

Lívia Körtvélyessy, Alexandra Bagasheva, Pavol Štekauer,
Salvador Valera, Ján Genči

49 Derivational networks in European languages: A cross-linguistic perspective

The previous chapters described the specific features and richness of derivational networks in each of the 40 languages included in our research sample. In this final chapter, these preceding chapters and the 1,200 derivational networks on which they are based serve as an important and rich source of data and observations for drawing relevant cross-linguistic conclusions on the similarities and differences among the languages, as well as those language genera and/or languages that are sufficiently represented in our sample. In particular, we examine and compare the maximum derivational networks (section 49.1), saturation values (49.2), consistency of derivations at the language level (49.3) and at the genera level (49.4), correlations between saturation values and the paradigmatic capacity (49.5), maximum and average numbers of orders of derivation (49.6), numbers of derivatives (49.7), correlations between semantic categories and orders of derivation (49.8), semantic categories with blocking effects (49.9), typical combinations of semantic categories (49.10), multiple occurrences of semantic categories (49.11), reversibility of semantic categories (49.12) and the reasons for structurally poor derivational networks (49.13). The data are evaluated in terms of word-classes and orders of derivation, with a special focus on the role of genera and/or families, morphological types and the nature of the word-formation systems of individual languages. It is hypothesized that each of these five factors has an impact on (the possibility of) the generalization of our data.

49.1 Maximum derivational networks

The parameter of the maximum derivational network (MDN) serves to identify the maximum number of possible derivatives, i.e. the maximum potential number of words derivable from a basic simple word. Given our research sample, it is related to 10 basic underived words selected according to the criteria described in section 1.3.1. The MDN is calculated as a sum total of all the highest numbers of derivatives for a given semantic category from among all 10 sample

Körtvélyessy, Lívia, Bagasheva, Alexandra, Štekauer, Pavol, Valera, Salvador and Genči, Ján. "49 Derivational networks in European languages: A cross-linguistic perspective". *Derivational Networks Across Languages*, edited by Lívia Körtvélyessy, Alexandra Bagasheva and Pavol Štekauer. Berlin, Boston: De Gruyter Mouton, 2020, pp. 485–608. <https://doi.org/10.1515/9783110686630-049>

words (in our case) of a given word-class (cf. Figure 1.4 and the related account of the calculation in section 1.2). This means that it shows the maximum number of derivatives that can theoretically be formed from each simple underived word, thus indicating the derivational potential of simple underived words belonging to the same word-class (with regard to the specific sample).

Analogically to the considerable differences between languages in terms of the number of affixation subtypes (Körtvélyessy et al. 2018) used for the formation of new complex words, substantial differences can be found in the MDNs. They are evident for the orders of both derivation and word-classes, and in what follows they are shown by word-class. Table 49.1 illustrates the situation in the word-class of nouns.

It is exclusively Slavic and Romance languages plus Basque that can be found among the top 10 languages for the 1st order derivation. The 2nd order situation is more varied: the top 10 languages include, in addition to these two main genera, the Uralic languages Hungarian and North Saami. In general, from the 2nd order onwards, the Uralic languages feature high values. Apart from Hungarian and North Saami, Finnish has a high MDN value, too. While the importance of Slavic and Romance languages gradually drops as the order of derivation grows, languages like Georgian and Turkish grow in importance.

The situation in the group of basic adjectives is similar. All top 10 positions in the 1st order of derivation are reserved for Slavic and Romance languages. Interestingly, the MDNs of the basic adjectives in Romance languages drop significantly from the 2nd order, from which point on none of them appear in the top 10. With the growing orders of derivation, the Uralic languages as well as Lithuanian, Georgian and Turkish grow in significance. In the latter two languages, this pattern is consistent with the one found in nouns, too.

The dominance of Slavic languages according to the parameter of MDNs among basic verbs is striking, which is primarily due to an extremely high number of prefixes expressing various shades of *aktsionsart*. In general, the MDNs of Germanic languages are low for all three word-classes. The differences between languages with the highest MDN values and the lowest ones are extremely large – much larger than in nouns and adjectives (Table 49.2).

In addition, Table 49.2 shows that the highest MDN value from among all word-classes and orders of derivation is found for Czech verbs in the 3rd order, where the MDN value is as high as 149. Slovak basic verbs produce 129 potential derivatives in their 2nd order, and the MDN of Serbian adjectives' 2nd order is 100. To remind the reader of the meaning of these values, they indicate the number of derivatives that can potentially be produced from each of the basic words of a particular word-class. The derivational potential in the languages with the highest MDN values is thus enormous.

Table 49.1: Languages with the highest and lowest MDN values, nouns.

Language	1st order		2nd order		3rd order		4th order	
	No. of derivatives	Language	No. of derivatives	Language	No. of derivatives	Language	No. of derivatives	Language
Italian	64	Croatian	76	Hungarian	40	Hungarian	15	
Croatian	63	Italian	70	Czech	33	Georgian	10	
Galician	54	Serbian	59	Croatian	29	Italian	9	
Serbian	51	Czech	57	Georgian	27	North Saami	8	
Danish	9	Icelandic	2	6 languages ^a	0	18 languages	0	
Maltese	8	Danish	1					
Chechen	6	Chechen	0					
Dargwa	4	Dargwa	0					

^aChechen, Danish, Dargwa, English, Frisian, and Norwegian.

Table 49.2: Languages with the highest and lowest MDN values, verbs.

	1st order	2nd order	3rd order	4th order			
Language	No. of derivatives	Language	No. of derivatives	Language	No. of derivatives	Language	No. of derivatives
Serbian	93	Slovak	129	Czech	149	Czech	64
Slovene	68	Serbian	124	Serbian	114	Serbian	55
Slovak	66	Georgian	122	Croatian	94	Slovene	39
Croatian	62	Czech	114	Georgian	74	Croatian	24
Danish	16	Norwegian	12				
Icelandic	15	Icelandic	8				
Norwegian	14	Maltese	5				
Maltese	9	Danish	1				

Table 49.3 provides a review of the average MDN values for all three word-classes by order of derivation. It is evident that the greatest potential for rich derivational networks is offered by basic verbs in each order of derivation. The average values significantly exceed those for nouns and adjectives. Adjectives seem to establish slightly better derivational conditions than nouns.

Table 49.3: Average MDNs for all three word-classes by orders of derivation.

Word-class	1st order	2nd order	3rd order	4th order	5th order
Nouns	28.45	25.89	11.03	2.76	0.53
Adjectives	27.47	26.45	15.03	4.55	0.66
Verbs	36.34	44.87	24.95	8.18	1.50

The distribution of the data can be advantageously represented by boxplots (see Diagrams 49.1–49.3 below). It is obvious that, with a rising order, the data are more scattered in each word-class, including the existence of outliers, i.e. the data which are at an abnormal distance from the median. Thus, the higher the order of derivation, the more scattered the MDN values. This is especially significant for the 3rd order of verbs. The boxplot for verbs also shows that four languages with MDN values above 100 cause a high average value of verbs in the 2nd order. With the exception of one language, these data are still on the whisker in comparison to the 3rd order, where languages with high MDN values are outliers. The number of outlier languages in the 3rd order of verbs is as many as five; in the case of adjectives, there are three outlier languages, and in the case of nouns, none of the languages display an outlier MDN value. This distribution confirms the data for the 3rd order in **Table 49.3** – the high average value for verbs is caused by the languages with an extremely high MDN value in comparison to the median value. The boxplots also show that the number of languages with an MDN above 50 is generally much higher for verbs, especially in the 2nd order of derivation. This confirms our interpretation of the average values in **Table 49.3** – the word-class of verbs has the highest potential for the derivation of new words.

Several important conclusions can be drawn from our previous discussion.

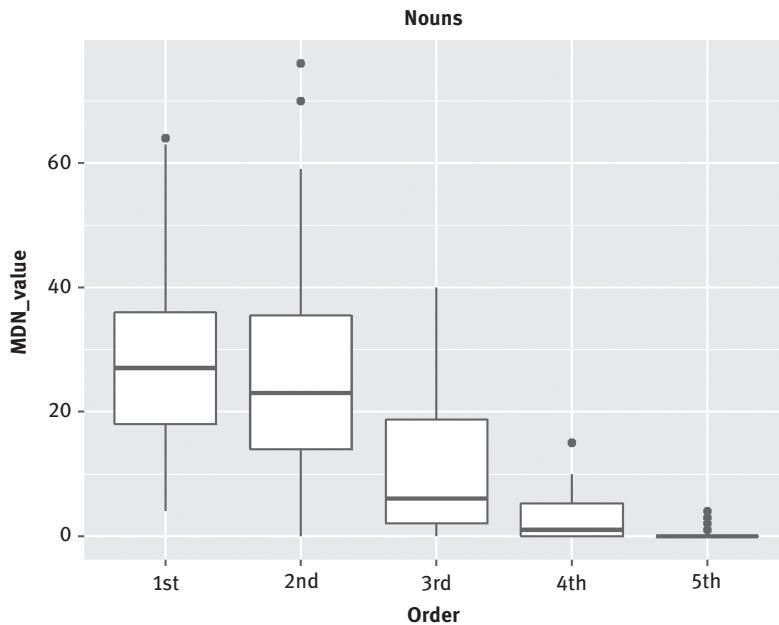


Diagram 49.1: Relation between MDNs and orders of derivation, nouns.

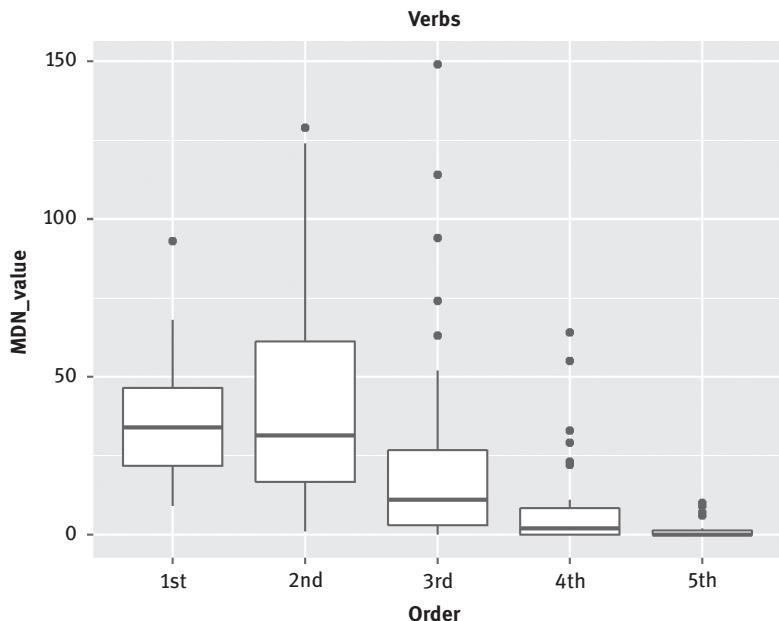


Diagram 49.2: Relation between MDNs and orders of derivation, verbs.

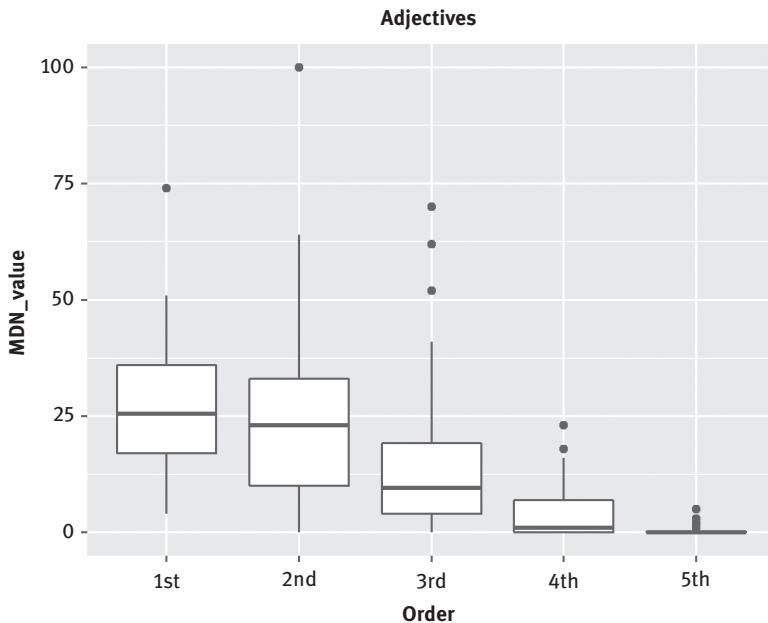


Diagram 49.3: Relation between MDNs and orders of derivation, adjectives.

Summary

- (i) The highest potential for deriving rich derivational networks is clearly bound to simple underived *verbs*. They have the highest MDN value in every order of derivation, and these values are significantly higher than in the other two word-classes.
- (ii) The derivational potential of simple underived nouns and adjectives is *similar*, and in some orders almost identical.
- (iii) The highest MDN values are featured mainly by *Slavic languages* and, in the higher orders of nominal derivation, also some other languages like Hungarian and Georgian.
- (iv) *High MDN values* seem to be typical of *synthetic and agglutinating languages*. However, instances like Dargwa, which has a minimum MDN value, suggest that there is no systematic correlation between the morphological type of a language and the nature of its word-formation system, as also observed by Štekauer (2012). This important finding will be highlighted in several places in this chapter.

- (v) If the orders of derivation are compared, the *highest derivational potential is bound to the first two orders* which, in the case of nouns and adjectives, are very similar. Then, as the orders increase, the possibilities for derivation significantly drop. For verbs, the 2nd order of derivation evidently dominates. The gaps between the average MDN values in individual orders of derivation are significant.

49.2 Saturation values

The saturation value (SV) parameter examines the degree to which the potential, expressed as the MDN value, is *actualized* by the individual sample words used in our research (cf. the explanation and examples in section 1.3.5). This parameter is advantageous in several respects, as it makes it possible to:

- (i) concentrate on individual words and compare the richness of derivation by orders of derivation;
- (ii) compare the saturation values of all sample words of a specific word-class and conclude on the degree of similarity/difference in the derivational richness of simple underived words of the same word-class that belong to the core vocabulary in all languages under research;
- (iii) compare nouns, verbs and adjectives as a whole and find out which of these three word-classes establishes the best derivational opportunities; and
- (iv) evaluate the data from the perspective of groups of languages (genera, families).

49.2.1 Cross-linguistic comparison of individual words by SV

For reasons of space, it is not possible to discuss all 30 words here. Therefore, we have picked out the noun *bone* (including its equivalents in the other sample languages) to illustrate

- a) the range of findings offered by this method of analysis, and, by implication,
- b) the possibilities of evaluating derivational networks inherent in the proposed approach.

The highest SV in the 1st order was identified for Dutch (81.82%). This means that *been*, the Dutch equivalent of *bone*, allows the actualization of over 80% of the derivational potential represented by the corresponding MDN for Dutch nouns. Dutch is followed by Greek (50%), Lithuanian (41.38%) and a group of

13 languages in the range of 30–40%, comprising five Slavic languages, two Germanic and two Romance languages, and Georgian (39.12%), Tatar (38.89%), Maltese (37.50%) and North Saami (34.78%). The language with the lowest SV in the 1st order is Icelandic (8.33%).

The 2nd order is dominated by Greek, which has an SV of as high as 69.57%, followed by French (66.67%), Dutch (60%), German (55.56%) and Slovak (48.89%).¹ The 2nd order for *bone* features much bigger differences in derivation among the sample languages in at least two respects:

- (i) as per the preceding data, the highest SVs significantly exceed those of the 1st order; and
- (ii) unlike the 1st order, there are languages without any derivatives: Icelandic, Tatar, Basque, Dargwa, and Chechen. Furthermore, there are languages with an SV of under 10%: Catalan, Spanish, Welsh, and Georgian.

The SVs for Galician were relatively low in the 1st and the 2nd orders (22.22% and 25.71%, respectively), but this increases substantially in the actualization of the 3rd order possibilities (66.67%). Greek also maintains a high SV level in the 3rd order (47.06%), and Slovak and North Saami are in a similar, but slightly lower range (40%). Apart from the five languages without derivations in the 2nd order, some other languages attest no derivations in the 3rd order, namely Italian, Portuguese, Spanish, Welsh, Latvian, Estonian and Maltese. The derivational capacity of the sample languages dramatically falls in the next orders of derivation.

Taking the total SV into consideration, Dutch is at the top with 65.22%, followed by Greek (56.45%), German (45.45%) and Slovak (41.60%). At the opposite end we find Icelandic, the only language with a total SV under 10%. All the data are summarized in Table 49.4. The languages are listed by language genera/families.

Based on the total SVs for the word *bone*, languages can be divided into three groups. The SVs in the first group (high) are above 40%. In the second group (medium), they range from 40% to 20%. The last group (low) covers languages with saturation levels below 20%. This division is useful for further typological evaluations.

SVs can be projected onto a *saturation map*. In Map 49.1 below, the green colour indicates the highest total SVs for *bone* (Dutch, French, German, Slovak

¹ The topmost position is assumed by Danish with 100%. However, this value follows from the fact that Danish has only one derivative in the 2nd order. Therefore, it is not taken into consideration here.

Table 49.4: SVs of bone and its equivalents in the sample languages.

	Language	1st order	2nd order	3rd order	4th order	5th order	Total saturation value
<i>kost</i>	Bulgarian	36.67	26.09	26.67	0	0	28
<i>kost</i>	Croatian	22.22	23.68	6.9	0	0	19.54
<i>kost</i>	Czech	32.26	24.56	30.3	16.67	50	27.91
<i>kość</i>	Polish	33.33	35.71	12.5	0	0	30.3
<i>кость</i>	Russian	29.41	38.89	23.08	0	0	30.61
<i>кост</i>	Serbian	31.37	33.9	7.41	0	0	26.21
<i>kost'</i>	Slovak	36.17	48.89	40.74	40	0	41.6
<i>kost</i>	Slovene	22.5	24.32	23.53	100	0	25
<i>кістка</i>	Ukrainian	16	23.81	38.89	33.33	0	25.35
<i>ben</i>	Danish	33.33	100	0	0	0	36.36
<i>been</i>	Dutch	81.82	60	0	0	0	65.22
<i>bone</i>	English	30	45.45	0	0	0	35.48
<i>bonke</i>	Frisian	22.22	50	0	0	0	27.27
<i>Knochen</i>	German	45	55.56	33.33	0	0	45.45
<i>bein</i>	Icelandic	8.33	0	0	0	0	6.67
<i>ben, bein</i>	Norwegian	27.78	14.29	0	0	0	24
<i>ben</i>	Swedish	35.71	29.41	16.67	0	0	28.95
<i>os</i>	Catalan	26.83	8.11	25	0	0	18.18
<i>os</i>	French	38.1	66.67	100	0	0	60
<i>óso</i>	Galician	22.22	25.71	66.67	0	0	25
<i>osso</i>	Italian	25	20	0	0	0	18.07
<i>osso</i>	Portuguese	36.11	26.92	0	0	0	31.25
<i>os</i>	Romanian	34.15	32.61	23.81	0	0	30.09
<i>hueso</i>	Spanish	19.57	6.06	0	0	0	13.1
<i>cnámh</i>	Irish	17.65	26.32	33.33	0	0	21.43
<i>asgwrn</i>	Welsh	17.14	6.67	0	0	0	13.73

Table 49.4 (continued)

	Language	1st order	2nd order	3rd order	4th order	5th order	Total saturation value
<i>kauls</i>	Latvian	30.43	37.5	0	0	0	30.95
<i>kaulas</i>	Lithuanian	41.38	39.13	25	0	0	39.29
<i>κόκκαλο</i>	Greek	50	69.57	47.06	0	0	56.45
<i>luu</i>	Estonian	22.58	24	0	0	0	19.4
<i>luu</i>	Finnish	24.24	35.29	19.05	0	0	24.74
<i>dákti</i>	North Saami	34.78	23.68	40	25	0	29.9
<i>cson</i>	Hungarian	25.93	25.58	12.5	20	0	20.16
<i>söyäk</i>	Tatar	38.89	0	0	0	0	16.28
<i>kemik</i>	Turkish	25	20	12.5	0	0	16.44
<i>hezur</i>	Basque	27.78	0	0	0	0	17.86
<i>ликка</i>	Dargwa	25	0	0	0	0	25
<i>da'ahk</i>	Chechen	16.67	0	0	0	0	16.67
<i>dzvali</i>	Georgian	39.29	9.09	3.7	0	0	15.31
<i>għadam</i>	Maltese	37.5	25	0	0	0	30.77

Table 49.5: Languages according to the total SV of bone from highest to lowest SVs.

SV	Languages
HIGH	Dutch, French, Greek, German, Slovak
MEDIUM	Bulgarian, Czech, Danish, Dargwa, English, Finnish, Frisian, Galician, Hungarian, Irish, Latvian, Lithuanian, Maltese, North Saami, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovene, Swedish, Ukrainian
LOW	Basque, Catalan, Chechen, Croatian, Estonian, Georgian, Icelandic, Italian, Spanish, Tatar, Turkish, Welsh

and Greek). The yellow areas are those with the lowest SVs (Croatian, Estonian, Catalan, Italian, Basque, Chechen, Turkish, Tatar, Georgian, Welsh, Spanish, and Icelandic). The areas in blue indicate languages with medium SVs.



Map 49.1: Distribution of languages according to their SV (bone).

Map 49.1 indicates that the languages with the highest SV for *bone* constitute a *homogeneous strip* (with the exception of Greek), stretching from the Atlantic coast to Central Europe.

Another important parameter concerns the number of derivational orders employed in individual languages. In the case of *bone*, there is no language that derives words in all five orders. Five languages derive words in four orders, 18 languages in three orders, 12 languages in two orders and five languages only in one order. It follows from Table 49.6 that the highest number of derivational orders correlates with *Slavic and Uralic languages*, i.e. *synthetic languages*.

Table 49.6: Classification of languages according to the number of orders of derivation (bone).

Order of derivation	Languages
5 orders	—
4 orders	Hungarian, North Saami, Slovak, Slovene, Ukrainian (5 languages)
3 orders	Bulgarian, Catalan, Croatian, Czech, Finnish, French, Galician, Georgian, German, Greek, Irish, Lithuanian, Polish, Romanian, Russian, Serbian, Swedish, Turkish (18 languages)
2 orders	Danish, Dutch, English, Estonian, Frisian, Italian, Latvian, Maltese, Norwegian, Portuguese, Spanish, Welsh (12 languages)
1 order	Basque, Chechen, Dargwa, Icelandic, Tatar (5 languages)

The sample languages significantly differ in the *distribution* of the actualized derivational potential across individual orders of derivation. If we concentrate on the first three orders, Greek's SV is 50% and above in each order. Ukrainian manifests increasing SVs as the order increases (16%, 23.81%, 38.89%) and a high SV even in the 4th order (33.33%). The opposite can be observed in Lithuanian (41.38%, 39.13%, 25.00%). There are languages with a kind of falling-rising SV, such as Catalan (26.83%, 8.11%, 25.00%), languages with a rising-falling SV, for example, Finnish (24.24%, 35.29%, 19.05%), and, importantly, languages that concentrate the derivation exclusively in the first two orders. Examples of this are numerous, including, for instance, Dutch, English, Italian, Spanish, and Estonian. In addition, there are languages that restrict their derivational activities to the 1st order (Icelandic and Tatar).

In the following analysis, the first three orders of derivation are taken into consideration. As a result, the sample of languages is reduced to 23 languages because 17 languages derive words from *bone* only in the first two orders (cf. Table 49.4).

The SVs of *bone* in these languages follow two basic tendencies: the SV either rises (the SV in the 2nd order is higher than the SV in the 1st order) or falls (the SV in the 2nd order is lower than the SV in the 1st order). Furthermore, for both of these basic cases, three patterns can be observed.

