

# Analogical classification in formal grammar

Matías Guzmán Naranjo

Draft  
of May 23, 2019, 11:45

# Empirically Oriented Theoretical Morphology and Syntax

Chief Editor: Stefan Müller

Consulting Editors: Berthold Crysmann, Laura Kallmeyer

In this series:

1. Lichte, Timm. Syntax und Valenz: Zur Modellierung kohärenter und elliptischer Strukturen mit Baumadjunktionsgrammatiken
2. Bîlbîie, Gabriela. Grammaire des constructions elliptiques: Une étude comparative des phrases sans verbe en roumain et en français
3. Bower, Claire, Laurence Horn & Raffaella Zanuttini (eds.). On looking into words (and beyond): Structures, Relations, Analyses

ISSN: 2366-3529

# Analogical classification in formal grammar

Matías Guzmán Naranjo


Guzmán Naranjo, Matías. 2019. *Analogical classification in formal grammar* (Empirically Oriented Theoretical Morphology and Syntax 5). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/186>

© 2019, Matías Guzmán Naranjo

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>  <http://langsci-press.org/graphics/ccby.eps>

ISBN: 978-3-96110-186-3 (Digital)

978-3-96110-187-0 (Hardcover)

ISSN: 2366-3529

DOI: [10.5281/zenodo.3191825](https://doi.org/10.5281/zenodo.3191825)

Source code available from [www.github.com/langsci/186](https://www.github.com/langsci/186)

Collaborative reading: [paperhive.org/documents/remote?type=langsci&id=186](https://paperhive.org/documents/remote?type=langsci&id=186)

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe<sub>La</sub>T<sub>E</sub>X

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin

 <http://langsci-press.org/graphics/storage/logo.pdf>

# Contents

Acknowledgments	v
Abbreviations	vii
1 Introduction	1
2 Remarks on analogy	3
2.1 The many meanings of analogy . . . . .	3
2.1.1 Single case analogy . . . . .	3
2.1.2 Proportional analogies . . . . .	6
2.1.3 Analogical classifiers . . . . .	9
2.1.4 Summing up . . . . .	14
2.2 The mechanism for analogy . . . . .	14
2.2.1 Simple rules . . . . .	15
2.2.2 Schemata . . . . .	18
2.2.3 Multiple-rule systems . . . . .	20
2.2.4 Neural networks and analogical modelling . . . . .	23
2.2.5 Analogy or rules . . . . .	25
2.2.6 Mental representations vs grammatical relations . . . . .	28
2.2.7 Summing up . . . . .	29
2.3 Missing pieces . . . . .	30
2.4 Final considerations . . . . .	33
3 Modelling analogy in grammar	35
3.1 Basic assumptions . . . . .	35
3.1.1 Feature structures . . . . .	35
3.1.2 Type hierarchies . . . . .	36
3.2 Analogy as type constraints . . . . .	40
3.2.1 Analogy is categorical . . . . .	42
3.2.2 Analogy runs through the hierarchy . . . . .	43
3.3 The (semi-)formal model . . . . .	45
3.4 Final remarks . . . . .	51

<b>4</b>	<b>Methodological notes</b>	<b>53</b>
4.1	On the general methodology . . . . .	53
4.2	Statistical models and methodology . . . . .	54
4.3	Analogical models using neural networks . . . . .	57
4.4	Measuring variable importance . . . . .	67
4.5	Clustering and distances between classes . . . . .	68
4.6	Summing up . . . . .	72
<b>5</b>	<b>Gender systems</b>	<b>73</b>
5.1	Masculine-feminine syncretism: Latin . . . . .	75
5.1.1	The Latin third declension . . . . .	75
5.1.2	Data . . . . .	76
5.1.3	Results . . . . .	77
5.2	Gender vs inflection class: Romanian . . . . .	79
5.2.1	The Romanian gender and plural system . . . . .	79
5.2.2	Modelling the system . . . . .	85
5.2.3	Data . . . . .	91
5.2.4	Results . . . . .	93
5.3	Interim conclusion . . . . .	103
<b>6</b>	<b>Hybrid classes</b>	<b>105</b>
6.1	Overabundant inflection: Croatian singular instrumental . . . . .	105
6.1.1	Modelling the system . . . . .	107
6.1.2	Materials . . . . .	108
6.1.3	Results . . . . .	108
6.2	Frequency and analogical similarity: Russian diminutives . . . . .	111
6.2.1	Russian diminutives . . . . .	111
6.2.2	Modelling the system . . . . .	114
6.2.3	The dataset . . . . .	115
6.2.4	Results . . . . .	116
6.3	Interim conclusion . . . . .	118
<b>7</b>	<b>Morphological processes and analogy</b>	<b>119</b>
7.1	Prefixes and gender: Swahili noun classes . . . . .	119
7.1.1	Materials . . . . .	122
7.1.2	Results . . . . .	123
7.2	Prefixes and inflection classes: Eastern Highland Otomi . . . . .	127
7.2.1	Verb classes in Eastern Highland Otomi . . . . .	127
7.2.2	Materials . . . . .	128

