–523. Springer: Cham.
Rodrigues, Alexandra S. & Pedro J. Rodrigues. 2017. Cross-paradigms or word formation patterns in interface: evidence from Portuguese. Paper presented at the ParadigMo Conference, Toulouse, 19–20 June.
Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.
Scalise, Sergio. 1988. Inflection and derivation. *Linguistics* 26. 561–581.
Ševčíková, Magda & Zdeněk Žabokrtský. 2014. Word-Formation Network for Czech. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the 9th International Language Resources and Evaluation Conference (LREC 2014)*, 1087–1093. Paris: ELRA.
Skalička, Vladimír. 2004–2006. *Souborne dilo I–III*. Edited by F. Čermák, J. Čermák, P. Čermák & K. Poeta. Praha: Karolinum.
Swadesh, Morris. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21 (2). 121–137.

Štekauer, Pavol. 2014. Derivational paradigms. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 354–369. Oxford: Oxford University Press.

Štekauer, Pavol. 2015. The delimitation of derivation and inflection. In P.O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, 218–234. Berlin: Mouton.

Ten Hacken, Pius. 2014. Delineating derivation and inflection. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 10–25. Oxford: Oxford University Press.

Van Marle, Jaap. 1995. The unity of morphology: on the interwovenness of the derivational and inflectional dimension of the word. In G. E. Booij & J. van Marle (eds.), *Yearbook of Morphology 1995*, 67–82. Dordrecht: Kluwer Academic Publishers.

Zych, Anna. 1999. *Struktura I semantyka polkich I roszjskich gniazd slowotwórczzch z przyimkami wyjściowymi nazywajacymi barwe*. Katowice: WUS.