Rising (including rising-falling)

14 languages show this pattern. Half of them are Slavic languages. In the first pattern, the lowest SV is in the 1st order, and the highest in the 3rd order. The second and third patterns are rising-falling ones. In both of them, it is the 2nd order of derivation that features the highest SV. They differ in their position of the 3rd order relative to the 1st order.

Falling (including falling-rising)

In contrast to the former pattern, only two languages here are Slavic (Bulgarian and Czech). There are two Romance languages and two Uralic languages (Hungarian and North Saami). As in the previous case, three patterns can be identified: Pattern 1 shows gradually falling SVs, while Patterns 2 and 3 are falling-rising types. In all patterns, the SV in the 1st order of derivation is always higher than in the 2nd order. In Patterns 1 and 2, the SV in the 1st order is the

Table 49.7: Patterns of rising SVs (bone).

	Description	Languages (Total 14)
Pattern 1		French, Galician, Irish, Turkish, Ukrainian (5 languages)
	LOW-MEDIUM-HIGH	
Pattern 2		Croatian, Finnish, German, Greek, Polish, Russian, Serbian (7 languages)
	MEDIUM-HIGH-LOW	
Pattern 3		Slovak, Slovene (2 languages)
	LOW-HIGH-MEDIUM	

Table 49.8: Patterns of falling SVs (bone).

	Description	Languages (Total 9)
Pattern 1		Georgian, Hungarian, Lithuanian, Romanian, Swedish (5 languages)
	HIGH-MEDIUM-LOW	
Pattern 2		Bulgarian, Catalan, Czech (3 languages)
	HIGH-LOW-MEDIUM	
Pattern 3		North Saami (1 language)
	MEDIUM-LOW-HIGH	

highest and in the 2nd order it drops to either the lowest or the medium value. In the last pattern, the highest value is in the 3rd order.

This was an example of the possibilities for an analysis at the level of a single word. A more interesting and more telling analysis would be one that covered all ten words of each word-class.

49.2.2 Cross-linguistic comparison of word-classes by saturation value

Nouns

For each order of derivation, we classified languages into three groups according to their SV (Table 49.9). This perspective shows that the highest average SV in the 1st order for all ten nouns was found in Dutch (51.82%), followed by Greek (39.95%), German (38.00%), Georgian (36.07%), Dargwa (35.00%) and Swedish (34.29%). The SV for Maltese in the 1st order is as low as 16.25%. Dargwa is a special case in this series: it ranks highly in the 1st order, but this is the only order for the derivation of basic simple nouns in this language. German tops the 2nd order with 48.89%, as the SV of Dutch drops significantly to 28.00%. While the 3rd order is dominated by North Saami with 34.40%, German still maintains a high SV in this order (30.00%). Languages with only one order of derivation (Chechen and Dargwa) have high SVs in the 1st order (for obvious reasons). Dargwa belongs to the high SV group and Chechen to the medium group. With the exception of Danish, a similar situation can be observed for other languages with two orders (English, Frisian, and Norwegian).

Verbs

Table 49.10 classifies languages according to the SVs of verb-based derivations. Verb-based derivations are characterized by high SVs across all orders. Thus, Greek dominates the first two orders with values approaching 50% and German the 3rd order with an SV of 40.00%. Even the 4th order's top value is as high as 30%. In general, the SVs of verb-based derivational networks are fairly high. In the 1st order of derivation, there are as many as 20 languages that actualize their potential to more than 30%. In the 2nd order, there are 11 such languages. It is interesting that the top 10 values for the individual orders include genetically diverse languages, i.e. we do not witness the dominance of a specific genus.

Adjectives

The adjective-based derivations show very high top-level SVs. The highest SV in the 1st order reaches 60% (Dargwa), in the 2nd order it is 40.40% (Greek), in the 3rd order it is 48.89% (Greek) and in the 4th order it has the same value as

Table 49.9: Classification of languages according to SV by orders of derivation, nouns.

	High SV	Medium SV	Low SV	No derivatives
1st order	Czech, Dargwa, Dutch, Finnish, Georgian, German, Greek, North Saami, Romanian, Russian, Swedish (11 languages)	Basque, Bulgarian, Catalan, Chechen, Croatian, English, Estonian, Frisian, Galician, Hungarian, Italian, Irish, Latvian, Lithuanian, Norwegian, Polish, Portuguese, Serbian, Slovak, Slovene, Spanish, Tatar, Turkish, Ukrainian, Welsh (25 languages)	Danish, French, Icelandic, Maltese (4 languages)	
2nd order	Czech, Dutch, English, Estonian, Finnish, Frisian, German, Greek, Hungarian, Italian, North Saami, Slovak, Turkish (13 languages)	Bulgarian, Catalan, Croatian, Danish, Georgian, Latvian, Lithuanian, Norwegian, Romanian, Serbian, Swedish, Ukrainian (12 languages)	Basque, French, Galician, Icelandic, Irish, Maltese, Polish, Portuguese, Russian, Slovene, Spanish, Tatar, Welsh (13 languages)	Chechen, Dargwa (2 languages)
3rd order	Croatian, Czech, Finnish, German, Bulgarian, Catalan, Estonian, Greek, Hungarian, Lithuanian, North Saami, Portuguese, Turkish (10 languages)	Georgian, Irish, Italian, Latvian, Polish, Romanian, Serbian, Spanish, Tatar, Ukrainian (9 languages)	Basque, Dutch, French, Galician, Icelandic, Maltese, Slovak, Slovene, Swedish, Ukrainian (15 languages)	Chechen, Danish, Dargwa, English, Frisian, Norwegian (6 languages)

4th	Croatian, German, North Saami order (3 languages)	Bulgarian, Czech, Finnish, Hungarian, Italian, Romanian, Serbian, Slovak, Turkish, Ukrainian (10 languages)	Catalan, Estonian, Georgian, Basque, Chechen, Danish, Dargwa, Polish, Russian, Slovene, Dutch, English, French, Frisian, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Spanish, Tatar, Welsh (20 languages)
5th	Bulgarian, Croatian, Czech, Finnish, Hungarian, North Saami, Slovak, Turkish, Ukrainian order (9 languages)		Basque, Catalan, Chechen, Danish, Dargwa, Dutch, English, Estonian, French, Frisian, Galician, Georgian, German, Greek, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovene, Spanish, Swedish, Tatar, Welsh (31 languages)

Legend:

- 1st order: HIGH = 30–52%; MEDIUM = 20–29%; LOW = <20%
 2nd order: HIGH = 26–50%; MEDIUM = 20–25%; LOW = <20%
 3rd order: HIGH = 20–34%; MEDIUM = 11–19%; LOW = <11%
 4th order: HIGH = 20–27%; MEDIUM = 11–19%; LOW = <11%
 5th order: ALL CATEGORIES = 10%

Table 49.10: Classification of languages according to SV by orders of derivation, verbs.

	High SV	Medium SV	Low SV	No derivatives
1st	Czech, Dutch, French, Georgian, Bulgarian, Catalan, Croatian, order Greek, Maltese, North Saami, Dargwa, English, Estonian, Finnish, Frisian, Galician, German, Hungarian, Icelandic, Italian, Latvian, Lithuanian, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish, Swedish, Tatar, Ukrainian (8 languages)		Basque, Chechen, Danish, Irish, Norwegian, Russian, Serbian, Welsh (8 languages)	
2nd	Croatian, Dutch, Georgian, order German, Greek, Hungarian, Latvian, Lithuanian, North Saami, Serbian, Turkish (11 languages)	Bulgarian, Catalan, Czech, Dargwa, English, Estonian, Finnish, Italian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish, Tatar (16 languages)	Basque, Chechen, Danish, French, Frisian, Galician, Icelandic, Irish, Norwegian, Russian, Spanish, Ukrainian, Welsh (13 languages)	Basque, Danish, Dargwa, Ukrainian (4 languages)
3rd	Croatian, Czech, Dutch, English, order Georgian, German, Greek, Latvian, Lithuanian, North Saami, Serbian, Slovak, Slovene, Turkish (14 languages)	Bulgarian, Catalan, Estonian, Finnish, Galician, Hungarian, Irish, Italian, Polish, Romanian, Russian, Spanish, Swedish, Tatar, Welsh (15 languages)	Chechen, French, Frisian, Icelandic, Maltese, Norwegian, Portuguese (7 languages)	Basque, Danish, Dargwa, Ukrainian (4 languages)

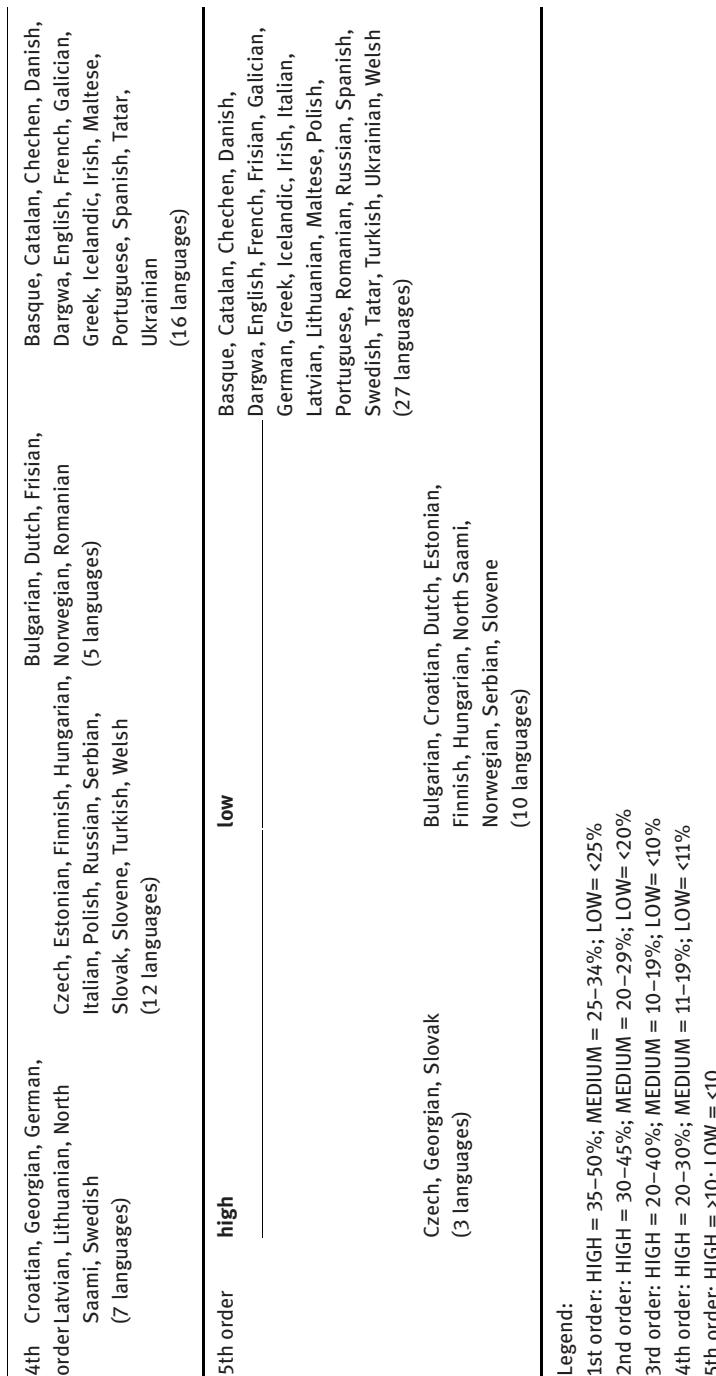


Table 49.11: Classification of languages according to SV by orders of derivation, adjectives.

	High SV	Medium SV	Low SV	No derivatives
1st	Dagwa, Dutch order French, Lithuanian, Maltese (5 languages)	Basque, Bulgarian, Catalan, Croatian, Czech, English, Estonian, Finnish, Frisian, Galician, Georgian, German, Greek, Hungarian, Icelandic, Irish, Italian, Latvian, North Saami, Norwegian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Swedish, Tatar, Turkish, Ukrainian, Welsh (33 languages)	Chechen, Danish (2 languages)	
2nd	Estonian, French, order Georgian, Greek, Hungarian, Lithuanian, North Saami, Turkish (8 languages)	Chechen, Croatian, Czech, Dutch, Finnish, Galician, German, Irish, Italian, Latvian, Maltese, Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Spanish, Swedish, Tatar, Ukrainian (21 languages)	Basque, Bulgarian, Catalan, Danish, English, Frisian, Icelandic, Norwegian, Russian, Welsh (10 languages)	Dargwa (1 language)
3rd	Estonian, Finnish, order Greek, Hungarian, Latvian, Lithuanian, North Saami, Turkish (8 languages)	Bulgarian, Croatian, Czech, Dutch, Georgian, German, Maltese, Tatar (8 languages)	Basque, Catalan, English, Frisian, Galician, Irish, Italian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Swedish, Ukrainian, Welsh (18 languages)	Chechen, Danish, Dargwa, French, Icelandic, Norwegian (6 languages)

4th Greek, Latvian order (2 languages)	Croatian, Estonian, Finnish, Lithuanian, Bulgarian, Catalan, Czech, Georgian, German, Hungarian, Italian, Maltese, Polish, Romanian, Serbian, Slovak, Slovene, Tatar, Ukrainian, Welsh (7 languages)	Basque, Chechen, Danish, Dargwa, Dutch, English, French, Frisian, Galician, Icelandic, Irish, Norwegian, Portuguese, Russian, Spanish (15 languages)
5th Czech, Finnish order (2 languages)	German, Georgian, Turkish (3 languages)	Bulgarian, Croatian, North Saami, Serbian, Slovene (5 languages)
		Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Spanish, Swedish, Tatar, Ukrainian, Welsh (30 languages)

Legend:

1st order: HIGH = 40–60%; MEDIUM = 20–39%; LOW = <20%

2nd order: HIGH = 30–41%; MEDIUM = 20–29%; LOW = <20%

3rd order: HIGH = 30–50%; MEDIUM = 20–29%; LOW = <20%

4th order: HIGH = 30–40%; MEDIUM = 20–29%; LOW = <20%

5th order: HIGH = 20–25%; LOW = <20

in the 1st order – 60% (Dargwa). The highest SV in the 5th order is 25% (Finnish). These high SVs in all five orders also influence the division of languages into high, medium and low groups.

Family-wise, the top position in the 1st order is assumed by diverse languages. The most significant finding of this sort comes from the 3rd order, where half of the top languages are represented by Uralic languages.

Total word-class SV

A more generalized view is given by the *total word-class SV*, i.e. by the total SV per word-class in a given language. It is calculated as a proportion of all actualized derivatives and the maximum derivational network for a given word-class as a whole (10 basic words in our case).

Nouns

Table 49.12 below divides languages into three groups.

Table 49.12: Classification of languages according to total saturation value, nouns.

Total word-class SV	Languages
High (30–40%)	Dargwa, Dutch, German, Greek, North Saami
Medium (20–29.99%)	Bulgarian, Catalan, Chechen, Croatian, Czech, English, Estonian, Finnish, Frisian, Galician, Georgian, Hungarian, Irish, Italian, Latvian, Lithuanian, Norwegian, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Swedish, Turkish, Ukrainian
Low (< 20%)	Basque, Danish, French, Icelandic, Maltese, Polish, Tatar, Welsh

These results can be projected onto the following saturation map (Map 49.2).

It follows from Map 49.2 that there is no homogeneous territory of languages featuring the highest total word-class SV (dark green colour). This group of languages is heterogeneous in terms of their genetic origin: two Germanic languages, one North Caucasian language, one Uralic language, and Greek. The vast majority of languages (27 out of 40) belong to the medium group, suggesting that *the total word-class SV between 20 and 29.99% is characteristic for the word-class of nouns*. In other words, cross-linguistically, the



Map 49.2: Distribution of languages according to their total word-class SV, nouns.

word-class of nouns seems to actualize between 20% and 29.99% of its potential derivational capacity.

Verbs

Table 49.13, featuring the total word-class SVs for verbs, confirms the previous observation that the majority of languages feature medium SVs.

Table 49.13: Classification of languages according to total saturation value, verbs.

Total word-class SV	Languages
High (30–46%)	Dargwa, Dutch, Georgian, German, Greek, Lithuanian, Maltese, North Saami, Turkish
Medium (20–29.99%)	Basque, Bulgarian, Catalan, Croatian, Czech, English, Estonian, Finnish, French, Frisian, Galician, Hungarian, Icelandic, Italian, Latvian, Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Spanish, Swedish, Tatar, Ukrainian
Low (<20%)	Chechen, Danish, Irish, Norwegian, Russian, Welsh

The situation in the word-class of verbs is similar to that of nouns in concentrating the majority of languages in the range of 20–20.99% (25 out of 40 languages). This is again very telling of general patterns in the derivational networks of basic verbs. The topmost group includes the same languages as that of nouns, plus Turkish, Georgian, Maltese and Lithuanian. This means that it is also the case for verbs that the highest SV is characteristic of a heterogeneous group of languages, regardless of their genetic origin or geographical location. As shown in [Map 49.3](#), these languages are scattered across Europe. In contrast, the light brown colour on the map indicates that languages with low SVs are mostly spoken in peripheral European areas and in Eastern Europe.



Map 49.3: Distribution of languages according to their total word-class SV, verbs.

Adjectives

Finally, [Table 49.14](#) and [Map 49.4](#) give the total SVs for adjectives. As with nouns and verbs, the majority of languages belong to the medium value (25 out of 40). With the exception of German, all the other topmost languages from the word-class of nouns feature high SVs in adjectives as well. In this word-class, however, the number of high-SV languages is higher than in nouns or verbs (12), thus leaving merely three languages in the lowest SV group. Nevertheless, they do not constitute a continuous territory. What is analogous with the previous word-classes is the geographically peripheral location (the eastern part of Europe) of the low-SV languages.

Table 49.14: Classification of languages according to total saturation value, adjectives.

Total word-class SV	Languages
High (30–60%)	Dargwa, Dutch, Estonian, French, Greek, Hungarian, Latvian, Lithuanian, Maltese, North Saami, Spanish, Turkish
Medium (20–29.99%)	Basque, Bulgarian, Catalan, Croatian, Czech, English, Finnish, Frisian, Galician, Georgian, German, Icelandic, Irish, Italian, Norwegian, Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Swedish, Tatar, Ukrainian, Welsh
Low (<20%)	Chechen, Danish, Russian

**Map 49.4:** Distribution of languages according to their total word-class SV, adjectives.

Word-class comparison

Tables 49.12–49.14 and Maps 49.2–49.4 enable us to draw some interesting conclusions about the tendencies of the actualization of derivational potential at the level of word-classes:

Summary

- (i) Languages actualize 20–30% of the derivational potential of a word-class. This value indicates the degree of predictability of derivational networks.

This 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.

- (ii) There is a core group of languages that keep high SVs across all three word-classes. They include Greek, Dutch, North Saami and Dargwa. This group might be completed by German, Turkish and Lithuanian, but these have high values in two word-classes and a medium SV in the third word-class.
- (iii) There is no geographically homogeneous territory in which the languages of the topmost SVs are spoken. These languages are of various genetic origins and are scattered across Europe. In contrast, low-SV languages arrange themselves in geographically peripheral areas.

Distribution of SVs across orders of derivation

Patterns of SV distribution I

What is not revealed by the previous analysis is the fact that there are languages with significant differences between SVs in various orders of derivation.

Nouns

Let us take, for example, Dutch. Its SV in the 1st order of derivation is nearly 52%, but in the 2nd order it drops to 28% and in the 3rd order it is only 10%. To take a contrasting case, the SVs for North Saami noun-based derivatives always keep the language in the highest group and, at the same time, they are well balanced (33.04%, 30.79%, 34.40%). Hungarian is pretty consistent too: 12th position in the 1st order (29.26%), 13th position in the 2nd order (27.68) and 14th position in the 3rd order (23.50%). Czech ranks highly in the first two orders with 33.22% and 30.00%, respectively, then its SV drops to 24.24%. The same is basically true of Catalan, but with much lower SVs. Contrary to this, Georgian ranks 3rd in the 1st order with 36.07%, but 21st in the 2nd order (22.73%) and 24th in the 3rd order (14.81%). In general, however, it can be observed that the majority of languages maintain a specific level of SV throughout individual orders of derivation without substantial oscillations. This enables us to identify specific cases in terms of the richness of derivational networks at individual orders of derivation for a homogeneous group of words (simple underived nouns belonging to the core vocabulary, in our case).

As in the case of *bone*, languages can be classified in terms of falling and rising SVs. In the word-class of nouns, the situation is surprisingly homogeneous. 34 languages derive new complex words in three orders of derivation. Out of these 34 languages, 26 follow the same pattern (a falling tendency, Pattern 1).

Table 49.15: Patterns of rising saturation values, nouns.

Description	Languages (Total 6)
Pattern 2 	Estonian, Finnish, German, Latvian, Slovak (5 languages)
Pattern 3 	Turkish (1 language)

Table 49.16: Patterns of falling saturation values, nouns.

Description	Languages (Total 28)
Pattern 1 	Basque, Bulgarian, Catalan, Croatian, Czech, Dutch, French, Galician, Georgian, French, Hungarian, Icelandic, Irish, Italian, Lithuanian, Maltese, Polish, Romanian, Russian, Serbian, Slovene, Spanish, Swedish, Tatar, Ukrainian, Welsh (26 languages)
Pattern 2 	Portuguese (1 language)
Pattern 3 	North Saami (1 language)

Five languages represent the rising pattern (Pattern 2). Each of the three remaining languages belongs to a different type.

Verbs

As in the case of nouns, languages may also differ in the consistency of their SVs. Croatian, for example, keeps its SV above 30% (30.16%, 30.90%, 29.68%) in the first three orders of derivation. In contrast, German has a relatively low SV in the 1st order of derivation (29.71%), but in the 2nd and 3rd orders it rises up to 40.00%. A significant drop can be observed in Spanish (30.91%, 17.93%, 14.00%). Welsh keeps its SVs between 20% and 10% in all three orders of derivation. 27 languages out of 36 follow the falling line, and 26 of them belong to the same pattern (Pattern 1). The second largest group consists of six languages and represents the rising pattern (Pattern 2).

Table 49.17: Patterns of rising saturation values, verbs.

	Description	Languages (Total 9)
Pattern 2		Croatian, Hungarian, Italian, Latvian, Slovene, Turkish (6 languages)
Pattern 3		North Saami, Serbian (2 languages)
Pattern 4		German (1 language)

Table 49.18: Patterns of falling saturation values, verbs.

	Description	Languages (Total 27)
Pattern 1		Bulgarian, Catalan, Czech, Dutch, English, Estonian, Finnish, French, Frisian, Galician, Greek, Chechen, Icelandic, Irish, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Spanish, Swedish, Tatar, Welsh (26 languages)
Pattern 3		Georgian (1 language)

Adjectives

Pattern 1 in the falling line is the most numerous group (22 languages). Interestingly, a new pattern (Pattern 4) occurs in adjectives. The SV starts off low in the 1st order and gradually rises with the derivation order. This situation was found in the cases of Turkish and Greek.

Word-class comparison

The above shows that SVs fall gradually with rising orders of derivation for the vast majority of languages. This occurs in 28 languages for nouns, and in 27 languages for both verbs and adjectives. This suggests that the derivation of fewer words with an increasing order of derivation is independent of the word-class. Moreover, it is Pattern 1, i.e. the gradually falling SV with the increasing order, that evidently dominates: it is present in 26 languages in both nouns and verbs, and 22 languages in adjectives.

Table 49.19: Patterns of rising saturation values, adjectives.

	Description	Languages (Total 6)
Pattern 2		Estonian, Georgian, Slovene (3 languages)
Pattern 3		North Saami (1 language)
Pattern 4		Greek, Turkish (2 languages)

Table 49.20: Patterns of falling saturation values, adjectives.