7.2.3	Results	128
7.3	Stem changing processes: Hausa plural classes	130
7.3.1	The Hausa plural system	130
7.3.2	Materials	131
7.3.3	Results	132
7.4	Interim conclusion	135
<b>8</b>	<b>Complex inflectional classes</b>	<b>139</b>
8.1	Multiple inheritance and cross-hierarchies: Spanish verbal inflection	139
8.1.1	Spanish inflection classes	139
8.1.2	Previous takes on the Spanish verbal system	143
8.1.3	Modelling the system	146
8.1.4	Materials	152
8.1.5	Results	153
8.2	Cross-classifications between plural and singular: Kasem	164
8.2.1	ATR in Kasem	166
8.2.2	A simple analysis of Kasem noun classes	167
8.2.3	Materials	180
8.2.4	Modelling the system	181
8.2.5	Methodological considerations	185
8.2.6	Results	188
8.3	Interim Conclusion	207
<b>9</b>	<b>Concluding remarks</b>	<b>209</b>
9.1	The path forward	209
9.1.1	The limits of analogy	209
9.1.2	Analogical classifiers or proportional analogies	209
9.1.3	The features of analogy	210
9.1.4	Coverage	210
9.2	Final considerations	211
	<b>References</b>	<b>213</b>
	<b>Index</b>	<b>235</b>
	Name index	235
	Language index	235
	Subject index	237







## 6 Hybrid classes

This short chapter explores the situations where morphology allows for two different, competing strategies to be applied to the same lexeme. In inflection, this is called overabundance, while in derivation the phenomenon is called derivational doublets. These systems can be modelled using hybrid classes, like the ones introduced in Chapter 3. As for the test cases, I look at Russian diminutives (derivational doublets) and Croatian instrumental singular markers (overabundance). The interesting aspect of these two cases is that there is partial overlap between class membership, which contrasts with the previous gender and inflection class cases, where a stem or word can only belong to one inflection class at a time. In cases of affix competition and overabundance, a single lexeme can belong to two different classes at the same time, and giving speakers a choice between two formally different, but semantically identical, markers. This produces hierarchies with a different shape, which has a clear effect on the analogical relations.

### 6.1 Overabundant inflection: Croatian singular instrumental

In BCS (Bosnian, Croatian and Serbian), a number of masculine nouns belonging to the first (or *-a*) declension present partial overabundance between the markers *-em* (/jem/) and *-om* (/om/) in the instrumental singular, as shown in (1):

- (1) a. grad-**om** ‘city’-INSTR
- b. muž-**em** ‘man’-INSTR
- c. princ-**om**/princ-**em** ‘prince’-INSTR

Importantly, not all nouns can alternate between the two markers:

- (2) a. kej-om ‘river bank’-INSTR
- b. \*kej-em
- c. muž-em ‘man’-INSTR
- d. \*muž-om

A rule of thumb for class assignment has been proposed in the literature already: “nouns ending in a palatal phoneme use *-em*, whereas all other nouns use *-om*. However, although this rule seems reasonably straightforward, there are some environments where doublets occur” (Lečić 2015: 377). Diachronically, this overabundance (Thornton 2011; 2010a), emerged due to the collapse of an older palatal /rʲ/, which justified the use of *-em*, with a modern non-palatal /r/ which justifies the use of *-om* (Lečić 2015). Mladenović (1977) (as cited by Lečić 2015) claims that *-om* is spreading to contexts where *-em* would be historically used.

Some modern grammars give extremely general descriptions of this alternation: “the masculine-neuter ending *-om* appears as *-em* after ‘soft’ consonants” (Alexander 2006: 85).<sup>1</sup> Similarly, other grammars seem to suggest that the alternation is purely phonological: “Stems ending in a palatal cause vowel alternation in the instrumental singular ending, e.g. *učenikom* ‘pupil’ - *prijateljem* ‘friend’...” (Kordić 1997: 12). Yet other works argue that the distinction between *-em* and *-om* is completely predictable from whether the noun ends in a hard (*-om*) or soft consonant (*-em*) (Hammond 2005: 146). Additional phonological environments of this alternation have been noted already:

Instrumental *-et* / *-em* is normal with stems in *-c*, where vocative has *-e/-e* and the first-palatalization alternation, as *ótac/otac* ‘father’, vocative *oce/оче*. *-ot/-om* tends to be kept in foreign words and names (*Kiš-ot/Kйш-om*) and in words with *e* in the preceding syllable: *padež-ot/падеж-om* ‘case’. (Brown 1993: 320)

As can be seen the idea that analogical relations help predict this particular alternation is not really new. However, Lečić (2015) convincingly shows that for the majority of the proposed prescriptive rules of where and when to use *-em* or *-om*, exceptions can be found in a corpus. This essentially means that there is no obvious categorical rule that correctly predicts whether a noun will take *-em*, *-om* or both. Secondly, and more importantly, the author shows that overabundant nouns, even when very infrequent with one of the two forms, are acceptable for speakers, whereas non-overabundant nouns (according to the corpus) are acceptable with only one of the two forms. This strongly indicates that there really are three classes of nouns in BCS: *-om* nouns, *-em* nouns and *-om/-em* nouns.

---

<sup>1</sup>The soft consonants in Croatian are: *č, š, ž, ć, đ, dž, nj, lj, j, c*.

### 6.1.1 Modelling the system

One approach how to model overabundance with type hierarchies can be achieved by employing hybridization (Guzmán Naranjo & Bonami 2016). Hybridization assumes that there are two basic types, *exclusively-em* and *exclusively-om*. Nouns of type *exclusively-em* can only take the marker *-em* and similarly for *exclusively-om*. Nouns of a hybrid type *em~om* can take either *-em* or *-om*. This hybrid type is related to the other two types as shown in Figure 6.1.

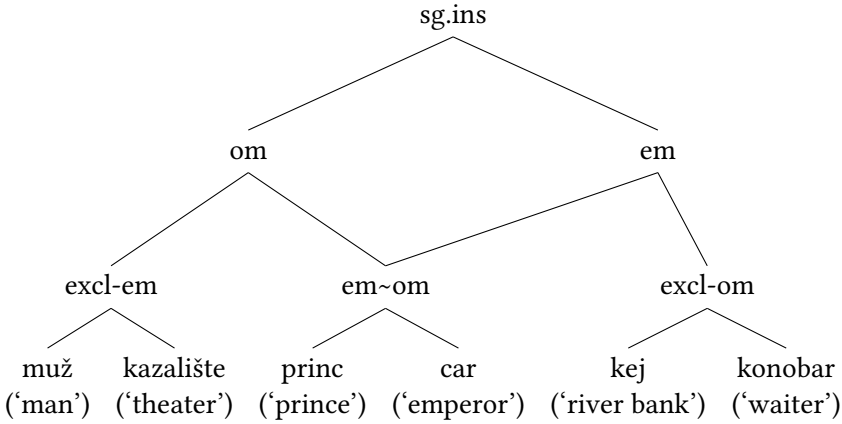


Figure 6.1: Hybridization in BCS nouns.