	Description	Languages (27)
Pattern 1		Basque, Bulgarian, Catalan, Croatian, Czech, English, Frisian, Galician, German, Irish, Italian, Maltese, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Spanish, Swedish, Ukrainian, Welsh (22 languages)
Pattern 2		Hungarian, Latvian, Lithuanian (3 languages)
Pattern 3		Finnish, Tatar (2 languages)

Patterns of SV distribution II

If the focus is laid on the classification of languages according to [Tables 49.12–49.14](#), i. e. into high, medium and low SV groups, and if we restrict our attention to the first three orders of derivation, we can observe the varying behaviours of languages.

Nouns

Some languages behave consistently across the orders of derivation. For example, Czech, Finnish, German, Greek and North Saami always have high SVs, while Bulgarian, Catalan, Latvian, Serbian and Ukrainian are always in the medium group. French, Icelandic and Maltese always feature low SVs. Eight languages

never reach the group of high SVs: in the first three orders of derivation, they feature the patterns medium-low-low (Basque, Galician, Spanish, Tatar, Welsh) and medium-low-medium (Irish, Polish, Slovene). In contrast, 10 languages never drop to the low saturation group. They adhere to one of the following patterns: medium-high-medium (Italian, Estonian, Slovak), high-medium-medium (Swedish, Romanian, Georgian), medium-high-high (Turkish, Hungarian) or medium-medium-high (Croatian, Lithuanian). Table 49.21 gives an overview of all the patterns occurring in nouns.

Table 49.21: Classification of languages by saturation values in the first three orders of derivation, nouns.

Patterns for the first three orders	Languages
H-H-H	Czech, German, Greek, Finnish, North Saami
M-M-M	Bulgarian, Catalan, Latvian, Serbian, Ukrainian
L-L-L	French, Icelandic, Maltese
H-M-M	Georgian, Romanian, Swedish
M-M-L	Basque, Galician, Spanish, Tatar, Welsh
M-L-M	Irish, Polish, Slovene
M-H-M	Estonian, Italian, Slovak
M-H-H	Hungarian, Turkish
M-M-H	Croatian, Lithuanian
H-H-L	Dutch
H-L-M	Russian
M-L-H	Portuguese

Legend: H – high SV, M – medium SV, L – low SV.

Verbs

In this word-class, too, there are languages that maintain a high SV throughout the first three orders of derivation (Greek, Turkish, Georgian, Dutch and North Saami). By contrast, French drops its SV in the 2nd order of derivation so dramatically that it has a low rank in the 2nd and 3rd orders (such that it has no derivatives in the 4th order). Czech oscillates between the high and medium

groups (but with very small differences in SVs), while Maltese drops its SVs continuously: in the 1st order of derivation its SV is high, in the 2nd it is medium, and in the 3rd order it is low.

Table 49.22: Classification of languages by saturation values in the first three orders of derivation, verbs.

Patterns for the first three orders	Languages
H-H-H	Dutch, Georgian, Greek, North Saami, Turkish
M-M-M	Bulgarian, Catalan, Estonian, Finnish, Italian, Polish, Romanian, Swedish, Tatar
L-L-L	Chechen, Norwegian
H-M-H	Czech
M-M-L	Portuguese
M-L-M	Galician, Spanish
M-H-H	Croatian, Latvian, Lithuanian, German
M-M-H	English, Slovene, Slovak
M-H-M	Hungarian
H-L-L	French
M-L-L	Frisian, Icelandic
L-L-M	Irish, Russian, Welsh
L-H-H	Serbian
H-M-L	Maltese

Adjectives

Table 49.23 below illustrates that only one language (Lithuanian) has high SVs in all three orders of derivation. None of the languages feature low saturation in the first three orders. The largest number of languages (12) have medium SVs in the 1st and 2nd orders, dropping into the low group in the 3rd order. The table also shows that while there are languages that are consistent in their SVs (Lithuanian, German, Croatian, Czech), there are also languages with considerable variations in their SVs.

Table 49.23: Classification of languages by saturation values in the first three orders of derivation, adjectives.

Patterns for the first three orders	Languages
H-H-H	Lithuanian
M-M-M	Croatian, Czech, German
M-M-L	Galician, Irish, Italian, Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Spanish, Swedish, Ukrainian
M-L-L	Basque, Catalan, English, Frisian, Russian, Welsh
H-M-M	Dutch, Maltese
M-L-M	Bulgarian,
M-H-H	Estonian, Greek, Hungarian, North Saami, Turkish
M-M-H	Latvian, Finnish, Tatar
M-H-M	Georgian

Comparison of word-classes

The data show that keeping a particular level of SV across the first three orders of derivation partly depends on the word-class and partly on the general derivational potential of individual languages. Two languages maintain the pattern H-H-H in nouns and verbs (Greek and North Saami) and two languages maintain the pattern M-M-M in nouns and verbs (Bulgarian and Catalan). There is no such correspondence between adjectives and the other two word-classes. In addition, there is no language that maintains the pattern L-L-L in at least two word-classes.

There are several prevailingly H languages. As mentioned above, they include Greek and North Saami as well as some other languages with predominantly high SVs across their orders of derivation in all three word-classes, which are Turkish, Czech, German, Hungarian and Lithuanian.

There are 10 prevailingly M languages: Bulgarian, Galician, Italian, Polish, Romanian, Slovak, Slovene, Spanish, Swedish and Tatar. Nine other languages with medium values in two word-classes can also be added to this group: Catalan, Croatian, Georgian, Latvian, Estonian, Finnish, Irish, Portuguese, and Serbian.

There is no prevailingly L language. Two languages have low SVs in two word-classes: low SVs occur in verb-based and adjective-based networks in Russian, and Welsh has medium SVs in nouns and low SVs in verbs and adjectives. The majority of low SVs can be found in languages with fewer than three orders of derivations in all or some word-classes, like Danish, Frisian, Icelandic,

English, Basque and Norwegian. An interesting case is that of French, which has fewer than three orders in adjectives; while it has low SVs in nouns, its verb-based network combines a high SV in the 1st order and low SVs in the 2nd and 3rd orders.

All this discussion can be represented by the following three diagrams, which illustrate the correlation between the number of orders and the total word-class SVs. Since the correlation coefficient for each word-class approaches 0 (nouns = – 0.01, verbs = 0.17, adjectives = – 0.22), there is no correlation between SVs and orders of derivation. The diagrams show that languages tend to have SVs between 20 and 30%.

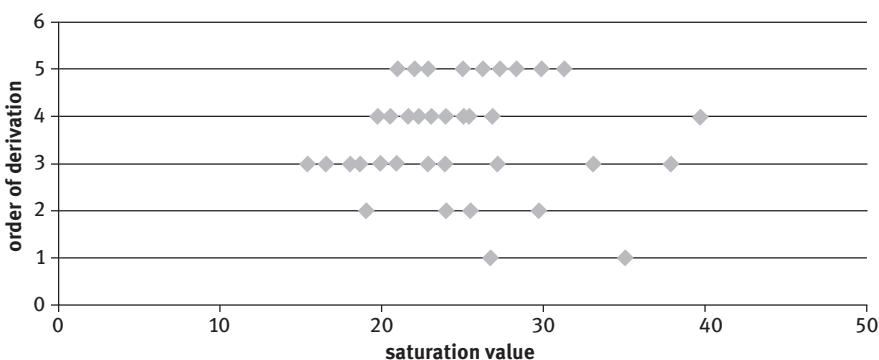


Diagram 49.4: Correlation of saturation values and orders of derivation, nouns.

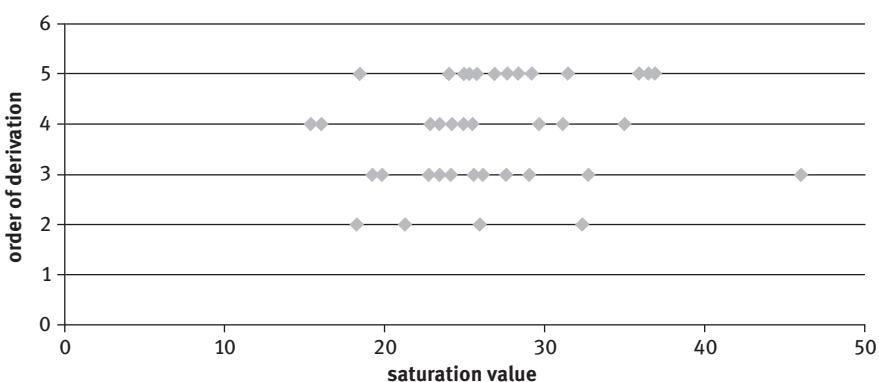


Diagram 49.5: Correlation of saturation values and orders of derivation, verbs.

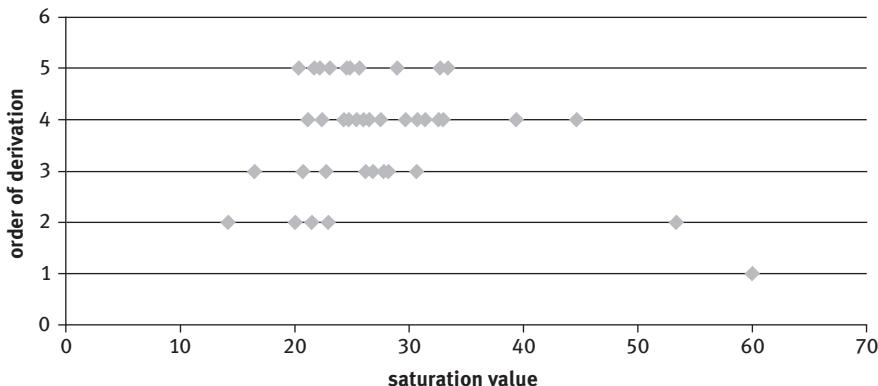


Diagram 49.6: Correlation of saturation values and orders of derivation, adjectives.

Summary

- (i) The fundamental tendency for the majority of languages is the falling tendency, i.e. the saturation value is indirectly proportional to the growing order of derivation. This tendency is independent of the word-class of the basic word.
- (ii) There are very few languages that maintain a consistent level in their saturation values across consecutive orders of derivation. The SV consistency of a particular language depends on the word-class of the basic word.
- (iii) The SVs' consistency across orders of derivation is not affected by the genetic affiliation of a language.
- (iv) In general, there is no correlation between saturation values and orders of derivation.

49.3 Consistency of derivation: Are the derivational networks predictable?

Any discussion of derivational networks should, by necessity, pursue the objective of answering one of the following central questions: are derivational networks predictable? If so, what is the degree of predictability? And what are the factors affecting the predictability of derivational networks? Therefore, further to the discussion in section 49.2, we aim to identify any patterns in saturation values for individual languages and individual orders of derivation in order to

find out whether the saturation values for all 10 words of a given word-class are *consistent*. Consistency implies *predictability*. Certainly, this is not an either/or question. Instead, predictability is a cline, determined by the extent of deviations from the average saturation value. This can be evaluated by means of the parameter of *standard deviation* (SD). If the standard deviation is under 10%, we will consider the derivation within a particular word-class of a given language to be predictable in a given order of derivation. The data enable us to draw the following conclusions.

49.3.1 Nouns

In Table 49.24, languages are divided into three groups according to the SD. The table shows that 13 languages are ranked in the group with SD <10 in the 1st order of derivation. Two languages from this group (Bulgarian and Croatian) occur in the same SD group in the 2nd order as well, this time accompanied by Ukrainian. There is no language with SD <10 in the 3rd order of deviation. Furthermore, Bulgarian and Croatian behave differently in the 3rd order. While Bulgarian smoothly slips into the medium group, Croatian makes a jump into the group with the lowest consistency. The number gradually drops with the growing order.

The SD >20 group shows the opposite tendency. In the 1st order of derivation, there are only three languages with an SD of above 20, but in the 2nd order there are 12 languages, and in the 3rd the number of languages is 13. Dutch occurs repeatedly in this group in each order of derivation. So do Danish and Dargwa. However, Dargwa has only one order of derivation and Danish two.

The medium SD group (10–19.9) shows a dropping tendency too, but the numbers are more balanced (1st order: 24 languages; 2nd order: 20 languages; 3rd order: 15 languages). Seven languages (Czech, Polish, Russian, Serbian, Slovene, Catalan, and North Saami) belong to this group in each order of derivation.

Map 49.5 shows the consistency in the word-class of nouns in the 1st order of derivation. The darker the green colour, the more consistent the SVs in a given language. There is a stretch of areas in dark green running through Europe from north to south. This stretch divides Europe into western and eastern parts. The three light green locations are isolated territories of Danish, Dargwa and Dutch.

Nouns, 1st order

There are 13 languages with SDs under 10.00: Bulgarian, Croatian, Slovak, German, Icelandic, Norwegian, Italian, Welsh, Latvian, Finnish, Hungarian,

Table 49.24: Classification of languages by standard deviation per order of derivation, nouns.

Order	1st	2nd	3rd	
SD				
0.1–9.9	Basque, Bulgarian, Croatian, Finnish, German, Hungarian, Icelandic, Italian, Latvian, Norwegian, Slovak, Turkish, Welsh	Bulgarian, Croatian, Ukrainian		
10–19.9	Catalan, Chechen, Czech, English, Estonian, French, Frisian, Galician, Georgian, Greek, Irish, Lithuanian, Maltese, North Saami, Polish, Portuguese, Romanian, Russian, Serbian, Slovene, Spanish, Swedish, Tatar, Ukrainian	Basque, Catalan, Czech, English, Frisian, Galician, Hungarian, Irish, Italian, Latvian, North Saami, Polish, Portuguese, Russian, Serbian, Slovak, Slovene, Spanish, Turkish	Bulgarian, Catalan, Czech, Finnish, Greek, Hungarian, Italian, North Saami, Polish, Russian, Serbian, Slovak, Slovene, Tatar, Ukrainian	
>20	Danish, Dargwa, Dutch	Danish, Dutch, Estonian, Finnish, French, Greek, Icelandic, Lithuanian, Norwegian, Romanian, Tatar, Welsh	Basque, Croatian, Dutch, Estonian, French, Galician, German, Icelandic, Irish, Latvian, Lithuanian, Portuguese Spanish, Swedish, Turkish, Welsh	
No derivatives		Chechen, Dargwa, Georgian, Maltese, Swedish	Danish, Dargwa Chechen, English, Frisian, Georgian, Maltese, Norwegian, Romanian	

Turkish and Basque. Their SDs range between 6 and 9.9, with Bulgarian featuring the most consistent SV across nouns (6.23). Importantly, no language exceeds an SD value of 20.

Nouns, 2nd order

Only three languages have an SD under 10.00: Bulgarian, Croatian and Ukrainian. Most of the other languages are within the range between 10.00 and 20.00. There are a few values above 20.



Map 49.5: Classification of languages by standard deviation in the 1st order, nouns.

Nouns, 3rd order

None of the languages with a standard deviation under 10 preserve their SV consistency in the 3rd order. The values for the individual nouns are apparently more scattered than in the 2nd order, and much more than in the 1st order.

49.3.2 Verbs

There are 11 languages in the SD <10 group in the 1st order of derivation. Only Serbian remains in the same group in the 2nd order, and there is no language in the 3rd order. As in the case of nouns, there is a falling tendency in terms of the number of languages.

In the least consistent group ($SD >20$), the 1st order is represented by one language (French), while the 2nd order counts seven languages and the 3rd order has 22 languages.

The highest number of languages occurs in the SD 10–19.9 group: 29 languages in the 1st order, 32 languages in the 2nd order and 15 languages in the 3rd order.

None of the languages, with the exception of Serbian, remain in the SD <10 group in the 2nd and 3rd orders of derivation. All languages (with the exception of Georgian) drop to the medium group. Afterwards, they follow one of the following options for the 3rd order:

- (i) they stay in the medium group (Croatian, Estonian, Hungarian);
- (ii) they fall into the SD >20 group (Latvian, North Saami, Turkish, Georgian);
or
- (iii) they do not have a 3rd order of derivation (Basque, Chechen).

French is the only language in the SD >20 group in the 1st order of derivation. In the 2nd and 3rd orders, it belongs to the medium group. Georgian is the only language that gradually falls from the SD <10 to the SD >20 group through the first three orders of derivation. The rest of the languages (Ukrainian, Danish, Dutch, Italian, Dragwa, and Maltese) start in the medium group and drop into the SD >20 group in the 2nd order. Afterwards, they either stay in this group or they do not have derivatives in the 3rd order. Languages in the medium group in the 1st order of derivation stay in the same group in the 2nd order. The prevailing number of Slavic languages, Tatar and Lithuanian also stay here in the 3rd order, while the majority of Germanic and Romance languages drop into the last deviation group, as do the Celtic languages and Greek.

Map 49.6 illustrates the consistency situation in the 1st order of derivation, where darker shades signal higher consistencies. Languages in dark brown are scattered throughout Europe. This is different from the situation with nouns. Nevertheless, the map confirms the observations that consistency in developing derivational networks is related to the genetic affiliation of a language.

Verbs, 1st order

In this case, 10 languages have an SD under 10.00: Serbian, Croatian, Estonian, Hungarian, Turkish, North Saami, Latvian, Basque, Georgian and Chechen. Serbian and Croatian have the lowest SDs (5.95 and 6.12, respectively). However, the SD of as many as 15 other languages falls within the range of 10.00 to 12.00. By implication, the predictability of derivation in this order of verbs appears to be very good, especially since there is only one language whose SD exceeds 20.00 (French: 21.00).

Table 49.25: Classification of languages by standard deviation per order of derivation, verbs.

Order	1st	2nd	3rd
SD			
0.1–9.9	Basque, Chechen, Croatian, Estonian, Georgian, Hungarian, Latvian, North Saami, Serbian, Turkish	Serbian	
10–19.9	Bulgarian, Catalan, Czech, Danish, Dargwa, Dutch, English, Finnish, Frisian, Galician, German, Greek, Icelandic, Irish, Italian, Lithuanian, Maltese, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Slovene, Spanish, Swedish, Tatar, Ukrainian, Welsh	Basque, Bulgarian, Catalan, Chechen, Croatian, English, Estonian, Finnish, Czech, English, Estonian, French, Frisian, Lithuanian, Polish, Finnish, French, Galician, German, Greek, Hungarian, Russian, Serbian, Slovene, Hungarian, Icelandic, Irish, Latvian, Lithuanian, North Maltese, Saami, Norwegian, Polish, Portuguese, Romanian, Spanish, Swedish, Tatar, Turkish, Welsh	Bulgarian, Croatian, Czech, French, Hungarian, Lithuanian, Polish, Russian, Serbian, Slovene, Swedish, Tatar
>20	French	Danish, Dargwa, Dutch, Georgian, Italian, Maltese, Ukrainian	Catalan, Chechen, Dutch, Frisian, Galician, Georgian, German, Greek, Icelandic, Irish, Italian Portuguese, Latvian, Maltese, North Saami, Norwegian, Romanian, Slovak, Spanish, Turkish, Welsh
No derivatives			Basque, Danish, Dargwa, Ukrainian

Verbs, 2nd order

No language is characterized by an SD below 10.00. The vast majority of them have a value between 10.00 and 20.00. High values of above 20.00 have been found for languages that do not have new derivatives in the 2nd order for several basic words (Danish, Ukrainian, Welsh, Dargwa and Maltese).



Map 49.6: Classification of languages by standard deviation in the 1st order, verbs.

Verbs, 3rd order

Like with nouns, none of the SD <10 languages preserve their SV consistency in the 3rd order. The most consistent languages in this order of derivation include Serbian (11.02), Swedish (12.75), Bulgarian (12.86), Slovene (13.01), French (15.27), four other Slavic languages (Czech, Russian, Croatian and Polish), three Uralic languages (Finnish, Estonian and Hungarian), English and Tatar, all of which have an SD value below 20.00.

49.3.3 Adjectives

In the word-class of adjectives, the group with SDs below 10 counts 12 languages in the 1st order of derivation, nine in the 2nd order and one in the 3rd order. This dropping tendency (higher order, fewer languages) is typical also of the medium group (20, 20 and 19 languages in the 1st, 2nd and 3rd orders, respectively). In the last group (above 20), the opposite tendency can be observed: there are two languages in the 1st order, 10 languages in the 2nd order and 20 languages in the 3rd order. Greek displays very consistent behaviour throughout the orders of derivation and is always in the SD <10 group. At the other end of the scale is Dutch, which is always in the >20 group. The behaviour of Slavic languages in the SD <10

group in the 1st order of derivation is interesting. In the 2nd order, they either fall into the medium SD group (Bulgarian, Croatian, Polish, Slovene) or they stay in the SD <10 group. Consequently, they then continue in the medium SD group in the 3rd order. This behaviour is in contrast to non-Slavic languages in the SD <10 group. Greek has been already mentioned, and Catalan and Turkish pass into the medium group in the 2nd order and the SD >20 group in the 3rd order. Hungarian stays in the SD <10 group in the 2nd order and then jumps into the SD >20 group in the 3rd order. Basque descends from the SD <10 group in the 1st order to the SD >20 group in the 2nd and 3rd orders. To sum up, Slavic languages never go lower than the medium group for the 1st and 2nd orders of derivation. By contrast, no Germanic language is present in the SD <10 group.

Table 49.26: Classification of languages by standard deviation per order of derivation, adjectives.

Order	1st	2nd	3rd
SD			
0.1–9.9	Basque, Bulgarian, Catalan, Estonian, Finnish, German, Greek Croatian, Greek, Hungarian, Greek, Hungarian, Polish, Russian, Slovene, Lithuanian, North Saami, Turkish, Ukrainian, Welsh Russian, Ukrainian		
10–19.9	Chechen, Czech, Danish, Bulgarian, Catalan, Bulgarian, Croatian, Dargwa, English, Estonian, Croatian, Czech, English, Estonian, Finnish, German, Finnish, French, Frisian, Frisian, Georgian, Irish, Lithuanian, North Saami, Galician, German, Irish, Italian, Latvian, Maltese, Polish, Romanian, Russian, Italian, Latvian, Lithuanian, Norwegian, Polish, Slovene, Swedish, Tatar, Maltese, North Saami, Romanian, Serbian, Slovak, Ukrainian Norwegian, Portuguese, Slovene, Swedish, Tatar, Romanian, Serbian, Slovak, Turkish Spanish, Swedish, Tatar		
>20	Dutch, Icelandic	Basque, Chechen, Danish, Basque, Catalan, Czech, Dutch, French, Galician, Danish, Dutch, English, Icelandic, Portuguese, Frisian, Galician, Georgian, Spanish, Welsh Hungarian, Irish, Italian, Latvian, Maltese, Portuguese, Serbian, Slovak, Spanish, Turkish, Welsh	
No derivatives	Dargwa		Chechen, Dargwa, French, Icelandic, Norwegian

Map 49.7 below shows how Europe is divided in the 1st order of derivation if the consistency criterion is taken into consideration, where the darkest shade indicates the highest consistency. Western European areas mostly feature a medium level of consistency (SD between 10–19.9). Central Europe is a transition area, and eastern and southern European languages appear to be very consistent.



Map 49.7: Classification of languages by standard deviation in the 1st order, adjectives.

Adjectives, 1st order

Six out of the nine Slavic languages feature very high levels of consistency in deriving new words from basic adjectives. The SD of the remaining three is slightly above 10.00. From among the other languages, values under 10.00 have been found for Catalan, Welsh, Greek, Turkish, Basque and Georgian, i.e. 12 languages in total. The lowest values are for Russian (5.98), Turkish (6.14), Welsh (6.78) and Catalan (6.88). A fairly high level of consistency of derivation from basic adjectives in the 1st order is supported by there being no language with an SD above 20.00.

Adjectives, 2nd order

The standard deviation of five languages is below 10.00: Russian (6.46), Ukrainian (9.17), German (8.46), Lithuanian (9.82), and Greek (9.87). Note that all the other Slavic languages have standard deviation levels slightly above 10.00. As with nouns and verbs, high standard deviations (above 20.00) are mainly (but not exclusively – cf. Spanish and Portuguese, for example) due to

the absence of derivatives within this order for several basic adjectives (e.g. in Danish, Dutch, Icelandic, French, Welsh, Basque, and Chechen).

Adjectives, 3rd order

Lithuanian is the only language with a high level of consistency in the 3rd order (6.68). While the standard deviation of German is also fairly low for this order of derivation (13.60), the derivation of most of Germanic languages is fairly unpredictable, which is especially due to the almost total absence of derivatives in this order. Relatively good values are characteristic of Uralic languages (Estonian, Finnish and North Saami), as well as Tatar, Greek and Romanian.

49.3.4 Comparison of word-classes

1st order

Croatian, Hungarian and Basque are in the most consistent group in the 1st order of derivation in each word-class. Czech, English, Frisian, Swedish, Galician, Portuguese, Spanish, Romanian, Irish, Lithuanian, and Maltese are always placed in the medium group. None of the languages occur repeatedly in the SD >20 group in the 1st order of derivation. The number of highly consistent languages (SD <10) per individual word-class is balanced (nouns – 13 languages; verbs – 10 languages; adjectives – 12 languages).

2nd order

None of the languages occur in the SD <10 group in each word-class. Ukrainian is the only language which can be found twice in this group (with nouns and adjectives). In comparison with the 1st order, the number of languages in the SD <10 group is relatively small: there are only three languages for nouns, one for verbs, and nine for adjectives. Czech, English, Frisian, Catalan, Irish, Latvian and Turkish are always in the medium group for each word-class, while Dutch and Danish are always in the SD >20 group in each word-class.

3rd order

Greek is the only language in the SD <10 group. Bulgarian, Czech, Polish, Russian, Slovene, Finnish and Tatar are always in the medium group. Dutch, Galician, Portuguese, Spanish, Irish, Latvian, Turkish are always in the SD >20 group.

Summary

- (i) Derivational networks are most predictable in the 1st order. This is evidenced by a relatively high number of languages with SDs under 10.00 in all three word-classes (almost one third of all the sample languages), as well as by a relatively high level of consistency of this value in many other languages. This observation runs counter to previous views that suggested that derivational categories are “not part of any well-organized categorial system” (Plank 1994: 1672).
- (ii) The consistency of the results falls as the order increases, which means that the derivational networks are much less predictable in higher orders of derivation.
- (iii) The three languages (Croatian, Turkish and Basque) that are highly consistent in the 1st order of derivation across all three word-classes ($SD < 10.00$) belong to different language families and different morphological types. To this group of languages with consistent derivation, we may add languages with SDs that are below 10.00 in two word-classes and slightly higher in the third class, such as Bulgarian, Polish and Welsh.
- (iv) A high level of consistency across all three orders in all three classes is rare, but it does occur in Bulgarian and Serbian (with the exception of the 3rd order of adjectives in the case of the latter).
- (v) Regardless of the word-class, the most populous category is always the medium SV group in the 1st order of derivation, which counts 26 languages for nouns, 24 languages for verbs and 33 languages for adjectives. These three word-class-based groups overlap in 20 languages. Thus, 20 languages have medium SVs in each word-class in at least two orders of derivation. By implication, a medium SV (20–30%) can be considered the most common pattern.

49.4 The level of genera

49.4.1 Comparison of the genera

This section only includes those genera that are represented by a sufficient number of languages in our sample, i.e. only three Indo-European genera: the Slavic, Germanic and Romance languages. A comparison of average SVs by order of derivation and word-class is summarized in Table 49.27.

Table 49.27 shows a falling pattern in SVs as the order of derivation grows without exception, i.e. in all three word-classes for each genus.

Table 49.27: Comparison of selected genera by order of derivation and word-class.

	Genus	1st order (%)	2nd order (%)	3rd order (%)
Nouns	Germanic	29.94	27.86	17.08
	Romance	26.95	19.26	13.69
	Slavic	27.73	23.27	16.97
Adjectives	Germanic	28.62	22.45	19.40
	Romance	34.13	27.62	13.58
	Slavic	26.93	22.46	17.42
Verbs	Germanic	28.62	22.45	19.40
	Romance	30.72	22.01	13.52
	Slavic	28.73	25.00	20.70

Table 49.27 also shows that there are no striking differences in the average SVs in the 1st order with the exception of adjectives, where the average SV of Romance languages is higher by almost 6%. The SVs in the other cases are remarkably similar, ranging between 26.95% and 30.72, which indicates a general tendency to fill out derivational networks in the 1st order of the Slavic, Germanic and Romance genera. Similar conclusions apply to the 2nd order, where there is only one large difference in average SV, viz. between the Germanic and the Romance languages in noun-based derivations (27.86% vs. 19.26%). The average SV values in the 2nd order tend to be lower than in the 1st order by about 5%. In other words, the derivational richness of the 2nd order of derivation is on average lower than that of the 1st order by 5%. This drops in the 3rd order by an additional 7%.

It appears, then, that, like individual languages, the genera are sensitive to the word-class of basic words and the order of derivation. Thus, while for noun-based derivations the ranking of the language genera in all three examined orders is Germanic, Slavic, then Romance, the situation for adjective-based derivations changes by order of derivation: the Romance genus dominates in the 1st and the 2nd orders, but its average SV in the 3rd order is lowest of all. Similarly, for verb-based derivations, the Romance genus dominates in the 1st order, but has by far the lowest average SV in the 3rd order. The 2nd and 3rd orders are dominated by the Slavic genus.

49.4.2 Inside the genera

The analysis in section 49.4.1 shows a fairly high degree of consistency in the average saturation values of the three Indo-European genera as a whole. A look inside the individual genera, however, provides us with a less consistent picture. Taking the Slavic genus as an example, the difference between the highest SV (Czech: 33.22%) and the lowest SV (Slovene: 23.00%) in the 1st order of nouns is over 10%. The same is true of the 2nd order: Czech and Slovak (both 30%) exceed Polish by almost 15%. A similar difference between the highest SV (Czech) and the lowest SV (Russian) can also be found in the 3rd order. These findings raise the question of whether this level of variance is a general phenomenon across the three genera in question. The answer to this can be found in [Table 49.28](#).

Table 49.28: Differences between the languages with the highest and lowest SVs.

	Genus	1st order (%)	2nd order (%)	3rd order (%)
Nouns	Germanic	32.93	28.89	20.00
	Romance	15.09	10.29	10.00
	Slavic	10.22	14.29	12.89
Adjectives	Germanic	25.00	14.62	20.00
	Romance	31.94	21.56	9.17
	Slavic	9.63	12.91	14.39
Verbs	Germanic	17.25	30.00	30.00
	Romance	5.08	11.91	7.75
	Slavic	22.63	16.15	14.49

Languages without a 3rd order of derivation are disregarded here.

[Table 49.28](#) shows us that the SV differences between individual languages are considerable in all orders of derivation and in all three word-classes. This is not surprising if one realizes that the word-formation system of each language of a particular genus developed differently throughout its history, as evidenced by the degree of diversification² calculated, *inter alia*, for the three

² “This parameter indicates to what degree the WF systems of genetically related languages differ in their structural richness. By implication, it enables us to compare language genera/

above-discussed Indo-European genera in order to identify the degree of diversification of word-formation systems of languages belonging to the same genus (Körtvélyessy et al. 2018).

Summary

- (i) There is an evident tendency for all language genera SVs: they fall in all three word-classes as the order of derivation grows without exception (for the examined genera).
- (ii) The average SVs for the examined genera are remarkably similar and indicate that each of the genera actualizes about 25–30% of the potential in the 1st order, about 22% in the 2nd order, and about 13–18% in the 3rd order.
- (iii) The genera tend to maintain very similar average SVs across the word-classes.
- (iv) A different picture is obtained by comparing SVs for languages inside the genera: the differences are considerable in all three word-classes and all three orders of derivation.

49.5 Correlation between the saturation value and the paradigmatic capacity

An interesting piece of information about derivational networks from the cross-linguistic point of view is the (non-)existence of *correlation* between two variables: the *saturation value* and the *paradigmatic capacity*. Let us recall that the saturation value is calculated as the proportion between the number of actual derivatives and the maximum (i.e. potential) derivational network. The paradigmatic capacity is determined by the number of derivatives from the word-formation base in a particular order of derivation. The question is whether there is any relation between the number of words derivable from the basic simple word in the 1st order of derivation and the saturation value in the next orders. As is generally known, a correlation reflects the interdependence of two functions or data sets. A value of 1 therefore means that one data set copies the other one, suggesting a maximum correlation. Growing values in one set imply

families in terms of the degree of diversification of their WF systems from their protolanguage” (Körtvélyessy et al. 2018: 315).

proportionally growing values in the other. A value of -1 means indirect proportionality, and zero implies no interdependence.

The correlation value for the 1st order is necessarily 1, because we calculate the correlation of the 1st order saturation by means of the 1st order data. Since the SVs in the 2nd, 3rd, 4th and 5th orders are related to the 1st order data, the correlation values indicate a trend of the respective order's saturation relative to the 1st order data.

49.5.1 Nouns

Taking the whole sample of languages into consideration, one cannot observe any systematic correlation. Languages differ from one another significantly in this respect, ranging from 0.92 (Estonian) down to 0.07 (Basque) in the 2nd order, from 0.88 (Estonian) to 0.04 (Spanish, Polish) in the 3rd order, etc.

22 languages feature a correlation value above 0.5 in the 2nd order of derivation, compared to only five languages in the 3rd order, one language (Estonian) in the 4th order and no languages in the 5th order. In each order, the highest correlation value (0.92, 0.88, and 0.85, respectively) is found for Estonian. Georgian and Turkish are the only other languages in which the correlation exceeds the value of 0.5 in both the 2nd and the 3rd orders. Correlations can differ substantially for the same language in different orders of derivation. For illustration, while the correlation of French nouns in the 2nd order is 0.87, it is only 0.38 in the 3rd order.

Indirectly proportional correlation is rare. While there are several languages with minus values and their number increases with the order of derivation from the 2nd order to the 4th order, the only significant values are for Ukrainian (-0.66) in the 3rd order, Ukrainian (-0.57) and Slovak (-0.53) in the 4th order, and Slovak (-0.54) in the 5th order.

49.5.2 Verbs

Verbs are characterized by the highest number of languages with 2nd order correlation values above 0.5 (28 languages). Latvian, Lithuanian, Dargwa and Polish exceed the correlation value of 0.9. Nine languages exceed the correlation value of 0.5 in the 3rd order, with Lithuanian at the top. Welsh, Georgian and North Saami also feature high correlation values in each of these three orders of derivation. This contrasts to their values in the other two word-classes. The correlation of, for example, Welsh in the 2nd order of adjectives is as low as 0.24, and in the 3rd and the 4th orders it is -0.10.

49.5.3 Adjectives

While the observation of considerable differences among languages' correlation values in each order of derivation has been confirmed for adjectives as well, what strikes one at first sight is that the correlations may differ significantly for different word-classes within the same language. While Estonian clearly dominates this parameter for nouns, its correlation values for adjectives are much lower for individual orders (0.55, -0.35, -0.12). The same is true of, for example, Serbian: its values for nouns in the first two orders (0.87, 0.53) are much higher than those for adjectives (0.58, 0.23).

The highest number of languages with a correlation above 0.5 can be found in the 2nd order of adjectives (22 languages). In this case, Slovak manifests the strongest correlation (0.87), followed by Croatian (0.83) and Czech (0.81). In the 3rd order, there are only three languages above 0.5 (Croatian 0.64, North Saami 0.56, Galician 0.53). In the 4th order, there are also three languages above 0.5 (Ukrainian 0.69, North Saami 0.63, Maltese 0.62).

49.5.4 Word-class comparison

Taking all three word-classes into consideration, only one language exceeds the correlation value of 0.50 in the 5th order of derivation: North Saami, with 0.68 for adjectives. North Saami is also an example of a language with high correlation values that are almost constantly above 0.50, with the exception of the 3rd and 4th orders of nouns.

Summary

- (i) No significant correlation seems to exist between saturation value and paradigmatic capacity for our sample of languages as a whole.
- (ii) The correlation depends on the interplay of three factors:
 - the specific language
 - the word-class of the basic word
 - the order of derivation

Generalizations are therefore difficult to make, and no clear tendencies emerge from the data in relation to a possible correlation between saturation value and paradigmatic capacity.

49.6 Maximum and average number of orders of derivation

Languages differ in the number of possible affixes attached to the basic word. This section examines the possibility of generalizing the number of orders of derivation with regard to the specific word-class of basic words, the morphological type of a specific language, the nature of its word-formation system, as well as its genetic affiliation.

49.6.1 Nouns

As it follows from Table 49.29, only nine languages derive nouns in five orders, including five Slavic, two Uralic, and Turkish. All these languages are usually described as synthetic languages, and three of them are agglutinating (Finnish, Hungarian, Turkish).

Table 49.29: Classification of languages according to the maximum number of orders of derivation, nouns.

No. of orders	Languages
5 orders	Bulgarian, Croatian, Czech, Finnish, Hungarian, Slovak, Turkish, Ukrainian (8 languages)
4 orders	Catalan, Estonian, Georgian, German, Italian, North Saami, Polish, Romanian, Russian, Serbian, Slovene, Swedish, Tatar (13 languages)
3 orders	Basque, Dutch, French, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Portuguese, Spanish, Welsh (13 languages)
2 orders	Danish, English, Frisian, Norwegian (4 languages)
1 order	Chechen, Dargwa (2 languages)

At the other end of the scale are two Nakh-Daghestanian languages (Chechen, Dargwa) with only one order of derivation. Two orders of derivation are available in Danish, English, Frisian and Norwegian. All of them are Germanic languages with isolating morphology.

49.6.2 Verbs

Each language has at least two orders of derivation. Basque, Danish, Dargwa and Ukrainian do not make use of more than two orders. The languages are genetically and geographically distant, and they also differ morphologically. Thus, no association between the number of orders of derivation and their classification can be observed. On the other hand, all the other Slavic languages in our sample make use of more than three orders of derivation. One of them has six orders of derivation, while five of them have five orders of derivation. All agglutinating languages in our sample have five derivational orders.

Table 49.30: Classification of languages according to the maximum number of orders of derivation, verbs.

No. of orders	Languages
6 orders	Slovak (1 language)
5 orders	Bulgarian, Croatian, Czech, Dutch, Estonian, Finnish, Georgian, Hungarian, Norwegian, Serbian, Slovene, Turkish (12 languages)
4 orders	Frisian, German, Italian, Latvian, Lithuanian, North Saami, Polish, Romanian, Russian, Swedish, Welsh (11 languages)
3 orders	Catalan, Chechen, English, French, Galician, Greek, Icelandic, Irish, Maltese, Portuguese, Spanish, Tatar (12 languages)
2 orders	Basque, Danish, Dargwa, Ukrainian (4 languages)
1 order	No languages

49.6.3 Adjectives

There is one language with only one order of derivation (Dargwa). Five languages have two orders of derivation, and nine languages have three orders. The biggest group is represented by languages with four orders of derivation. Five orders of derivation were identified in ten languages.

The low number of orders of derivation in Dargwa is balanced out by a high SV in the 1st order. It is the highest SV in this word-class (60%). A low number of orders compensated with a high SV in the 1st order can also be observed in French. French has only two orders of adjective-based derivations. The SV in the 1st order is the second highest (57.14). On the other hand, three

other languages with a low number of orders of derivation have low SVs in both orders of saturation, e.g. for Danish it is 15% in the 1st order and 10% in the 2nd. This means that it is not possible to speak about a general tendency to compensate for a low number of orders of derivation with a high SV in these orders. Analogically, there is no strict correlation between a high number of orders of derivation and high SVs. While, for example, Turkish, Czech and North Saami, with five orders of derivation in each word-class, are characterized by high SVs in the first three orders in all three word-classes, Greek, with high SVs across word-classes and orders of derivation, has only three orders for nouns and verbs and four for adjectives. From this, it follows that SVs do not necessarily correlate with the number of derivational orders.

Table 49.31: Classification of languages according to the maximum number of orders of derivation, adjectives.

No. of orders	Languages
5 orders	Bulgarian, Croatian, Czech, Finnish, German, Georgian, North Saami, Serbian, Slovene, Turkish (10 languages)
4 orders	Catalan, Estonian, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Romanian, Slovak, Swedish, Tatar, Ukrainian, Welsh (15 languages)
3 orders	Basque, Dutch, English, Frisian, Galician, Irish, Portuguese, Russian, Spanish (9 languages)
2 orders	Chechen, Danish, French, Icelandic, Norwegian (5 languages)
1 order	Dargwa (1 language)

Examples of words with six and five orders of derivation, ranging over all three word-classes, are given in examples (1) and (2), respectively.

(1) *6 orders*

Slovak: *pri-s-ťah-ov-alec-k-y*
 DIR-REMOVE-pull-DURATIVE-AGENT-?-MANNER
 ‘immigration.ADV’

(2) *5 orders*

Noun-based
 Ukrainian

Ha-йм-ен-ува-иң-ячк-о

RESULTATIVE-name-CAUS-CAU-CAUS-DIM-INFLECTION

‘nice, short name’

Hungarian

meg-szem-ély-es-ít-és

RESULTATIVE-eye-NOMINALIZER-ATTRIBUTIVE-CAUSATIVE-NOMINALIZER

‘personalization’

Finnish

päiv-it-t-y-mättö-myys

day-CAUS-CAUS-ANTICAUS-PRIVATIVE-STATE

‘being impossible to be updated by itself’

Verb-based

Dutch

on-weet-en-schap-elijk-heid

NEG-know-INFINITIVE-NOM-ADJ-NOM

‘unscientificness’

Estonian

tead-v-us-ta-matu-s

know-ABILITY.ADJ-ABILITY.NOUN-CAUSATIVE-PRIVATIVE-STATIVE

‘unconsciousness, subliminality’

Hungarian

meg-ad-ó-z-tat-ás

RESULTATIVE-give-NOMINALIZER-DURATIVE-CAUSATIVE-NOMINALIZER

‘taxing.N’

Adjective-based

Czech

nej-úzk-ost-n-ě-ji

AUGMENTATIVE-narrow-ABSTRACTION-RELATIONAL-MANNER-AUGMENTATIVE

‘most anxiously’

Georgian

ga-mo-u-sc'or-eb-l-oba

ACTION-ACTION-PRIVATIVE-straight-RESULT -ABSTRACTION

‘irreparability’

North Saami

njulg-e-st-adda-goahti-n

right/straight- CAUSATIVE-SUBITIVE-ITERATIVE- INCHOATIVE-ACTION

'straightening (out) quickly several times'

49.6.4 Comparison of word-classes

Our data show that the maximum number of orders of derivation, i.e. the maximum number of affixes attached to a simple underived base, is six. Five orders occur with all three word-classes, including eight languages for noun-based affixation, 10 languages for adjective-based affixation, and 12 languages for verb-based affixation. By implication, the verbal base appears to be the most productive source of affixation.

Five orders of derivation have been identified in all three word-classes for Bulgarian, Croatian, Czech, Turkish and Finnish, i.e. none of the Germanic or Romance languages in the sample produce affixation chains of a maximum length in all three word-classes. German and Danish can produce five-affix words from adjectives, and Norwegian also can from verbs. No Romance language can derive words with five affixes. On the other hand, Romance languages are very consistent in the number of orders of derivation: in all three word-classes, they have either three or four orders of derivation. The only exception is French with two orders in the case of adjectives. This word-formation feature is dominated by Slavic and Uralic languages. At the opposite end of the scale, there are two languages that can attach only one affix to the basic noun (Chechen, Dargwa), while English, Frisian and Norwegian can attach two affixes. With adjectives, Dargwa is a single-affixation language; Chechen, French, Icelandic and Norwegian cannot attach more than two affixes. With verbs, only two affixation steps are possible in Dargwa, Basque and Ukrainian.

These results are confirmed by the average numbers of affixation steps. Out of 11 languages with an average value of three or more, seven are Slavic, three are Uralic, and then there is also Turkish. A similar dominance has also been found for adjectives and verbs.

Verbs and adjectives serve as the most prolific starting point for affixation processes, having on average 2.78 and 2.76 affixation steps, respectively, for the whole language sample. These values are higher than the average of 2.46 affixation steps for nouns.

Obviously, the highest number of languages can be found in the 1st order of derivation, and the lowest number in the 5th order. The number of languages that can derive new words in individual orders of derivation are more or less

Table 49.32: Number of languages per order of derivation.

Order of derivation	1st	2nd	3rd	4th	5th
Nouns	40	38	34	20	8
Verbs	40	40	36	24	12
Adjectives	40	31	34	25	10

equally distributed in terms of word-classes. For example, in the 1st order, each of the 40 sample languages derives new words; in the 3rd order, the difference is merely two languages; and in the 4th and 5th orders, it is four languages. This balanced representation of languages for each order of derivation in each word-class is violated by adjectives in the 2nd order.

The sample languages were divided into morphological types. According to Sapir (1921), two criteria can be applied: the index of synthesis and the index of fusion. While the index of synthesis considers how many morphemes a word is built from, the index of fusion focuses on the technique or the process of building words. In our case, the number of derivational orders expresses the number of attached affixes. Consequently, we can discuss the index of synthesis and its relation to the number of derivational orders. The morphological types of individual languages were identified on the basis of various sources (Müller et al. 2015–2016; Štekauer et al. 2012; *Ethnologue*). We are aware that this kind of classification cannot be precise, because it is not possible to match one language as a whole with one type. In our approach, we take the type of prevailing feature as a determinant of the morphological language type.

Given these methodological restrictions, three morphological types were identified in our sample: 25 inflectional languages (Bulgarian, Catalan, Croatian, Czech, Dutch, French, Galician, German, Greek, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Ukrainian, Welsh); 10 agglutinating languages (Basque, Dargwa, Estonian, Finnish, Georgian, Hungarian, Chechen, North Saami, Tatar, Turkish); and five analytical languages (Danish, English, Frisian, Norwegian, Swedish). Based on the maximum number of orders of derivation, the sample languages were divided again into three further groups:

- (i) languages with a high number of orders (4–5);
- (ii) languages with a medium number of orders (3); and
- (iii) languages with a low number of orders (1–2).

Table 49.33: Classification of languages by the number of orders of derivation and word-classes.

	Nouns	Verbs	Adjectives
HIGH	Bulgarian, Catalan, Croatian, Czech, Estonian, Finnish, Georgian, German, Hungarian, Italian, North Saami, Polish, Romanian, Russian, Serbian, Slovak, Slovene, Swedish, Tatar, Turkish, Ukrainian (21 languages)	Bulgarian, Croatian, Czech, Bulgarian, Catalan, Dutch, Estonian, Finnish, Frisian, Georgian, German, Hungarian, Italian, Latvian, Greek, Hungarian, Italian, Lithuanian, North Saami, Polish, Romanian, Russian, Serbian, Slovak, Slovene, Swedish, Turkish, Welsh (23 languages)	Bulgarian, Croatian, Czech, Estonian, Finnish, Georgian, German, Hungarian, Italian, Latvian, Lithuanian, Maltese, North Saami, Polish, Romanian, Russian, Serbian, Slovak, Slovene, Swedish, Tatar, Turkish, Ukrainian, Welsh (26 languages)
MEDIUM	Dutch, French, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Portuguese, Spanish, Welsh (12 languages)	Basque, Catalan, English, French, Galician, Greek, Icelandic, Irish, Maltese, Portuguese, Spanish, Tatar (12 languages)	Basque, Dutch, English, Frisian, Galician, Irish, Portuguese, Spanish (8 languages)
LOW	Basque, Chechen, Danish, Dargwa, English, Frisian, Norwegian (7 languages)	Danish, Dargwa, Chechen, Norwegian, Ukrainian (5 languages)	Chechen, Danish, Dargwa, French, Icelandic, Norwegian (6 languages)

The following languages are always in the HIGH group:

- (i) Inflectional: Bulgarian, Croatian, Czech, German, Italian, Polish, Romanian, Serbian, Slovak, Slovene
- (ii) Agglutinating: Estonian, Finnish, Georgian, Hungarian, North Saami, Turkish
- (iii) Analytic: Swedish

These languages can be completed with those which occur in the MEDIUM group only once, e.g. Catalan features a high number of derivational orders for nouns and adjectives, and a medium number for verbs. All these languages are either inflectional (Catalan, Latvian, Lithuanian, Russian, Welsh) or agglutinating (Tatar).

Based on this summary, it can be concluded that a high number of derivational orders is associated with inflectional and agglutinating types of languages. There are four groups of exceptions to this:

- (i) There are four inflectional languages (Galician, Irish, Portuguese, Spanish) which are MEDIUM in each word-class. Interestingly, three of them are Romance languages. Altogether, there are six Romance languages in our sample. While two of them follow the above-mentioned tendency specified for the relation between inflectional languages and the number of derivation orders, four do not. Obviously, Romance inflectional languages behave differently from Slavic inflectional languages. By implication, it is not only the morphological type of a language but also its genetic affiliation that influences its derivational nature.
- (ii) There are four other inflectional languages that do not have a high number of derivational orders: Icelandic, Dutch, Ukrainian and Greek. Dutch and Greek occur in the MEDIUM group twice and once in the HIGH group, so they are still very close to meeting the criteria for the above-mentioned tendency. The same is true of Ukrainian, which occurs twice in the HIGH group and once (verbs) in the LOW group. Icelandic is in the MEDIUM group twice and once in the LOW group.
- (iii) There are three agglutinating languages (Basque, Dargwa and Chechen) that feature very low numbers of derivational orders (Dargwa is always in the LOW group; Basque and Chechen are in the LOW group twice and once in the MEDIUM group).
- (iv) Swedish is usually described as an analytical language. In our sample, it is the only analytical language that features a high number of derivational orders. The remaining four analytical languages are not consistent in their behaviour: Norwegian has a LOW number of derivational orders for nouns and adjectives but a HIGH number for verbs; English occurs twice in the MEDIUM group (verbs, adjectives) and once in the LOW group; Frisian is HIGH for verbs, MEDIUM for adjectives and LOW for nouns; and finally, Danish is always LOW. Interestingly, all analytical languages are in the LOW group for nouns. If they are ever in a HIGH group (Frisian and Norwegian), it is for verbs.

Summary

- (i) The maximum number of orders of derivation in our sample of languages is six.
- (ii) In most languages with five orders of derivation, the 5th order is saturated in the word-class of verbs. In general, the verbal base correlates with a high number of orders of derivation.

- (iii) If the average values of derivational orders across all three word-classes are taken into consideration, the verb-based and adjective-based networks are significantly richer than the noun-based networks.
- (iv) Inflectional and agglutinating languages tend to have a high number of derivational orders, which naturally follows from the ‘synthetic’ nature of their morphological systems.
- (v) Genetically, this corresponds to Slavic and Uralic languages. None of the Germanic or Romance languages reach the level of five orders of derivation.
- (vi) While some languages compensate for their low number of orders of derivation with a high SV in the 1st order, this is not a hard and fast rule.
- (vii) Conversely, high SVs are also present in languages with a high number of orders of derivation.
- (viii) The lack of absolute homogeneity of languages belonging to the same morphological type can also be accounted for by: (a) the unequal development of the word-formation systems of languages belonging to the same language genus (see above, section 49.4.2); and (b), as demonstrated by Štekauer (2012), the typology of word-formation systems does not correspond with the morphological typology. In other words, “there does not seem to be any strong and systematic correlation between the traditional morphological classification and the use of word-formation rules in these languages” (Štekauer 2012: 725).

49.7 Number of derivatives

The number of derivatives shows the extent of the actually realized potential of derivational networks. As such, it is an important indicator of their richness.

49.7.1 Average number of derivatives

Table 49.34 shows that the most prolific base for the derivation of complex words is the verb. The average of verb-based derivatives clearly outnumbers the figures for adjectives and nouns. Furthermore, the 2nd order of verb-based derivations features the highest average number of derivatives in general. The values for adjective-based derivations are slightly higher than those for noun-based derivations. This result tallies with the data on the maximum derivational networks for individual word-classes, the data on the average number of orders of derivation, as well as the data on the correlation between

Table 49.34: Average number of derivatives by order of derivation for all languages.

	1st order	2nd order	3rd order	4th order	5th order
Nouns	7.71	6.26	2.21	0.43	0.06
Verbs	10.82	12.84	6.13	1.46	0.18
Adjectives	8.17	6.86	3.56	0.86	0.09

SVs and paradigmatic capacity in terms of the number of languages with a correlation above 0.5.

Language-wise, the highest average number of derivatives occurs in the 2nd order of Serbian verbs (39.10), followed by Croatian, Slovak, Georgian, Czech and Lithuanian, all of which exceed 30 derivatives. Interestingly, the average value for verbs is higher in the 2nd order than in the 1st order. Verb-based derivations are dominated by Slavic languages (Serbian, Croatian, Czech and Slovak). The maximum average values per language in adjective-based derivations are much lower than those in verbs and do not exceed 20 words, with the sole exception of the Serbian 2nd order (20.10). While the first three orders are clearly dominated by Slavic languages, Finnish gradually gains in importance as the order of derivation grows.

While the previously mentioned Slavic languages maintain high positions in noun-based derivations throughout all five orders, the roles of Italian in the first two orders and of Hungarian in the 3rd and the 4th orders are strikingly significant.

49.7.2 Sensitivity to absence of a word-formation process

Körtvélyessy et al. (2018) evaluated a sample of European languages by way of, *inter alia*, the calculation of the SV of individual word-formation processes.³ This enables us to test the interrelation between the absence of a given word-formation process or its low SV in word-formation and the number of derivatives. The question, therefore, is whether such an absence has an impact on the richness of a derivational network. For this purpose, we chose to analyze the

³ The saturation value in word-formation reflects the structural richness of a word-formation process. It is calculated as the proportion between the number of word-formation types actually employed for the formation of new complex words and the number of possible word-formation types within a given word-formation process.

class of verbs, owing to the phenomenon of *aktionsart* that is reflected in a large number of derivatives coined by prefixes.

In the following analysis, the SVs of prefixation identified by Körtvélyessy et al. (2018) are compared with the number of derivatives in the word-class of verbs. Two groups of languages are contrasted: languages with a high prefixation SV (high PrefSV), namely Romanian, Catalan, Dutch, German, and Maltese, and languages with a low or zero prefixation SV (low PrefSV), including Estonian, Hungarian, Tatar, and Finnish. Languages with a high PrefSV are, with the exception of Maltese, Indo-European languages from two language genera (Romance and German). Languages with a low PrefSV are Uralic and Altaic. Both Tatar and Turkish are members of the Turkic genus. Hungarian is an Ugric language; Estonian and Finnish belong to the Finnic genus. Thus, if Maltese is excluded, the Romance and Germanic languages are at one end of the scale (high PrefSV), and Turkic, Ugric and Finnic at the other (low PrefSV). Interestingly, languages of both groups are homogeneous also from the perspective of the morphological classification of languages. Romanian, Catalan, Dutch and German from the high PrefSV group are synthetic, while all languages from the low PrefSV are agglutinating languages.

Table 49.35 below orders languages according to their prefixation SVs. The number of derivatives is given as an average value.

Table 49.35: Languages by SV and number of derivatives.

Prefixation SV	Language	Number of derivatives				
		1st order	2nd order	3rd order	4th order	5th order
47.06	Romanian	12.3	11.2	2.9	0.2	0
41.18	Catalan	11.5	8.6	0.5	0	0
41.18	Dutch	5.4	4.8	0.8	0.2	0.1
41.18	German	10.4	12.4	6.4	1.2	0
41.18	Maltese	3.5	1.3	0.1	0	0
11.76	Estonian	11.6	11.9	4.7	1.2	0.2
11.76	Finnish	11.3	10.5	3.7	1.2	0.1
0	Hungarian	11.9	16.1	5.8	1.1	0.2
0	Tatar	8.9	6.7	1.5	0	0
0	Turkish	9.9	14.5	7.7	1.6	0.2

Vertical evaluation

In the 1st order of derivation, the number of derivatives is not proportional to the corresponding prefixation SV. While Romanian and Catalan are from the high PrefSV group and Estonian, Finnish and Hungarian belong to the low PrefSV group, their average numbers of derivatives are roughly identical.

The same observation applies to the 2nd order of derivation: all Uralic languages have a high number of derivatives and so do Romanian and Catalan. Besides genetic classification, the morphological type of a language plays an important role. Finnish, Hungarian and Estonian are agglutinating languages, while Romanian and Catalan are synthetic. From the 3rd order of derivation downwards, all languages with a high PrefSV show a decline and the number of derivatives is very low: in the 5th order there is only one derivative in Dutch, and there are no derivatives in Romanian, Catalan or German. In the low PrefSV languages, a different pattern of behaviour (with the exception of Tatar) can be observed: the number of derivatives drops with the higher degree of derivation in these languages as well but, if the position of a language (from the perspective of the number of derivatives) is taken into consideration, Turkish and the Uralic languages occupy the top positions in both the 4th and 5th orders.

What factors, then, affect the number of derivatives in individual languages? No doubt, there is not just a single reason. Instead, a combination of factors appears to be at play.

It is primarily, but not exclusively, a language's genetic membership that influences the number of derivatives rather than the absence or presence of prefixation in its word-formation system. As is evident from Table 49.35, German and Dutch have identical prefixation SVs, but significantly differ in their average number of derivatives in the 1st order. Similar cases abound. The *morphological type of a language* is also influential, but we must take into account that there is no systematic relation between the morphological type of a language and the nature of its word-formation system (Štekauer 2012). It is also necessary to reiterate the fact that the *word-formation systems of individual languages of the same genus diversified in the course of their development from their protolanguage*. Last but not least, we must mention the *compensation strategies* discussed in detail in individual language-specific chapters: each language finds its own word-formation strategy to compensate for its low/limited affixation capacity.

Horizontal evaluation

Generally, the number of derivatives drops as the degree of derivation increases. There are exceptions to this rule, however: German (high PrefSV) has a higher number of derivatives in the 2nd order, and so do Estonian, Hungarian and Turkish (low PrefSV). Thus, it can be concluded that this dropping line is not typical of languages with low prefixation SVs. With the exception of Tatar, languages with a low PrefSV have derivatives in each order of derivation. It is not so in the case of languages in the high PrefSV group: Catalan and Maltese do not derive new words in the 4th order, and in the 5th order, only one language from this group derives new words: Dutch (in fact, there is only one derivative for the verb *weten* ‘to know’).

The highest number of derivatives has been found for Hungarian in the 2nd order of derivation. In the 1st order it is Romanian (12.3), closely followed by Hungarian, Estonian and Finnish. From the 3rd order of derivation it is always a language from the low PrefSV group that assumes the top position. Thus, languages with low or no prefixation derive more verbs in individual orders of derivation.

Discussion

Disregarding Tatar, it can be concluded that languages that lack prefixation (or wherein the saturation value of prefixation is very low) derive new words in all orders of derivation. Languages with high prefixation SVs are typical of balanced numbers of derivatives in the 1st and 2nd orders, and the number of their derivatives dramatically drops from the 3rd order. A higher number of derived words in individual orders of derivation is typical of languages that lack prefixation or have low prefixation SVs. These observations can also be associated with the morphological type of languages and their genetic affiliation. All five languages with no prefixation (or with very low prefixation SVs) are agglutinating and fall into two language families. With the exception of Maltese, all languages from the high PrefSV group belong to the Indo-European language family representing two genera (Romance and Germanic), and they are synthetic languages.

Summary

- (i) Verb-based derivational networks feature the highest number of derivatives.
- (ii) The highest numbers of derivatives are produced by Slavic languages.

- (iii) The absence of a particular derivational process does not seem to affect the number of derivatives because it is usually compensated for by other word-formation processes.
- (iv) The number of derivatives appears to be the result of the interplay of several factors: the genetic affiliation of the language, which is reflected in its morphological type (synthetic type), which is itself reflected in high MDN values and high orders of derivation.

49.8 Correlations between semantic categories and orders of derivation

Even though they are not as rigorous and straightforward as form-based conclusions, some observations are worth mentioning that relate to emergent patterns based on the comparative semantic categories employed for describing the networks in the sample of 40 European languages.

Two methodological clarifications are in order here. A correlation between an order of derivation and semantic categories here means a consistent representation of a specific semantic category by derivations from the majority of base words within a word-class group. Second, a majority here means an incidence between 10 (all base words) and 7 or 6 base words.

DIMINUTIVE, QUALITY, PRIVATIVE, RELATIONAL and ACTION are the semantic categories that are most clearly correlated with the 1st order of derivation from nominal bases across the language sample. Thus, for example, DIMINUTIVE correlates with the 1st order of derivation of Spanish nouns (e.g. *hueso* ‘bone’ vs. *huesito* ‘small bone’) and adjectives (e.g. *delgado* ‘thin’ vs. *delgadito* ‘slightly thin’). DIMINUTIVE is actually the semantic category in which most correlations can be found in the sample: DIMINUTIVE is correlated with the 1st order of derivation in nine Slavic languages, six Romance languages, three Germanic languages, two Baltic languages, two Uralic languages and in Greek, Tatar and North Saami, i.e. in 25 languages from the sample altogether.

Even though a clear correlation for DIMINUTIVE can be identified in the majority of languages, AUGMENTATIVE is neither so copiously represented, nor so clearly correlated with the 1st order. Even though the two semantic categories are supposedly members of a common supercategory – EVALUATIVE – and denote opposite markedness or direction from a perceived standard, only DIMINUTIVE can be said to be strongly represented across language types, families, genera, and areal distribution. This is understandable, as witnessed by the findings presented by Körtvélyessy (2015: 119): out of the 71 European languages in her sample, 60 have

an evaluative morphology; 39 of these languages have both DIMINUTIVES and AUGMENTATIVES, while 21 languages only have DIMINUTIVES. No language has AUGMENTATIVES without also having DIMINUTIVES. Her results thus give additional support to the statement by Bakema and Geeraerts (2000: 106), who maintain that “[A]UGMENTATIVES are less widespread than DIMINUTIVES. The two categories are related by an implicational universal: the existence of AUGMENTATIVES in a language implies the presence of DIMINUTIVES, but the reverse does not hold.”

The semantic categories that are better represented in 1st order derivations from nominal bases in the sample are QUALITY, PRIVATIVE and RELATIONAL. QUALITY is correlated with 1st order derivatives from nominal bases in 21 languages, while PRIVATIVE is in 14 and RELATIONAL in 11, as in the following examples:

(3) QUALITY

English: *fire fiery*

PRIVATIVE

Slovak: *bezmenný bezzubý*
‘nameless’ ‘toothless’

RELATIONAL

Bulgarian: *kuče kučeški*
‘dog’ ‘related to a dog/of a dog’

Besides these clear associations, no further discernible patterns can be established.

Table 49.36: Correlation between the 1st order of derivation and semantic categories.

DIMINUTIVE	QUALITY	PRIVATIVE	RELATIONAL
Basque			
Bulgarian	Bulgarian		Bulgarian
Catalan	Catalan		Catalan
Croatian			Croatian
Czech			Czech
Chechen			
Dutch			
	English	English	

Table 49.36: (continued)

DIMINUTIVE	QUALITY	PRIVATIVE	RELATIONAL
Estonian		Estonian	
	Finnish	Finnish	
Frisian			
Galician		Galician	
German	German	German	
	Georgian	Georgian	
Greek			
Hungarian	Hungarian	Hungarian	
	Icelandic		
	Irish		
Italian	Italian	Italian	Italian
Latvian		Latvian	
Lithuanian		Lithuanian	Lithuanian
North Saami	North Saami	North Saami	
	Norwegian		
Polish	Polish		
Portuguese			
Romanian		Romanian	
Russian	Russian		
Serbian	Serbian		
	Swedish		
Slovak		Slovak	Slovak
Slovene			Slovene
Spanish			
Tatar		Tatar	
	Turkish	Turkish	
Ukrainian	Ukrainian		
	Welsh		Welsh

Examples of complex words derived from a nominal base which include these semantic categories in the 1st order are given in (4):

(4) DIMINUTIVE

Polish: *pies-ek*

dog-DIM

'doggie'

ACTION

Spanish: *ojo-ar*

eye-ACTION

'to eye'

QUALITY

Icelandic: *nafn-laus*

name-QUALITY

'nameless'

PRIVATIVE

Turkish: *kemik-siz*

bone-PRIVATIVE

'boneless'

As the order of derivation increases, the number of semantic categories for which a bias to correlation can be detected decreases. Only two semantic categories can be said to show a bias to correlation with the 2nd order of derivation from noun bases: ACTION and STATIVE. The incidence of these correlations was established in 15 languages for the former category and in nine for the latter. Table 49.37 represents the incidence of the most frequent semantic categories that correlate with the 2nd order of derivation from nominal bases.

Examples of complex words derived from a nominal base which include these semantic categories in the 2nd order are given in (5):

(5) ACTION

Croatian: *zub-ar-iti*

tooth-AGENT-ACTION

'to perform a dentist's work'

STATIVE

English: i) *tooth-less-ness*

ii) *tooth-i-ness*

Table 49.37: Correlation between the 2nd order of derivation and semantic categories in noun-based derivational networks.

ACTION	STATIVE
Catalan	
	English
Finnish	Finnish
Galician	
German	German
Greek	
Hungarian	Hungarian
	Icelandic
Irish	
Italian	
Lithuanian	
North Saami	
Norwegian	Norwegian
Portuguese	
	Russian
Swedish	Swedish
Turkish	Turkish
Ukrainian	

While Slavic languages stand out disproportionately in the sample in relation to the saturation of DIMINUTIVE in the 1st order of noun-based derivations, Germanic languages seem to be the ones wherein the majority of STATIVE derivations are consistently saturated in the 2nd order, while ACTION is saturated in the 2nd order in the greatest number of Romance languages.

As far as derivations from verb bases are concerned, the following tendencies are noticeable: there is a correlation between the 1st order of derivation and the semantic categories ACTION and AGENT (in 24 and in 21 languages, respectively). The remaining two semantic categories that are systematically saturated in the

1st order of derivation from verb bases, RESULTATIVE and ABILITY, are represented in 15 and 11 languages, respectively. Unlike in noun-based derivatives, DIMINUTIVE is correlated with the 1st order of derivation from verb bases in only seven languages, four of which are Slavic languages (Bulgarian, Russian, Slovak and Slovene), with the remainder being two Romance languages (Galician and Italian) and Greek. As with derivations from nominal bases, the pattern of an increase in the order of derivation being associated with a decrease in the number of semantic categories for which a bias to correlation can be detected is still present. Only two semantic categories can be said to show a tendency to correlate with the 2nd order of derivation from verb bases: ACTION and AGENT. ACTION seems to be frequently realized by 2nd order verb-based derivatives, being so in 14 languages; AGENT is realized in 10 languages. In the 2nd order, the tendencies are not as strong or clear as in the 1st order. No clear and consistent patterns can be found for the 3rd order of derivation in the set of derivational networks from verb bases. Table 49.38 summarizes the established patterns in the saturation of semantic categories per order of derivation in verb-based derivational networks.

The distribution of the first two categories in the 1st order of derivation is as follows: ACTION in 15 Indo-European languages (six Germanic, five Romance, two Baltic, one Celtic, and one Slavic), three Uralic languages, two Nakh-Daghestanian languages, one Altaic language, Basque, and one Afro-Asiatic language. For AGENT, the distribution is as follows: 15 Indo-European languages (six Romance, five Germanic, two Baltic, one Slavic, and one Celtic), two Nakh-Daghestanian languages, Basque, one Uralic language, and one Altaic language. It appears that Germanic languages tend towards ACTION, while Slavic languages seem to be significantly underrepresented. This can be accounted for by a methodological decision (cf. Chapter 2) that affected the data gathering stage, namely, to consider the derivation of ACTION (*nomina actionis*) to be considered so regular and predictable in Slavic languages that it was to be treated as falling under the scope of inflection. The absolute absence of any Slavic languages among the languages in which verb-based derivatives systematically saturate the semantic category ABILITY cannot be explained away by a methodological decision, however. This conspicuous absence is definitely associated with the features of this group of languages.

It is also worth noting that ACTION appears in both the 1st and 2nd order of derivation from verbal bases in the following languages: Latvian, Lithuanian, Estonian, Finnish, Hungarian, German, Icelandic, Norwegian, Catalan, Russian, and Turkish. This presupposes the occurrence of this semantic category in subsequent orders of derivation (see section 49.9). As for the semantic category AGENT, it appears repeatedly in subsequent orders of derivation from verb bases in the following languages: Latvian, Lithuanian, Georgian, Finnish, Swedish, Italian, and Romanian.

Table 49.38: Semantic categories per order of derivation in verb-based derivational networks.

Language	1st order of derivation			2nd order of derivation		
	ABILITY	ACTION	AGENT	DIMINUTIVE	RESULTATIVE	ACTION
Catalan	ACTION	AGENT			RESULTATIVE	
Chechen	ACTION					
Croatian				DIRECTIONAL	TERMINATIVE	ABSTRACTION
Czech				FINITIVE	ITERATIVE	ABILITY
Dargwa			AGENT		RESULTATIVE	
Dutch			AGENT			
English	ABILITY	ACTION	AGENT			
Estonian	ABILITY	ACTION			RESULTATIVE	
Finnish	ACTION	AGENT				ACTION
French	ABILITY	AGENT				Agent
Frisian	ACTION	AGENT			RESULTATIVE	
Galician	ABILITY	ACTION	AGENT	DIMINUTIVE		
Georgian	ACTION	AGENT				AGENT
German	ABILITY	ACTION	AGENT			ACTION

(continued)

Table 49.38 (continued)

Language	1st order of derivation			2nd order of derivation				
	DIMINUTIVE							
Greek								
Hungarian	ACTION	DIMINUTIVE			RESULTATIVE			
Icelandic	ACTION				ACTION			
Irish	ACTION	AGENT				ACTION AGENT		
Italian	ABILITY	ACTION	AGENT	DIMINUTIVE	RESULTATIVE			
Latvian	ACTION	AGENT				ACTION AGENT		
Lithuanian	ACTION	AGENT				ACTION AGENT		
Maltese	ACTION				RESULTATIVE			
North saami	ACTION	AGENT				ACTION		
Norwegian	ACTION				RESULTATIVE			
Polish	DIMINUTIVE			RESULTATIVE				
Portuguese	ACTION	AGENT				ACTION		
Romanian	ABILITY	AGENT				AGENT		
Russian	ACTION				ACTION			
Serbian	DIMINUTIVE			RESULTATIVE				
Slovak				RESULTATIVE				
				ACTION				

Slovene		AGENT		
Spanish	ABILITY	ACTION		RESULTATIVE
Swedish	ABILITY	ACTION	AGENT	
Ukrainian		DIMINUTIVE		RESULTATIVE
Tatar		AGENT		
Turkish		ACTION		RESULTATIVE
Welsh	ABILITY			ACTION

In adjective-based derivations, the most conspicuous tendency of correlation in the 1st order is detected for the categories STATIVE, MANNER, DIMINUTIVE and AUGMENTATIVE, which are represented in 23, 21, 17 and 15 languages, respectively. Table 49.39 lists the languages in which the correlation has been established.

Table 49.39: Correlation between semantic categories and the 1st order of derivation from adjective bases.

MANNER	STATIVE	DIMINUTIVE	AUGMENTATIVE
Basque	Basque	Basque	
Bulgarian	Bulgarian	Bulgarian	
		Catalan	Catalan
Czech			
	Chechen		
Dargwa	Dargwa		
	Dutch		
English	English		
Estonian			
Finnish	Finnish	Finnish	
French	French		
Galician	Galician	Galician	Galician
Georgian	Georgian	Georgian	
	German		
Greek	Greek		
Hungarian			
Icelandic	Icelandic		
	Irish		
Italian	Italian		
Latvian			
Lithuanian	Lithuanian		
	Maltese		

Table 49.39 (continued)

MANNER	STATIVE	DIMINUTIVE	AUGMENTATIVE
North Saami			
Norwegian			
Polish		Polish	
	Portuguese	Portuguese	Portuguese
	Romanian	Romanian	Romanian
Russian			
Slovak	Slovak	Slovak	
Slovene			Slovene
Spanish	Spanish	Spanish	Spanish
Tatar			
	Turkish	Turkish	Turkish
Swedish			Swedish
Ukrainian	Ukrainian	Ukrainian	Ukrainian
Welsh			

The distribution is as follows: STATIVE in 14 Indo-European languages (six Germanic, four Romance, three Slavic, and one Celtic), two Nakh-Daghestanian languages, one Kartvelian language, Basque, and three Uralic languages. MANNER appears in 18 Indo-European languages (six Slavic, four Romance, three Uralic, two Baltic, two Germanic, and Greek), two Nakh-Daghestanian languages, and Basque. Unlike in noun-based derivatives, in adjective-based ones, the occurrence of both AUGMENTATIVE and DIMINUTIVE is almost balanced across languages – AUGMENTATIVE systematically appears in 1st order derivatives from adjective bases in 15 languages and DIMINUTIVE in 16. Worth mentioning is also the fact that the semantic category ACTION makes consistent appearances in both the 2nd and 3rd order of derivation, in 17 and 12 languages respectively. No other remarkable patterns can be isolated for the 2nd and 3rd orders.

Summary

- (i) In general, correlations appear to depend on the base.
- (ii) Correlations between semantic categories and orders of derivation occur mainly in the first two orders of derivation: they occur steadily in virtually all languages in the 1st order, and in most languages in the 2nd order.
- (iii) By contrast, beyond the 2nd order of derivation, correlations become obscure: in several languages a correlation between the 3rd order and specific semantic categories has been established, but the semantic categories are diverse and only six languages coincide here as regards which semantic categories correlate. From that order onwards, no clear correlations were observed.
- (iv) Regarding semantic categories, DIMINUTIVE appears to be correlated with the 1st order of derivation from all three types of base.

49.9 Semantic categories with blocking effects

There has been some fairly extensive research into the combinability of affixes and the blocking effects of affixes (so-called terminal affixes). The underlying idea of this direction of investigation concerns the fact that affixes cannot be combined arbitrarily. Instead, there are strict rules in each language that govern the possibility of the attachment of various affixes to the base (such as various base-driven and affix-driven approaches; cf. Fabb 1988; Plag 1996; Giegerich 1999; Gaeta 2005, among others) as well as the ordering and combinations of affixes (cf., for example, Aronoff and Fuhrhop 2002; Hay and Plag 2004; Manova and Aronoff 2010; Saarinen and Hay 2014; Manova 2015). While the approaches to affix ordering vary (Manova and Aronoff (2010) identify eight approaches: phonological, morphological, syntactic, semantic, statistical, psycholinguistic, cognitive and templative), the point of departure in each of them is affix. Our approach is different and, in a way, complementary. We take a semantic category (which, usually, can be represented by more than one synonymous affix) as the starting point and examine their combinability as well as their capacity to block any subsequent derivation. Certainly, in a form of typological research, we necessarily have to try to identify any regularities and tendencies across languages. Our findings are presented in sections 49.9 and 49.10.

Blocking is reported to be conditioned by specific suffixes and also by specific semantic categories. The former is reported for languages like Estonian and Norwegian in the language sample, but the focus of this section is on the

latter, i.e. blocking by semantic categories, such as blocking by the semantic category MANNER in Portuguese (e.g. *novamente* ‘newly’) or by the semantic category STATIVE in German (e.g. *Schneidbarkeit* ‘cuttability’).

Semantic categories with blocking effects are reported in all the languages of the sample except Welsh. A bias towards specific blocking categories can also be identified, such that the same semantic category has a blocking effect through all the word-classes, regardless of the order of derivation. Thus, for example, in North Saami the category ACTION blocks further derivation in verbs from the 1st order onwards, and in nouns and adjectives from the 2nd order onwards. This can be attested for specific semantic categories across languages as follows.

- (i) Finnish: STATIVE
- (ii) Georgian: AGENT
- (iii) Hungarian: MANNER PROCESS
- (iv) Icelandic: ACTION PROCESS
 - AGENT QUALITY
 - MANNER STATIVE
- (v) Lithuanian: ACTION
- (vi) North Saami: ACTION PROCESS
 - AGENT
- (vii) Slovene: DIMINUTIVE FEMALE
- (viii) Spanish: MANNER
- (ix) Swedish: AGENT INSTRUMENT
 - AUGMENTATIVE MANNER
 - ENTITY PATIENT
- (x) Ukrainian: AGENT

Again, considerable variation can be identified here, ranging from languages where blocking is reported for only one category (Finnish, Georgian, Lithuanian, Spanish, Ukrainian) to languages where up to six categories have a blocking effect (Icelandic, Swedish). Remarkably, the number of semantic categories with blocking effects through all word-classes is comparatively low: 12 categories out of a possible 41. It is also remarkable how, of these blocking categories, the category AGENT occurs in half of the languages reported above.

By word-class, in adjective-based derivations, the most conspicuous associations in the 1st order are detected for STATE and DIMINUTIVE, as follows:

- (i) DIMINUTIVE: Czech Greek Slovene
 - Georgian Maltese Turkish

(ii) STATE:	Bulgarian	Georgian	Maltese
	Chechen	Hungarian	Tatar
	Finnish	Icelandic	Ukrainian

In the 2nd order of derivation, they are for AGENT, MANNER and PROCESS, as follows:

(i) AGENT:	Georgian	Lithuanian	Swedish
	Icelandic	North Saami	Ukrainian
(ii) MANNER:	Finnish	Lithuanian	Spanish
	Hungarian	Polish	
(iii) PROCESS:	Hungarian	Maltese	
	Icelandic	North Saami	

In the 3rd order of derivation, they are for ACTION and MANNER, as follows:

(i) ACTION:	Hungarian	Lithuanian	Turkish
	Icelandic	North Saami	Ukrainian
(ii) MANNER:	Czech	Greek	Spanish
	Finnish	Hungarian	

In the 4th order of derivation, it is for ACTION, as follows:

(i) ACTION:	Lithuanian	North Saami
-------------	------------	-------------

These data are summarized in [Table 49.40](#).

According to this, Hungarian, Icelandic, Lithuanian and North Saami record consistent blocking by semantic categories across orders of derivation within the adjective word-class.

In the orders of derivation where several semantic categories are reported to have a blocking effect, the tendency is for each language to have only one semantic category with a blocking effect, e.g. Bulgarian with respect to STATE in the 1st order. However, some languages report several, e.g. Georgian reports three (DIMINUTIVE, STATE, AGENT) and Maltese reports two (DIMINUTIVE, STATE), both in the 1st order. Similarly, Hungarian, Lithuanian and North Saami report two categories each in the 2nd order, and Hungarian does so again in the 3rd order. Few languages report blocking in the 4th order of derivation. Otherwise,

Table 49.40: Semantic categories with blocking effects in adjectives.

Language	Order of derivation						
	1st order		2nd order		3rd order		
Bulgarian	STATE						
Chechen	STATE						
Czech	DIMINUTIVE				MANNER		
Finnish	STATE		MANNER		MANNER		
Georgian	DIMINUTIVE	STATE	AGENT				
Greek	DIMINUTIVE				MANNER		
Hungarian	STATE		MANNER	PROCESS	ACTION		
Icelandic	STATE		AGENT	PROCESS	ACTION		
Lithuanian	AGENT		MANNER	ACTION			
Maltese	DIMINUTIVE	STATE	PROCESS				
North Saami	AGENT		PROCESS	ACTION	ACTION		
Polish	MANNER						
Slovene	DIMINUTIVE						
Spanish	MANNER				MANNER		
Swedish	AGENT						
Tatar	STATE						
Turkish	DIMINUTIVE				ACTION		
Ukrainian	STATE		AGENT	ACTION			

no pattern of distribution can be identified here in relation to language type, language family, or genus or areal specification.

In noun-based derivations, the most conspicuous associations in the 1st order are detected for DIMINUTIVE, MANNER and SIMILATIVE, as follows:

- (i) DIMINUTIVE: Georgian Polish Turkish
Hungarian North Saami Ukrainian
Latvian Slovene
Lithuanian Tatar

(ii) MANNER:	Chechen	North Saami
	Icelandic	Swedish

(iii) SIMILATIVE:	Georgian	Tatar
	Slovene	Turkish

In the 2nd order of derivation, they are for ACTION and MANNER, as follows:

(i) ACTION:	Hungarian	North Saami
	Latvian	Turkish

(ii) MANNER:	Czech	Hungarian	Spanish
	Finnish	Latvian	Swedish

In the 3rd order of derivation, it is for ACTION, as follows:

(i) ACTION:	Hungarian	North Saami
	Lithuanian	Turkish

In the 4th order of derivation, it is again for ACTION, as follows:

(i) ACTION:	Hungarian	Turkish
	North Saami	

These data are summarized in [Table 49.41](#).

According to this, Hungarian, North Saami and Turkish record consistent blocking by semantic categories across orders of derivation within the noun word-class.

As with adjectives, in the orders of derivation where several semantic categories have a blocking effect, many languages report only having one semantic category with a blocking effect, e.g. Chechen with respect to MANNER in the 1st order. However, some languages report several, e.g. Georgian, Slovene, Tatar and Turkish report two blocking categories in the 1st order, and Hungarian and Latvian also report two blocking categories in the 2nd order. Few languages report blocking in the 3rd or 4th order of derivation. Otherwise, no pattern of distribution can be identified here in relation to language type, language family, or genus or areal specification.

In verb-based derivations, the most conspicuous associations in the 1st order are detected for ACTION, AGENT ENTITY and LOCATION, as follows:

Table 49.41: Semantic categories with blocking effects in nouns.

Language	Order of derivation						
	1st order		2nd order		3rd order		
Chechen	MANNER						
Czech	MANNER						
Finnish	MANNER						
Georgian	DIMINUTIVE	SIMILATIVE					
Greek							
Hungarian	DIMINUTIVE		ACTION	MANNER	ACTION		
Icelandic		MANNER					
Latvian	DIMINUTIVE		ACTION	MANNER			
Lithuanian	DIMINUTIVE			ACTION			
North Saami	DIMINUTIVE	MANNER	ACTION	ACTION	ACTION		
Polish	DIMINUTIVE						
Slovene	DIMINUTIVE		SIMILATIVE				
Spanish				MANNER			
Swedish		MANNER		MANNER			
Tatar	DIMINUTIVE		SIMILATIVE				
Turkish	DIMINUTIVE		SIMILATIVE	ACTION	ACTION		
Ukrainian	DIMINUTIVE						

- (i) ACTION: Basque Lithuanian Slovene
 Latvian North Saami Tatar
- (ii) AGENT: Greek Icelandic Swedish
 Hungarian North Saami Tatar
- (iii) ENTITY: Greek Swedish
 North Saami Ukrainian
- (iv) LOCATION: Bulgarian Greek Ukrainian
 Georgian Swedish

In the 2nd order of derivation, they are for ACTION and DIMINUTIVE, as follows:

(i) ACTION:	Hungarian Icelandic	North Saami Tatar
-------------	------------------------	----------------------

(ii) DIMINUTIVE:	Czech Greek	Slovene Tatar
------------------	----------------	------------------

In the 3rd order of derivation, they are for ACTION and MANNER, as follows:

(i) ACTION:	Hungarian Latvian	North Saami Turkish
-------------	----------------------	------------------------

(ii) MANNER:	Czech Finnish	Hungarian Latvian	Spanish Swedish
--------------	------------------	----------------------	--------------------

In the 4th order of derivation, it is for ACTION, as follows:

(i) ACTION:	Hungarian	North Saami
-------------	-----------	-------------

These data are summarized in Table 49.42.

Table 49.42: Semantic categories with blocking effects in verbs.

Language	Order of derivation					
	1st order		2nd order		3rd order	4th order
Basque	ACTION					
Bulgarian			LOCATION			
Czech				DIMINUTIVE	MANNER	
Finnish						
Georgian			LOCATION			
Greek	AGENT	ENTITY	LOCATION	DIMINUTIVE		
Hungarian	AGENT			ACTION	MANNER	ACTION
Icelandic	AGENT			ACTION		
Latvian	ACTION			ACTION		

Table 49.42 (continued)

Language	Order of derivation				
	1st order		2nd order		3rd order
Lithuanian	ACTION				
North Saami	ACTION	AGENT	ENTITY	ACTION	ACTION
Slovene	ACTION			DIMINUTIVE	
Spanish					MANNER
Swedish		AGENT	ENTITY	LOCATION	
Tatar	ACTION	AGENT		ACTION	DIMINUTIVE
Turkish					ACTION
Ukrainian		ENTITY	LOCATION		

According to Table 49.42, Greek, Hungarian, North Saami, Swedish and Tatar record consistent blocking by semantic categories across orders of derivation within the verb word-class.

As for adjectives and nouns, in the orders of derivation where several semantic categories have a blocking effect, the tendency is for languages to have only one semantic category with a blocking effect, even if some languages report several, e.g. Greek, North Saami, Swedish, Tatar and Ukrainian in the 1st order, Tatar in the 2nd order, and Hungarian in the 3rd order. Few languages report blocking in the 4th order of derivation. Otherwise, no pattern of distribution can be identified here in relation to language type, language family, or genus or areal specification.

The opposite scenario, i.e. the identification of semantic categories that do not occur at all in specific languages, is also possible, but the combinations are too many to list here. A set of semantic categories that rarely have a blocking effect in the sample can, however, be listed here. The following selections are reported only in one case, regardless of the word-class or the order of derivation:

- (i) ATTENUATIVE, in the 2nd order of Greek nouns
- (ii) ANTICAUSATIVE, in the 1st order of Maltese verbs
- (iii) CONCOMITANT, in the 1st order of Greek verbs
- (iv) CUMULATIVE, in the 2nd order of Ukrainian verbs
- (v) DISTRIBUTIVE, in the 3rd order of Polish verbs
- (vi) INCEPTIVE, in the 1st order of Ukrainian verbs
- (vii) POSSESSIVE, in the 1st order of Hungarian nouns

Summary

- (i) All languages in the sample, irrespective of genus or family, report semantic categories with blocking effects. The only exception is Welsh.
- (ii) Semantic categories do not report blocking effects through all the word-classes in the same order of derivation. DIMINUTIVE, MANNER and ACTION cut across word-classes, albeit in different orders of derivation for the three word-classes.
- (iii) Blocking in the 4th order of derivation is reported to be base-insensitive and is limited to both one semantic category (ACTION) and few languages (Hungarian and North Saami for all three word-classes, and Lithuanian and Turkish for nouns).
- (iv) Several languages record consistent blocking by semantic categories across orders of derivation, e.g. Hungarian and North Saami.
- (v) In the orders of derivation where several semantic categories have a blocking effect, the tendency is for languages to have only one semantic category with a blocking effect, except for some languages which consistently report several, regardless of the word-class or order of derivation, e.g. Hungarian and Tatar.

49.10 Typical combinations of semantic categories

By ‘typical combinations of semantic categories’, we mean the occurrence of semantic categories in successive orders of derivation in the networks of five or more of the base words in a particular word-class, e.g. in Catalan (nouns as base): LOCATION-base > ACTION > RESULTATIVE (e.g. *magatzem* ‘warehouse’ > *emmagatzemar* > *desemmagatzemar*); in Croatian (nouns as base): AGENT (or EXPERIENCER, PATIENT) – POSSESSIVE (e.g. *pas* ‘dog’ > *pset-ar* ‘dog holder’ – *pset-ar-ev* ‘dog holder’s’), AGENT – FEMININE – POSSESSIVE (*zub* > *zub-ar* – *zub-ar-ica* – *zub-ar-ič-in* ‘female dentist’s’); and in English (verbs as base): combinations of ABILITY in the 1st order with ABILITY and PRIVATIVE in the 2nd order (*cuttable* > *cuttability*, *cuttable* > *uncuttable*). Typical combinations of semantic categories are reported for all the languages in the sample except Welsh. A number of languages also do not attest typical combinations for specific word-classes, e.g. there are no typical combinations for adjectives in Dargwa, for nouns in Chechen, Dargwa, Maltese and Polish, or for verbs in Irish.

The combinability of semantic categories (which, for methodological reasons, can only be traced in combinations between 1st and 2nd order derivatives onwards) is astoundingly varied, and no recurrent patterns could be established across the language sample for typical combinations across the board. Still, some cases can be highlighted.

- (i) Each word-class starts out with specific typical combinations, such that the same sequence is typically not found across the three word-classes in the sample, except in certain languages, as in example (5) and as summarized in Table 49.43.

Table 49.43: Languages in which typical combinations occur regardless of word-class.

Language	Typical combination		
	Adjectives	Nouns	Verbs
Czech	QUALITY + ABSTRACTION	QUALITY + ABSTRACTION	QUALITY + ABSTRACTION
	QUALITY + PATIENT	QUALITY + PATIENT	QUALITY + PATIENT
	RELATIONAL + MANNER	RELATIONAL + MANNER	RELATIONAL + MANNER
Georgian		ACTION + AGENT	ACTION + AGENT
		ACTION + CAUSATIVE	ACTION + CAUSATIVE
		ACTION + RESULT	ACTION + RESULT
Greek	ACTION + SATURATIVE + PATIENT	ACTION + SATURATIVE + PATIENT	
	ACTION + ITERATIVE + PATIENT	ACTION + ITERATIVE + PATIENT	
	ACTION + ATTENUATIVE + PATIENT	ACTION + ATTENUATIVE + PATIENT	
Icelandic	QUALITY + QUALITY	QUALITY + QUALITY	QUALITY + QUALITY
Irish	CAUSATIVE + ACTION	CAUSATIVE + ACTION	
Latvian		CAUSATIVE + FINITIVE + ACTION	CAUSATIVE + FINITIVE ACTION
North Saami	RESULTATIVE + PROCESS	RESULTATIVE + PROCESS	
	RESULTATIVE + INCHOATIVE + PROCESS	RESULTATIVE + INCHOATIVE + PROCESS	

(5) Icelandic

Noun:	<i>nafnlaus</i>	>	<i>nafnleysi</i>
	‘nameless’		‘anonymity’
Verb:	<i>brennanlegur</i>	>	<i>torbrennanlegur</i>
	‘burnable’		‘difficult to burn’
Adjective:	<i>langsamur</i>	>	<i>langsamlega</i>
	‘prolonged’		‘by far’

No associations can be identified here with regard to language type, language family, or genus or areal distribution.

- (ii) The opposite scenario, i.e. typical combinations which are specific to word-classes, can also be found in several languages, but with considerably more restrictions, as in example (6) and as summarized in [Table 49.44](#).

Table 49.44: Word-class-specific typical combinations across languages.

Language	Typical combination		
	Adjectives	Nouns	Verbs
Chechen	QUALITY + STATE		
English	QUALITY + STATE		
Ukrainian	QUALITY + STATE		

(6) Chechen:

Adjective	<i>vokkha</i>	>	<i>vokkhalla</i>
‘old, senior, elder’			‘an old age, seniority, eldership’

It follows from [Table 49.44](#) that it is only the class of adjectives that manifests typical combinations of semantic categories across languages. No associations can be identified here with regard to language type, language family, or genus or areal distribution.

- (iii) The various possible arrangements and the size of the sample do not allow us even to hint at combinations that otherwise typically occur. It is, however, possible to identify typical starting categories that are particularly frequent in each word-class ([Table 49.45](#)) and typical starting categories that

occur in each word-class even if they are not particularly frequent in a language (Table 49.46).

Table 49.45 above shows specific biases of certain languages towards specific categories in specific word-classes, e.g. Bulgarian adjectives towards PATIENT, in contrast to North Saami adjectives towards RESULTATIVE or Tatar adjectives towards UNDERGOER, to name just one of the possible contrasts that can be found. More relevantly, it also shows, as also confirmed in Table 49.46, that certain categories stand out for their frequency regardless of their word-class or language, such as:

- a. ABILITY, AGENT, ACTION, CAUSATIVE, QUALITY or SIMILATIVE in the adjective word-class. Examples of combinations that occur in several languages are:
 - QUALITY + AUGMENTATIVE in German, Swedish and Ukrainian
 - PRIVATIVE + MANNER in Italian, Polish and Russian
 - SIMILATIVE + MANNER in Czech, Latvian, Lithuanian and Polish
- b. AGENT, QUALITY, PRIVATIVE and SIMILATIVE in the noun word-class. Examples of combinations that occur in several languages are:
 - QUALITY + STATIVE in Bulgarian, English, Norwegian and Serbian
 - CAUSATIVE + ACTION in Basque, Irish, Norwegian and Spanish
- c. AGENT, CAUSATIVE, DIRECTIONAL, FINITIVE, PROCESS and QUALITY in the verb word-class. Examples of combinations that occur in several languages are:
 - AGENT + FEMALE in Bulgarian, Dutch, German, Serbian and Slovene
 - CAUSATIVE + ACTION in Finnish and Swedish
 - CAUSATIVE + AGENT in Finnish and Swedish

No pattern can be identified by language type, language family, or genus or areal distribution for these typical combinations.

(iv) Despite the apparent bias towards a set of starting categories, the possible observations of combinations are thus very limited:

- a. Certain combinations with a range of subsequent categories can be attested at least four times across languages in the language sample for a specific word-class:
 - i. In the adjective word-class:
 - QUALITY + STATE (+) in Chechen, English, German and Ukrainian
 - ii. In the verb word-class:
 - AGENT + FEMALE (+) in Bulgarian, Croatian, German, Serbian and Slovak

Table 49.45: Initial categories in typical combinations across languages.

Language	Typical combination	
	Adjectives	Nouns
Bulgarian		PATIENT
Catalan		QUALITY
Croatian	AGENT	AGENT
Czech		
Finnish		
Galician		
Georgian	ACTION	ACTION
German	ABILITY	QUALITY
Greek	ACTION	ACTION
Hungarian		QUALITY
Icelandic		QUALITY
Lithuanian		CAUSATIVE PROCESS
North Saami		RESULTATIVE
Portuguese		
Romanian	CAUSATIVE	CAUSATIVE
Russian		
Tatar		UNDERGOER
Ukrainian		QUALITY

The presented semantic categories in Table 49.45 are characterized by a frequency above three attested combinations per language, regardless of word-class.

Typical combination			
Verbs			
ENTITY		ACTION	
		AGENT	
		DURATIVE	
		CAUSATIVE	PROCESS
		AGENT	
		ACTION	
QUALITY	ABILITY	AGENT	QUALITY
		QUALITY	
		QUALITY	
PROCESS		DIRECTIONAL	
		RESULTATIVE	CAUSATIVE
DIMINUTIVE		CAUSATIVE	
		ACTION	
		AGENT	UNDERGOER
		ACTION	

Table 49.46: Initial categories in typical combinations across languages.

Language				Semantic categories that	
	Adjectives			Nouns	
Bulgarian					
Catalan			QUALITY		
Chechen			QUALITY		
Croatian	AGENT				AGENT CAUSATIVE
Czech					AGENT
Dargwa					
English	ABILITY		QUALITY		PRIVATIVE
Finnish					
Galician					
Georgian	ACTION			ACTION	PRIVATIVE
German	ABILITY	AGENT	QUALITY	ABILITY	AGENT
Greek	ACTION				ACTION
Hungarian			QUALITY		
Icelandic			QUALITY		
Irish	CAUSATIVE		QUALITY		CAUSATIVE
Latvian	CAUSATIVE			SIMILATIVE	
Lithuanian	CAUSATIVE	PROCESS		SIMILATIVE	
Maltese	CAUSATIVE				
North Saami					PRIVATIVE
Polish		PRIVATIVE		SIMILATIVE	
Portuguese					
Romanian					
Russian					
Slovene	CAUSATIVE				
Swedish					
Tatar					
Ukrainian			QUALITY		

The presented initial semantic categories are with a frequency above four attested combinations, regardless of language.

start typical combinations

Verbs

QUALITY	AGENT	DIRECTIONAL	
ACTION			
ACTION			
	AGENT		
QUALITY SIMILATIVE		DIRECTIONAL FINITIVE	SIMILATIVE
	AGENT		
		PRIVATIVE	
	CAUSATIVE		PROCESS
QUALITY			PRIVATIVE
QUALITY SIMILATIVE ABILITY	AGENT		QUALITY
QUALITY			QUALITY
QUALITY			PROCESS
QUALITY	CAUSATIVE	DIRECTIONAL FINITIVE	
PROCESS		DIRECTIONAL FINITIVE	
	CAUSATIVE		
		PROCESS	
	CAUSATIVE		
			QUALITY
	AGENT	DIRECTIONAL FINITIVE	
	CAUSATIVE		
	AGENT CAUSATIVE		

- b. Certain combinations are attested only once in the language sample, even regardless of the word-class, which is sometimes because their starting category occurs only once, even in the same language:
 - i. In the noun word-class:
TEMPORAL + QUANTITY (+) in Catalan
 - ii. In the verb word-class:
ANTICAUSATIVE + ENTITY in Icelandic
CONCOMITANT + PATIENT in Greek
DISTRIBUTIVE + DURATIVE (+) in Czech

Summary

- (i) Typical combinations of semantic categories are reported for most of the languages in the sample, and also regularly for each word-class. However, each word-class starts out with its own specific typical combinations, such that the same sequence is typically not found across the three word-classes in the sample.
- (ii) Few typical combinations specific to word-classes can be found across languages, e.g. QUALITY + STATE in adjectives in Chechen, English and Ukrainian, so few biases of certain languages towards typical combinations of specific categories in specific word-classes can be detected.
- (iii) Some starting categories that occur in each word-class can be identified as showing categories that stand out for their frequency regardless of word-class and language, e.g. QUALITY.
- (iv) Unique combinations can be attested in the sample too, e.g. ANTICAUSATIVE + ENTITY occurs only once in the sample (Icelandic).
- (v) No pattern can be identified by language type, language genus, or family or areal distribution, but certain combinations with a range of subsequent categories can be attested across languages.

49.11 Multiple occurrence of semantic categories

The multiple occurrence of semantic categories in subsequent orders of derivation is a relatively frequent phenomenon in the 40 European languages under study. This phenomenon is manifested in the recurrence of the same semantic category in successive orders of derivation, as in the examples below:

- Basque: e.g. QUALITY – QUALITY: *epel*_A ‘warm’ > *epelki*_{ADV} ‘warmly’ > *epelkiro*_{ADV} ‘warmly’; *zuzen*_A ‘straight’ > *zuzenki*_{ADV} ‘straightly’ > *zuzenkiro*_{ADV} ‘straightly’; *berri*_A ‘new’ > *berriro*_{ADV} ‘newly, again’ > *berriroki*_{ADV} ‘newly, again’.
- Croatian: e.g. QUALITY – QUALITY: *zl-o-ba-n* ‘malicious’ – *na-zl-o-ba-n* ‘malicious’.
- Frisian: e.g. *witt* ‘know’ > ABSTRACTION in 1st order verb: *witten* > ABSTRACTION *wittenskip* ‘science’ > ABSTRACTION in 4th order: *wittenskiplikens* ‘scientific character’.
- Georgian: e.g. ACTION – ACTION: *tvaliereba*_V ‘to look at, examine’ > *atvaliereba*_V ‘to look up’; QUALITY – QUALITY: *moč'rili*_A ‘cut’ > *amoč'rili*_A ‘cut out’; AGENT – AGENT: *msroleli*_N ‘thrower’ > *amsroleli*_N ‘who throws up’; CAUSATIVE – CAUSATIVE: *srolineba*_V ‘let someone throw’ > *asrolineba*_V ‘let someone throw up’.

Multiple occurrence has been identified in 26 out of the 40 languages: Basque, Bulgarian, Catalan, Chechen, Dargwa, Finnish, Frisian, Georgian, German, Greek, Hungarian, Icelandic, Latvian, North Saami, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Swedish, Tatar, Turkish and Ukrainian.

These languages cover all language types and all language families, with the exception of the Afro-Asiatic family (Maltese), represented in the sample, as shown in Tables 49.47 and 49.48.

Table 49.47: Languages where the multiple occurrence of semantic categories is attested by language type.

Language type	Languages
Inflectional	Bulgarian, Catalan, German, Greek, Icelandic, Latvian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovene, Spanish, Ukrainian
Agglutinating	Basque, Chechen, Dargwa, Finnish, Georgian, Hungarian, North Saami, Tatar, Turkish
Analytical	Frisian, Swedish

Of these, Table 49.47 maintains in the main the proportions of the language types represented in the sample, whereas Table 49.48 lends itself to further comment in that the proportions of the language sample are not replicated in that Slavic languages, compared with other language genera of the Indo-European

Table 49.48: Languages where the multiple occurrence of semantic categories is attested by language family.

family, display multiple occurrence in seven out of nine languages in this genus, in contrast to four out of eight Germanic languages or three out of seven Romance languages.

With regard to the opposite case, i.e. the languages wherein multiple occurrence is not recorded, Tables 49.49 and 49.50 show that no relevant pattern can be identified, as the languages arrange themselves across language types and language families or genera in an approximate proportion to the number of languages of each case represented in the sample.

Table 49.49: Languages where the multiple occurrence of semantic categories is not attested by language type.

Language type	Languages
Inflectional	Dutch, French, Galician, Italian ^a
Agglutinating	Estonian
Analytical	Danish, English

^aHere recorded as negligible.

Table 49.50 is congruent with Table 49.46 in that it contains a fair number of the most represented language genera, i.e. Germanic and Romance, and in that it shows the opposite results to Table 49.46, i.e. a disproportional underrepresentation of Slavic languages.

The above does not take into consideration 10 languages, not listed above, for which occasional, rare, sporadic or multiple occurrences of semantic categories are reported. These languages are Croatian, Czech, Estonian, Irish, Lithuanian, Maltese, Norwegian, Romanian, Tatar and Welsh.

There is considerable variation as to which categories occur multiple times, in which orders of derivation they occur, and for which word-classes. Languages thus display arrangements that may differ considerably in several ways:

- (i) There may be considerable variation within languages with regard to semantic categories and how they occur multiple times with respect to word-classes, e.g.:
 - a. Greek records multiple occurrences of the category DIMINUTIVE in all three word-classes, whereas German displays multiple occurrences of AUGMENTATIVE in adjectives and verbs, but not in nouns. The latter, a rather specific pattern, may find similar arrangements in other languages, e.g. in Tatar, where the multiple occurrence of the semantic

Table 49.50: Languages where the multiple occurrence of semantic categories is not attested by language family.

Afro-Asiatic	Altaic	Basque	Language family/genus					Uralic	
			Baltic	Celtic	Germanic	Greek	Romance		
			Estonian		Danish Dutch English		French Galician Italian	Kartvelian	Nakh-Daghestanian

category CAUSATIVE is recorded in adjectives and verbs but, like German for AUGMENTATIVE, not in nouns.

- b. Bulgarian records multiple occurrences of the categories QUALITY in nouns and PLURIACTIONALITY in verbs, but none in adjectives, whereas Latvian, for example, records multiple occurrences of SIMILATIVE precisely in adjectives.
- (ii) There may be considerable variation between languages with regard to semantic categories and how they occur multiple times with respect to word-classes, e.g.:
 - a. Frisian reports multiple occurrences of a limited set of semantic categories per word-class:
 - i. Adjectives: ACTION
 - ii. Nouns: QUALITY
 - iii. Verbs: ABSTRACTION
 - b. This contrasts sharply with the set of semantic categories recorded for Polish or for Slovene, of which, for brevity, only the former is shown for illustration:
 - i. Adjectives: DIMINUTIVE, LOCATION, MANNER, QUALITY, REFLEXIVE and RESULTATIVE
 - ii. Nouns: COLLECTIVE, DIMINUTIVE, LOCATION, QUALITY, RELATIONAL and RESULTATIVE
 - iii. Verbs: ACTION, AGENT, AUGMENTATIVE, DIMINUTIVE, DISTRIBUTIVE, FEMALE, ITERATIVE, LOCATION, REFLEXIVE, RELATIONAL, and RESULTATIVE

The possibility of the multiple occurrence of a semantic category in successive orders of derivation can be explored in several senses. The first is whether the categories reoccur successively with or without a different intermediate semantic category. The two cases are recorded in the sample.

- (i) Multiple occurrence of a semantic category without a different intermediate category, e.g.:
 - a. Frisian:
 - i. Adjectives: ACTION + ACTION
 - ii. Nouns: QUALITY + QUALITY
 - iii. Verbs: ABSTRACTION + ABSTRACTION + ABSTRACTION
- (ii) Multiple occurrence of a semantic category with a different intermediate category, e.g.:

- a. Slovene:
 - i. Adjectives: PROCESS + PROCESS
MANNER + MANNER
 - ii. Nouns: ACTION + ACTION
DIMINUTIVE + DIMINUTIVE
PROCESS + FINITIVE + PROCESS
RELATIONAL + ENTITY + RELATIONAL
 - iii. Verbs: QUALITY + QUALITY
FINITIVE + QUALITY + PRIVATIVE + QUALITY

Concerning which semantic categories occur on multiple occasions, certain ones stand out in this respect because they appear consistently, as shown in Table 49.51.

Table 49.51: Most frequently recorded semantic categories with multiple occurrences.

Semantic category	Languages
QUALITY (10 languages)	Bulgarian, Catalan, Chechen, Frisian, Georgian, Icelandic, Polish, Serbian, Slovene, Ukrainian
ACTION (11 languages)	Catalan, Chechen, Frisian, Georgian, Icelandic, North Saami, Polish, Slovene, Spanish, Swedish, Ukrainian
DIMINUTIVE (6 languages)	Bulgarian, Greek, Polish, Portuguese, Romanian, Slovene
CAUSATIVE (9 languages)	Chechen, Finnish, Georgian, North Saami, Romanian, Swedish, Tatar, Turkish, Ukrainian

These categories occur multiply in both language types and language families and genera, as shown in the following table.

The data are not sufficient to make any typological conclusions or generalizations. Table 49.42 above maintains in the main the proportions of the language types represented in the sample, except that it lends itself to further comment in that the proportions of the language sample are not replicated by:

- (i) DIMINUTIVE, in that its multiple occurrence is not recorded in any agglutinating or analytical language; and
- (ii) CAUSATIVE, in that it is recorded in more agglutinating than inflectional languages, despite their different proportions in the language sample.

The set of tables above hints at different behaviours for each of the main language genera:

Table 49.52: Semantic categories for which multiple occurrence is attested by language type.

Language type	Semantic categories			
	QUALITY	ACTION	DIMINUTIVE	CAUSATIVE
Inflectional	Bulgarian, Catalan, Icelandic, Polish, Serbian, Slovene, Ukrainian	Catalan, Icelandic, Polish, Slovene, Spanish, Portuguese, Ukrainian	Bulgarian, Greek, Polish,	Romanian, Ukrainian
Agglutinating	Chechen, Georgian	Chechen, Georgian,		Chechen, Finnish, Georgian, North Saami, Tatar, Turkish
Analytical	Frisian		Frisian, Swedish	Swedish

- (i) Germanic languages do not maintain their proportion with respect to other language genera in this regard, except for the category ACTION and, to a much lesser degree, QUALITY.
- (ii) Romance languages do not maintain their proportion with respect to other language genera, except for the categories ACTION and DIMINUTIVE, both to a low degree.
- (iii) Slavic languages maintain their proportion with respect to other language genera for the semantic categories QUALITY, ACTION and DIMINUTIVE, but not for CAUSATIVE.

This confirms what Table 49.51 suggests concerning the multiple occurrence of semantic categories in Slavic languages in general, in this case also with regard to certain individual semantic categories (QUALITY, ACTION and, less markedly, DIMINUTIVE), but not with respect to others (CAUSATIVE). The limited data do not allow similar specific claims to be made with regard to individual semantic categories in Altaic, Nakh-Dagestanian or Uralic languages, which are also suggested as being language families in which the phenomenon of multiple occurrence is identified (Table 49.48).

Various other patterns are hinted at by the data, but again these are of a tentative kind, in the absence of a bigger data set, e.g.:

- (i) Multiple occurrences of the semantic category PROCESS are only recorded in inflectional languages (Catalan, Polish, Russian, and Slovene), but not in all the inflectional languages wherein multiple occurrence is recorded (not in Bulgarian, German, Greek, Icelandic, Latvian, Portuguese, Romanian, Serbian, Slovak, Spanish or Ukrainian).

Table 49.53: Multiple occurrences of the semantic category QUALITY by language family.

			Language family/genus					
Afro-Asiatic	Altaic	Basque	Indo-European			Kartvelian	Nakh-Daghestanian	Uralic
			Baltic	Celtic	Germanic	Greek	Romance	Slavic
			Frisian		Catalan	Bulgarian	Georgian	Chechen
			Icelandic			Polish		
					Serbian			
					Slovene			
					Ukrainian			

Table 49.54: Multiple occurrences of the semantic category ACTION by language family.

			Language family/genus					
Afro-Asiatic	Altaic	Basque	Indo-European			Kartvelian	Nakh-Daghestanian	Uralic
			Baltic	Celtic	Germanic	Greek	Romance	Slavic
			Frisian		Catalan	Polish	Georgian	Chechen
			Icelandic		Spanish	Slovene		
			Swedish			Ukrainian		

Table 49.55: Multiple occurrences of the semantic category DIMINUTIVE by language family.

		Language family/genus							
		Indo-European			Kartvelian			Nakh-Daghestanian	Uralic
Afro-Asiatic	Altaic	Basque	Baltic	Celtic	Germanic	Greek	Romance	Slavic	
						Greek	Portuguese	Bulgarian	
							Romanian	Polish	
								Slovene	

Table 49.56: Multiple occurrences of the semantic category CAUSATIVE by language family.

- (ii) Multiple occurrence of the semantic category ENTITY are only recorded in Slavic languages (Russian, Serbian, Slovak, and Slovene), but not in all the Slavic languages wherein multiple occurrence is recorded (not in Bulgarian, Polish or Ukrainian).

Specific specialized combinations of languages and semantic categories can also be identified in one of two ways:

- (i) With regard to the specialization of languages, in that several languages record multiple occurrences of very few semantic categories, e.g.:

- a. Basque: MANNER and STATE
- b. Finnish: CAUSATIVE
- c. German: AUGMENTATIVE
- d. Greek: DIMINUTIVE
- e. Latvian: SIMILATIVE
- f. Portuguese: DIMINUTIVE
- g. Romanian: CAUSATIVE and DIMINUTIVE
- h. Serbian: ENTITY and QUALITY
- i. Spanish: ACTION
- j. Turkish: CAUSATIVE

- (ii) With regard to the specialization of semantic categories, in that certain semantic categories occur repeatedly in very few or in just one language, e.g.:

- a. AUGMENTATIVE: German and Polish
- b. DISTRIBUTIVE: Polish
- c. FEMALE: Polish
- d. FINITIVE: Slovene
- e. INSTRUMENT: Polish and Russian
- f. ITERATIVE: Polish and Tatar
- g. LOCATION: Polish
- h. PERCEPTIVE: Tatar
- i. PLURIACTIONAL: Bulgarian
- j. REFLEXIVE: Polish and Russian
- k. SIMILATIVE: Latvian
- l. UNDERGOER: Tatar

The multiple occurrence of semantic categories thus gives rise to a number of combinations between languages/language families and categories. Of these, the clearest association is between Slavic languages and the repetitive occurrence of semantic categories in subsequent orders of derivation.

Summary

- (i) Semantic categories occur on multiple occasions in successive orders of derivation, especially in Slavic languages, but which categories do so and how they arrange themselves within languages may vary considerably. This holds both across word-classes and within word-classes.
- (ii) This multiple occurrence may be mediated by an intervening category, or not. Semantic categories may reoccur successively with or without a different intermediate semantic category, as in the sequences ACTION + ACTION vs. PROCESS + FINITIVE + PROCESS, both found in Slovene nouns.
- (iii) Differences can be found between language genera as regards the multiple occurrence of semantic categories, e.g. Germanic and Romance languages do not maintain their proportions compared with other languages, except for the semantic categories ACTION and QUALITY in the former case, and ACTION and DIMINUTIVE in the latter case. The opposite is found for Slavic languages: they do maintain their proportion compared with other genera for the semantic categories QUALITY, ACTION and DIMINUTIVE, but not for CAUSATIVE.
- (iv) The data are not sufficient to make any typological conclusions or generalizations, except in specific cases, e.g. multiple occurrences of DIMINUTIVE are not recorded in any agglutinating or analytical language, while those of CAUSATIVE are recorded in more agglutinating than inflectional languages, despite their different proportions in the language sample.
- (v) Very specific combinations of languages and semantic categories can be identified, such that several languages record multiple occurrences of very few semantic categories, and certain semantic categories occur in very few or in just one language.
- (vi) There is no correlation between the number of orders of derivation and the multiple occurrences of semantic categories. This applies to all word-classes. It is best represented in languages with four and five orders of derivation, though it has also been registered in languages with only two orders of derivation, e.g. adjectives in Chechen or nouns in Frisian.

49.12 Reversibility of semantic categories

The reversibility of semantic categories in subsequent orders of derivation, i.e. the occurrence of derivatives of both AB and BA orders of two semantic categories in a language, is not a frequent or well-represented phenomenon in the 40 European languages under study. Reversibility has been identified in 14 out of

the 40 languages: in their derivational networks, Croatian, English, Estonian, Finnish, German, Greek, Hungarian, Maltese, Romanian, Slovene, Spanish, Tatar, Turkish and Welsh display a reversed ordering of semantic categories in subsequent orders of derivation.

These languages cover all language types and five out of the seven language families represented in the sample, as shown in Tables 49.57 and 49.58.

Table 49.57: Languages where the reversibility of semantic categories can be attested by language type.

Language type	Languages
Inflectional	Croatian, German, Greek, Maltese, ^a Romanian, Slovene, Spanish, Welsh
Agglutinating	Estonian, Finnish, Hungarian, Tatar, Turkish
Analytical	English

^aMaltese is generally not recognized as an inflectional language. At best, it could be described as a hybrid language combining introflexion (root-and-vowel pattern) and affixation. Adhering to Sapir's typological classes, we consider it best, with the qualifications made here, to include Maltese in the group of languages with inflectional typological characteristics.

Of these, Table 49.57 maintains in the main the proportions of the language types represented in the sample, whereas Table 49.58 lends itself to further comment in that the proportions of the language sample are not replicated by:

- (i) Germanic, Romance or Slavic languages with respect to less represented language genera or families in the total language sample, but which attest the same number of languages in this regard, e.g. Turkish from the Altaic family and Uralic.
- (ii) Altaic languages, insofar as the two languages of the group included in the sample are represented here.
- (iii) Latvian and Lithuanian, which are not reported here for the reason that, in Baltic languages, instances of reversibility are perceived as and reported in terms of alternative derivational interpretations – e.g. Lithuanian *deg-ti* 'burn (intr.)' > suffixal CAUSATIVE, *deg-in-ti* 'burn (tr.)' > prefixal FINITIVE *su-deg-in-ti* 'burn down (tr.)'.

Other languages not listed above report occasional or exceptional reversibility (Czech) or, in the case of the Baltic genus, the phenomenon is reinterpreted in different terms.

Table 49.58: Languages where the reversibility of semantic categories can be attested by language family.

		Language family/genus							
Afro-Asiatic	Altaic	Basque	Indo-European				Kartvelian	Nakh-Daghestanian	Uralic
			Baltic	Celtic	Germanic	Greek			
Maltese	Tatar	Estonian	Welsh	English	Greek	Romanian	Croatian		Finnish
	Turkish			German		Spanish	Slovene		Hungarian

Among these, the possibilities are rather limited in number and refer to one or two combinations of reversible semantic categories. Only in three languages (Romanian, Tatar, Welsh) are the categories not limited to one or two pairs of semantic categories that permit a sequential exchange, as summarized in Table 49.59.

Table 49.59 shows that the languages wherein most reversible combinations of semantic categories occur are as follows:

(i) Romanian:

- a. ABILITY + ITERATIVE
- b. CAUSATIVE + ENTITY
- c. PRIVATIVE + STATIVE
- d. QUALITY + STATIVE

(ii) Tatar:

- a. ABSTRACTION + QUALITY CAUSATIVE + PLURIACTIONALITY/RECIPROCAL/UNDERGOER/REFLEXIVE/PROCESS
- b. ITERATIVE + RECIPROCAL
- c. QUALITY + STATIVE
- d. AGENT + STATIVE
- e. ABSTRACTION + QUALITY
- f. QUALITY + UNDERGOER

(iii) Welsh:

- a. ABSTRACTION + PRIVATIVE
- b. ABSTRACTION + QUALITY
- c. ABSTRACTION + REFLEXIVE
- d. CAUSATIVE + QUALITY
- e. PRIVATIVE + QUALITY

The limited amount of data does not allow the identification of patterns in the distribution of these reversible combinations by language type, by language genus or family, or by areal distribution. Similarly, there is no pair of semantic categories that are clearly the most frequent pair across the board. The most frequent reversible categories are:

- (i) CAUSATIVE (in three combinations and in five languages)
- (ii) PRIVATIVE (in five combinations and in five languages)
- (iii) QUALITY (in five combinations and in five languages)

The most frequent combinations of reversible categories are as follows, but again no pattern can be identified in their distribution due to the limited amount of data:

- (i) CAUSATIVE + PROCESS
- (ii) QUALITY + STATIVE

Table 49.59: Languages where the reversibility of semantic categories can be attested.

Language	Semantic categories (in one of the possible orders)	
Croatian	ATTENUATIVE RELATIVE	
English	ABILITY PRIVATIVE	
Estonian		CAUSATIVE PROCESS
Finnish		QUALITY MANNER
German	ABILITY PRIVATIVE	PATIENT SATURATIVE
Greek		QUALITY ^a STATIVE
Hungarian		
Maltese		CAUSATIVE PROCESS
Romanian	ABILITY ITERATIVE	CAUSATIVE ENTITY
Slovene		PRIVATIVE MANNER
Spanish		QUALITY STATIVE
Tatar		CAUSATIVE + various types e.g. REFLEXIVE, UNDERGOR
Turkish	AGENT STATE	CAUSATIVE PROCESS
Welsh	ABSTR. PRIVATIVE	CAUSATIVE QUALITY
	ABSTR. QUALITY	PRIVATIVE QUALITY

^aIn this case, STATIVE/ABSTRACTION.

Summary

- (i) The systematic reversibility of semantic categories is represented in five out of the seven language families of the sample.
- (ii) No patterns can be identified as regards language type, language genus or family, or areal distribution, due to the limited amount of data.

49.13 Reasons for structurally poor derivational networks

The size and diversity of the data sample used here also have an effect on the overall picture as regards, firstly, whether the derivational networks can be considered comparatively rich or poor and, secondly, the reasons for comparatively poorer networks.

Restrictions for richer derivational networks have been identified in 16 out of the 40 languages: Basque, Chechen, Danish, Dargwa, Dutch, English, French, Frisian, Icelandic, Irish, Latvian, Maltese, Norwegian, Portuguese, Tatar and Welsh.

These languages cover all language types and five out of the seven language families represented in the sample, as shown in Table 49.60.

Table 49.60: Languages for which restrictions on derivational networks have been identified by language type.

Language type	Languages
Inflectional	Dutch, French, Icelandic, Irish, Latvian, Maltese, Portuguese, Welsh
Agglutinating	Basque, Chechen, Dargwa, Tatar
Analytical	Danish, English, Frisian, Norwegian

Notably, the proportion of language types is not maintained here, and analytical languages stand out as not producing rich derivational networks (4 out of 5), compared to agglutinating languages (4 out of 10) and inflectional languages (8 out of 25). Although these percentages do not lend themselves to strong statements, they contrast sharply with what appears to be a rather even distribution as regards language genus or family, as shown in Table 49.61.

Table 49.61: Languages for which restrictions on derivational networks have been identified by language family.

It is worth noting, however, that no Slavic language reports poor derivational networks, especially when compared with the other two language genera that are the same approximate size and report limited derivational networks too (Germanic and Romance).

Several languages make less use of the derivation resources studied here only by comparison with other languages, e.g. Dargwa, which is reported to produce apparently poor derivational networks, but only when compared with the networks of other languages. This is not relevant in this section, however.

In the languages where poor derivational networks are reported, two major arguments are cited:

- (i) Limitations as a result of methodological decisions. These are as a consequence of three decisions:
 - a. The sample under study, as the lexical entries that make up the sample for the production of derivational networks may not lend themselves to derivation, especially in:
 - i. Dutch
 - ii. Finnish, as the entries of the sample do not cover all possible morphological and semantic types
 - iii. German, due to the representation of certain semantic categories in verbs (even if some of these have been interpreted not only as derivation but also as compounding)
 - iv. Maltese
 - v. Tatar, although in this case, specific entries are referred to rather than the sample in general
 - b. The processes excluded from the concept of the ‘derivational paradigm’, that is, processes that play a major role in the formation of new words in a given language but which are not covered by this piece of research. The processes cited in this respect are relatively few, and refer systematically to the following:
 - i. Suppletion or morphologically simple forms⁴:
 - 1. French
 - ii. Conversion, which is reported to be responsible for a large amount of word-formation in the following languages⁵:
 - 1. Basque, specifically with regard to conversion to verbs
 - 2. Dargwa
 - 3. English

⁴ This is also cited as a constraint on richer derivational paradigms in Romanian.

⁵ This is also cited as a constraint on richer derivational paradigms in Dutch and Romanian.

4. French, especially for derivation of nouns from adjectives and verbs
 5. Icelandic
 6. Portuguese
- iii. Compounding, which is reported to be responsible for a large amount of word-formation in the following languages:
 1. English
 2. French
 3. Icelandic
 4. Irish
 5. Latvian, specifically for nouns
 6. Norwegian
 7. Portuguese
 8. Tatar
 9. Welsh
 - iv. Reduplication, especially for the expression of evaluative meaning:
 1. French
 - v. Participial forms, which occur in:
 1. Finnish
 2. Portuguese
 - vi. Bound forms or affixoids, which occur in:
 1. Icelandic
 2. Welsh
 - vii. Particle verbs, which are cited as being frequent only for Dutch.
- c. The attestation method, which may in some languages create difficulties due to limited resource availability:
 - i. Underdeveloped lexicographical resources:
 1. Chechen
 - ii. Irregular attestation, such that entries that are not attested in corpora could have been attested in dictionaries or by native speakers:
 1. Maltese
- (ii) Limitations inherent to each language. These are language-specific conditions, and cover the following:
 - a. Constraints in the formation of neologisms:
 - i. Chechen
 - b. Constraints in the use of derivation for word-formation:
 - i. Danish

- ii. Frisian, in this case also with a rapidly decreasing number of derivatives from one order of derivation to the next
 - iii. Slovak, as nouns are reported to function as motivated units instead of as motivating units
- c. The influence of bilingualism, specifically where a contact language supplies word-formation resources that partly replace the word-formation resources of the language under study:
 - i. Chechen
 - d. The influence of native and borrowed morphology:
 - i. English
 - ii. Irish
 - iii. Portuguese, in this case relating to the borrowing of morphemes, not whole words
 - e. The use of compensating naming strategies, specifically:
 - i. Morphological resources:
 - 1. Icelandic, especially with regard to a rich use of inflection for the representation of some of the semantic categories under study
 - 2. Swedish, for the semantic categories not represented, even if, in general, the derivational networks of Swedish are not comparatively poor
 - 3. Ukrainian, even if the derivational capacity of the semantic categories recorded is quite rich
 - 4. Basque, especially with regard to affective palatalization and reduplication
 - ii. Syntactic resources, like phrases:
 - 1. French, e.g. with constructions involving several word-classes
 - 2. Frisian, e.g. with prepositional phrases where the derivational meaning is represented lexically
 - 3. Icelandic
 - 4. Irish
 - 5. Norwegian
 - 6. Swedish, again for the semantic categories not represented, even if, in general, the derivational networks of Swedish are not comparatively poor
 - 7. Ukrainian, even if the derivational capacity of the semantic categories recorded is quite rich
 - 8. Welsh

- iii. Semantic resources, like semantic extension/narrowing, borrowing, or the reuse of old vocabulary⁶:
- 1. Irish

Both of these types of limitations lend themselves to further analysis. Some limitations, specifically the role of alternative resources for the expression of certain semantic categories, are both methodological and language-specific, and could have been listed as either. The focus is, thus, on point b) of the methodological limitations and point e) of the language-specific limitations, in that they hint at the strategies employed as compensating mechanisms, or, more precisely, as alternative naming procedures to derivation.

Three cases stand out here. The first is compounding, which is reported in nine languages, but no relevant pattern can be found therein because the languages in question replicate, in the main, a cross-section of the entire language sample:

- (i) as regards language type, of these nine languages, six are inflectional, two are analytical, and one is agglutinating.
- (ii) as regards the language genus or family, of these nine languages, three are Germanic, two are Celtic, two are Romance, one is Baltic, and one belongs to the Altaic family.

These data are shown in Tables 49.62 and 49.63.

Table 49.62: Languages where compounding is reported as an alternative process for derivation and is partly responsible for comparatively poor derivational networks by language type.

Language type	Languages
Inflectional	French, Icelandic, Irish, Latvian, Portuguese, Welsh
Agglutinating	Tatar
Analytical	English, Norwegian

Table 49.63 lends itself to further comment in that the proportions of the language sample are not replicated by:

⁶ This is also cited as a constraint on richer derivational paradigms in Latvian, Romanian and Ukrainian.

Table 49.63: Languages where compounding is reported as an alternative process for derivation and is partly responsible for comparatively poor derivational networks by language family.

		Language family/genus								
		Indo-European			Kartvelian			Nakh-Daghestanian		Uralic
Afro-Asiatic	Altaic	Basque	Baltic	Celtic	Germanic	Greek	Romance			
Tatar			Latvian	Irish Welsh	English			French Portuguese		
					Icelandic					
						Norwegian				

- (i) Celtic languages, insofar as the two languages of the sample are cited here.
- (ii) Slavic languages, insofar as none of the nine languages of the sample are cited here.

The second case that stands out here is conversion, but to a lesser degree: it is reported in six languages where no relevant pattern can be found, because the languages in question replicate, in the main, a cross-section of the entire language sample:

- (i) as regards language type, of these six languages, four are inflectional, one is analytical, and one is agglutinating.
- (ii) as regards the language genus or family, of these six languages, three are Germanic, two are Romance, and the last one is Basque.

These data are shown in Tables 49.64 and 49.65.

Table 49.64: Languages where conversion is reported as an alternative process for derivation and is partly responsible for comparatively poor derivational networks by language type.

Language type	Languages
Inflectional	Dutch, French, Icelandic, Portuguese
Agglutinating	Basque
Analytical	English

Just as for compounding, the latter table (Table 49.67) lends itself to further comment in that the proportions of the language sample are not replicated by Slavic languages, insofar as none of the nine languages of the sample are cited here.

The third case is the use of syntactic structures for the expression of semantic categories, often in the form of phrases involving lexical bases that instead use derivation in other languages of the sample.⁷ These cases are shown in Tables 49.66 and 49.67.

This table hints that the language sample is not fully represented in this respect, in that agglutinating languages are not reported as using this resource.

Table 49.67 lends itself to further comment in that the proportions of the language sample are not replicated by:

⁷ This is also cited as a constraint on richer derivational paradigms in Swedish and Ukrainian.

Table 49.65: Languages where conversion is reported as an alternative process for derivation and is partly responsible for comparatively poor derivational networks by language family.

Afro-Asiatic	Altaic	Basque	Language family/genus								
			Baltic	Celtic	Germanic	Greek	Indo-European	Romance	Slavic	Kartvelian	Nakh-Daghestanian
		Basque			Dutch English Icelandic		French Portuguese				

Table 49.66: Languages where syntactic resources are reported as an alternative process for derivation and are partly responsible for comparatively poor derivational networks by language type.

Language type	Languages
Inflectional	French, Icelandic, Irish, Welsh
Agglutinating	
Analytical	Frisian, Norwegian

- (i) Multiple language families, insofar as only genera from the Indo-European family are cited here.
- (ii) Celtic languages, insofar as the two languages of the sample are cited here.
- (iii) Romance and Slavic languages, insofar as only one of their seven and nine languages are cited here.

The sample also reveals constraints in the representation of certain semantic categories, which results in poorer derivational networks. These constraints refer to the rare use of the categories represented in the languages listed in Table 49.68.

Finnish is used as an exemplar here to showcase lack of or rare occurrence of specific semantic categories in languages. There is great variability in terms of the specific semantic categories reported by different languages.

Without making a mountain out of a molehill, it can safely be suggested that comparative semantic categories are very useful as a means to an end, i.e. they are indispensable for generating derivational networks and establishing correlations between orders of derivation and preferences for their expression in respective orders per language and per group. Despite the enormous diversity in their combinability, such comparative concepts can be used as a criterion for descriptive generalizations for particular groups.

Summary

- (i) Constraints on derivational networks are relatively rare compared with derivational networks, where no major restrictions apply.
- (ii) Slavic languages stand out by not reporting poor derivational networks in any case.
- (iii) Alternatives to derivation refer mainly to compounding, conversion, and syntactic structures.

Table 49.67: Languages where syntactic resources are reported as an alternative process for derivation and are partly responsible for comparatively poor derivational networks by language family.

Afro-Asiatic	Altaic	Basque	Language family/genus							
			Baltic	Celtic	Indo-European	Germanic	Greek	Romance	Slavic	Kartvelian
					Irish Welsh	Frisian	Icelandic	French		
									Norwegian	

Table 49.68: An illustration of semantic categories not attested or reported to be rare in languages.

Languages						
Finnish	Frisian	Georgian	German	Norwegian	Ukrainian	Welsh
ANTICAUSATIVE						
AUGMENTATIVE						
COLLECTIVE						
DIMINUTIVE						
ENTITY						
EXPERIENCER						
INSTRUMENTATIVE						
LOCATION						
REFLEXIVE						
RESULTATIVE						
SUBITIVE						

49.14 Conclusions

The typological research presented in this chapter is based on all the language-specific descriptions in the preceding chapters. Our observations can be summarized as follows:

- (i) There are considerable *differences among languages in their derivational capacity*, which is reflected in the number of derivatives in derivational networks. It suffices to compare any derivational network of, for example, Croatian to a derivational network of, for example, English. This difference usually amounts to several dozen derivatives.
- (ii) If we compare the average maximum derivational network (MDN) values by word-class and by order of derivation, it is obvious that the derivational potential of simple underived nouns and adjectives is very similar, and in some orders almost identical. *Verbs clearly have the highest MDN values* in every order of derivation, and they are significantly higher than those of the other two word-classes. This is especially due to the extreme derivational potential of those languages

- that employ prefixes for the expression of the category of *aktionsart*. One possible explanation for this supremacy of verbs can be sought in the *derivational construal* (Croft 2012: 17) potential of verbs in view of the fact that, onomasiologically speaking, they are the locus of the lexical semantic encoding of event structure (i.e. they can function as lexicalized construal carriers of both aspectual and causal structure⁸).
- (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. High consistency across all three orders in all three classes is rare, but it does occur in Bulgarian and Serbian. When restricted to the 1st order, highly consistent networks in all three word-classes have been identified for Croatian, Turkish and Basque, and, to an extent, Bulgarian, Polish and Welsh.
 - (iv) The richness of derivational networks is *sensitive to the order of derivation*.
 - (v) The richness of derivational networks is probably also sensitive to the semantics of base words, but this hypothesis is in need of further empirical corroboration.
 - (vi) 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.
 - (vii) There is a *core group* of languages that keep *high saturation values (SVs) across all three word-classes*. This comprises Greek, Dutch, Dargwa and North Saami. These might be joined by German, Turkish and Lithuanian, which have high values in two word-classes and a medium SV in the third word-class.
 - (viii) There is an unambiguous tendency for SVs to fall gradually as the order of derivation rises in *all three word-classes*. It occurs in 28 languages for nouns, and 27 languages for both verbs and adjectives. This suggests that the tendency to derive fewer words as the order of derivation increases is independent of the word-class.
 - (ix) The SVs 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*.

⁸ See Croft (2012) for an elaborate account of the way in which verbs can, in terms of construction grammar and cognitive linguistics, lexically map various profiles of a concept via the derivational construal.

- (x) A *medium SV* (20%–30%) can be considered the most typical SV for all word-classes and for the first three orders of derivation.
- (xi) There is no geographically homogeneous territory in which the languages of the topmost SVs are spoken. These languages are of various genetic origins and are scattered across Europe. What can be considered as a general tendency, however, is *the use of low-SV languages in geographically peripheral areas of Europe*.
- (xii) The data suggest that derivational networks are *most predictable in the 1st order*. This is manifested by a high number of languages with a level of standard deviation (SD) below 10.00 as well as by the generally relatively high consistency of this value in the other languages. The consistency of results falls as the order grows, which means that derivational networks are much less predictable in the higher orders of derivation.
- (xiii) The correlation between SV and paradigmatic capacity may differ significantly in the same language in different word-classes and different orders of derivation.
- (xiv) 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 that reach five orders of derivation in all three word-classes, none of which belong to the Romance or Germanic genera. The average number of affixation steps is very similar for verb-based and adjective-based derivations (2.78 and 2.76, respectively). This figure is lower for nouns (2.46).
- (xv) In terms of the total number of derivatives, verbs have the most prolific base. The average number of verb-based derivatives is clearly greater than the figures for adjectives and nouns. This word-formation feature is dominated by Slavic and Uralic languages. The values for adjective-based derivations are slightly higher than those for noun-based derivations. This result tallies with the data on the maximum derivational networks for individual word-classes, the data on the average number of orders of derivation, as well as the data on the correlation between the SV and the paradigmatic capacity.
- (xvi) The non-existence of a particular word-formation process or a minimum number of word-formation types of a particular word-formation process does not correlate with the richness of a derivational network.
- (xvii) 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 a smaller number of derivational orders than Slavic languages. While Nakh-Daghestanian languages, classified as agglutinating, tend to have a very low number of

- derivational orders, Uralic languages, which are also agglutinating, feature high numbers of orders of derivation. Analytical languages are not consistent in their behaviour. Generally, however, they have a lower number of derivational orders, especially in the case of nouns.
- (xviii) Correlations between semantic categories and orders of derivation are reported in almost all languages as systematic occurrences in the 1st order of derivation and by the majority in the 2nd order of derivation, although only those semantic categories with a notable occurrence (i.e. those that were present in a significant number of the languages in the sample) were discussed above.
 - (xix) Regarding correlations, **DIMINUTIVE** appears to be correlated with the 1st order of derivation from all three types of bases. Otherwise, it appears that the correlations are base-sensitive. In several languages, a correlation between the 3rd order of derivation and specific semantic categories has been established, but the semantic categories are so diverse that there are no more than six languages in which the same category correlates with this order of derivation. No clear correlations for the 4th and 5th orders of derivation were observed. Only one language (Norwegian) reports a correlation between the 5th order of derivation from all three bases (i.e. nominal, verbal and adjectival) and the semantic category **STATIVE**.
 - (xx) Semantic categories with blocking effects are reported in all the languages of the sample except Welsh. However, few semantic categories report blocking effects through all the word-classes. Similarly, few languages report blocking in the 4th order of derivation. Several languages record consistent blocking by semantic categories across orders of derivation, e.g. Hungarian and North Saami. In the orders of derivation where several semantic categories have a blocking effect, the tendency is for languages to have only one semantic category with a blocking effect; however, there are some languages that consistently report several blocking effects, regardless of the word-class or order of derivation, e.g. Hungarian and Tatar.
 - (xxi) No language type, language genus, or family or areal distribution pattern can be identified as regards the distribution of semantic categories with a blocking effect, but a set of semantic categories that rarely have a blocking effect in the sample can be listed:
 - a. **ATTENUATIVE**, in the 2nd order of Greek nouns
 - b. **ANTICAUSATIVE**, in the 1st order of Maltese verbs
 - c. **CONCOMITANT**, in the 1st order of Greek verbs
 - d. **CUMULATIVE**, in the 2nd order of Ukrainian verbs
 - e. **DISTRIBUTIVE**, in the 3rd order of Polish verbs

- f. INCEPTIVE, in the 1st order of Ukrainian verbs
 - g. POSSESSIVE, in the 1st order of Hungarian nouns
- (xxii) Typical combinations of semantic categories are reported for most of the languages in the sample (though not for Welsh) and regularly for each word-class (though not, for example, for adjectives in Dargwa, nouns in Chechen, Dargwa, Maltese and Polish, or verbs in Irish). Each word-class starts out with specific typical combinations, such that the same sequence is typically not found across the three word-classes in the sample, except in certain languages, e.g. QUALITY + ABSTRACTION in Czech adjectives, nouns and verbs, or ACTION + AGENT in Georgian nouns and verbs. No associations can be identified here with regard to language type, language genus, or family or areal distribution. Few typical combinations that are specific to word-classes can be found across languages: an exception is QUALITY + STATE in Chechen, English, German and Ukrainian adjectives. Again, no associations can be identified here with regard to language type, language genus, or family or areal distribution.
- (xxiii) Typical starting categories that are frequent in each word-class can be identified to show biases of certain languages towards specific categories in specific word-classes, e.g. Bulgarian adjectives towards PATIENT. Typical starting categories that occur in each word-class can be identified to show categories that stand out due to their frequency, regardless of word-class and language, e.g.:
- a. In adjectives, ABILITY, ACTION, AGENT, CAUSATIVE, QUALITY and SIMILATIVE
 - b. In nouns, AGENT, QUALITY, PRIVATIVE and SIMILATIVE
 - c. In verbs, AGENT, CAUSATIVE, DIRECTIONAL, FINITIVE, PROCESS and QUALITY
- No pattern can be identified by language type, language genus, or family or areal distribution for these typical combinations, but certain combinations that have a range of subsequent categories can be attested at least four times across languages:
- d. In adjectives, QUALITY + STATE (+) in Chechen, German, English and Ukrainian
 - e. In verbs, AGENT + FEMALE (+) in Bulgarian, Croatian, German, Serbian and Slovak
- Certain combinations are attested only once, even regardless of the word-class, e.g. TEMPORAL + QUANTITY (+) in Catalan nouns or ANTICAUSATIVE + ENTITY in Icelandic verbs, to name only two examples.
- (xxiv) The multiple occurrence of a semantic category (i.e. its recurrence in successive orders of derivation) is a relatively frequent phenomenon in the 40 European languages under study. It is comparatively higher in Slavic languages than in other language genera or families in the sample.

- (xxv) There is considerable variation in the multiple occurrences of semantic categories, ranging from languages where only one category reoccurs to languages where over ten categories may reoccur for one word-class. Semantic categories may reoccur successively with or without a different intermediate semantic category. Differences can be found between language types as regards the multiple occurrence of semantic categories:
- DIMINUTIVE is not recorded in any agglutinating or analytical language.
 - CAUSATIVE is recorded in more agglutinating than inflectional languages, despite their different proportions in the language sample.
- (xxvi) Differences can be found between language genera as regards the multiple occurrence of semantic categories, which hint that:
- Germanic languages are biased towards the category ACTION and, to a much lesser degree, QUALITY.
 - Romance languages are biased towards the categories ACTION and DIMINUTIVE, both to a low degree.
 - Slavic languages are biased towards the categories QUALITY, ACTION and DIMINUTIVE, but not for CAUSATIVE.
- Very specific combinations can be identified, such that several languages record multiple occurrences of very few semantic categories, and certain semantic categories occur in very few or in just one language.
- (xxvii) The systematic reversibility of semantic categories is not a characteristic property of European languages. The languages that display semantic reversibility do this only with regard to one or two categories, so only exceptionally do a higher number of categories or combinations allow reversibility. No patterns can be identified as regards language type, language genus, or family or areal distribution, due to the limited amount of data.
- (xxviii) Constraints on derivational networks are relatively infrequent compared with derivational networks where no major restrictions apply. Slavic languages stand out due to not reporting any cases of poor derivational networks, especially compared with the other two language genera that are their same approximate size and report limited derivational networks (Germanic and Romance). Poor derivational networks may be due to methodological decisions or language-inherent issues. The method used constrained the derivational networks due to the lexical sample used, due to what was considered to fall within the scope of derivation, and due to limitations in the attestation method used. Alternatives to derivation refer mainly to compounding, conversion and syntactic structures, but apparently comparatively less frequently in agglutinating languages and, as mentioned above, in Slavic languages.

References

- Aronoff M. & N. Fuhrhop. 2002. Restricting suffix combinations in German and English: Closing suffixes and the monosuffix constraint. *Natural Language and Linguistic Theory* 20. 451–490.
- Bagasheva, Alexandra. 2017. Comparative semantic concepts in affixation. In J. Santana-Lario and S. Valera (eds.), *Competing patterns in English affixation*, 33–65. Bern: Peter Lang.
- Bakema, P. & D. Geeraerts. 2000. Diminution and augmentation. In G. E. Booij, Ch. Lehmann & J. Mugdan (eds.), *Morphologie/Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word Formation*, Vol. 1, 1045–1052. Berlin: de Gruyter.
- Croft, W. 2012. *Verbs. Aspect and Causal Structure*. Oxford: Oxford University Press.
- Fabb, N. 1988. English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory* 6. 527–539.
- Gaeta, L. (2005). Combinazioni di suffissi in Italiano. In M. Grossmann & A. Thornton (eds.), *La Formazione delle Parole. Atti del XXXVII Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI)*, 229–247. Roma: Bulzoni.
- Giegerich, H. J. 1999. *Lexical strata in English. Morphological causes, phonological effects*. Cambridge: Cambridge University Press.
- Hay J. & I. Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22. 565–596.
- Körtvélyessy, Lívia. 2015. *Evaluative Morphology from a Cross-Linguistic Perspective*. Newcastle: Cambridge Scholars Publishing.
- Körtvélyessy, L., P. Štekauer, J. Genčí & J. Zimmermann. 2018. Word-Formation in European languages. *Word Structure* 11 (3). 313–358.
- Manova, S. (ed.). 2015. *Affix ordering across languages and frameworks*. Oxford: Oxford University Press.
- Manova, S. & Aronoff, M. 2010. Modelling affix order. *Morphology* 20 (1). 109–131.
- Müller, P.O., Ohnheiser, I., Olsen, S. & Rainer, F. (eds.). 2015–2016. *Word-Formation. An International Handbook of the Languages of Europe*. Berlin: Mouton de Gruyter.
- Plag, I. 1996. Selectional restrictions in English suffixation revisited. A reply to Fabb (1988). *Linguistics* 34. 769–798.
- Plank, F. 1994. Inflection and derivation. In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*, vol. 3, 1671–1678. Oxford: Pergamon Press.
- Saarinen, P. & J. Hay. 2014. Affix ordering in derivation. In R. Lieber & P. Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 370–383. Oxford: Oxford University Press.
- Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.
- Štekauer, P. 2012. Morphological types vs. word-formation: any correlation? In E. Cyran, H. Kardela & B. Szymanek (eds.), *Sound structure and sense. Studies in memory of Edmund Gussmann*, 711–728. Lublin: KUL.
- Štekauer, P., S. Valera & L. Körtvélyessy. 2012. *Word-Formation in the World's Languages. A Typological Survey*. Cambridge: Cambridge University Press