In the present approach there are no constraints being inherited in the hierarchy in Figure 6.1, the approach at hand simply organizes nouns according to the markers they can take in the instrumental singular. Relevant constructions or rules would then introduce the appropriate markers for each type. This can be illustrated schematically in (3):

- (3) a.  $[\text{STEM}(X^{em})\text{-em}] \leftrightarrow [\text{SEM}(X) + \text{Inst} + \text{Sg}]$   
 b.  $[\text{STEM}(X^{om})\text{-om}] \leftrightarrow [\text{SEM}(X) + \text{Inst} + \text{Sg}]$

Because lexemes like *princ* and *car* belong to types *om* and *em*, both constructions can apply to them. Other implementations are possible, of course, but the important point is that the hierarchy in Figure 6.1 expresses that nouns that can take both markers share properties with those that can take only one.

A complex issue that arises with hybrid hierarchies is what happens to the analogical filters in such cases. The analogical function for some leaf type contains all the generalizations, as well as exceptions, that determine whether any

given lexeme belongs to said type or not. In terms of the model of analogy as a type constraint, the type *em~om* inherits all analogical constraints from *em* and *om* as:  $[em\sim om\ PHON] = [em\ PHON] \wedge [om\ PHON]$ . This means that nouns *em~om* will end up looking like nouns from the classes *excl-om* and *excl-em*, because they must satisfy the same constraints.

The prediction this approach makes is that we expect the confusion between *em~om* and each of the two exclusive classes *excl-em* and *excl-om* to be relatively higher than the confusion between the two exclusive classes (*excl-em* and *excl-om*) themselves.

### 6.1.2 Materials

I extracted all 13227319 instances of instrumental singular nouns from the Web Corpus of Croatian, Bosnian and Serbian (Ljubešić & Klubička 2014) (1.9 billion tokens). From these, 6575746 are masculine nouns. After removing clear mistakes (punctuation marks, etc.), the total number of types was 227263. The number of types which appeared with either *-em* or *-om*, or both was 186443. The final data-set (after removing cases that appeared with multiple spellings) contained 180987 nouns, with 39245 nouns (22%) taking *-em*, 137290 nouns (76%) taking *-om*, and 4452 nouns (2%) taking both markers. Because we cannot know from a corpus whether a noun is **not** over abundant, there is always the risk of have many false negatives, particularly in the lower frequency cases (since it is possible that they are overabundant but there were not enough cases in the corpus for it to appear with both markers). There is a large imbalance in the type frequency of each of the three classes. I only used *-em* nouns with a frequency of more than 60, *-om* nouns with a frequency of more than 500, and *-em/-om* nouns with a frequency of more than 100 to address both problems. This process produces a dataset with 3138 *-om* nouns, 2056 *-em* nouns and 1293 *-em/-om* nouns. These numbers are somewhat arbitrary, but they produce a more balanced sample and help control for false negatives. By selecting only the more frequent nouns, there is a higher probability that the class assignment is correct.

I take the stem of the nouns to be the instrumental singular minus the *-em* or *-om* endings. I performed no orthography corrections for this data-set.

### 6.1.3 Results

The model is rather simple for this case. The predictors are the last two segments and the number of consonant clusters in the noun: `class ~ final.1 + final.2`

+ `n_cluster`<sup>2</sup>. The results of the model can be seen in Table 6.1 and the corresponding statistics in Table 6.2.

Table 6.1: Confusion Matrix for the model predicting instrumental singular in Croatian nouns.

	Reference		
Prediction	em	em~om	om
em	1887	445	25
em~om	147	502	188
om	22	346	2925

Table 6.2: Overall statistics for Confusion Matrix Table 6.1.

Overall statistics:
Accuracy : 0.8192
95% CI : (0.8096, 0.8285)
No Information Rate : 0.4873
Kappa : 0.7051

We see that the model does predict fairly well the declension class of these nouns, although it makes a large number of mistakes when predicting *-em/-om*. Most relevant here is the degree of confusion between the three classes. It can be observed that *excl-om* and *excl-em* have a small rate of confusion between them. The greater amount of confusion is between *em~om* and *excl-em*, and *em~om* and *excl-om*. This is shown more clearly in Figure 6.3. The Y axis shows the percentage of predicted classes for each class. We see here that *excl-em* is rarely predicted to be *excl-om* and viceversa. Meanwhile, *em~om* is often predicted to be *excl-em* and *excl-om*.

This distribution is exactly what the model predicts, but it also makes sense from a historical perspective. As mentioned above, it was *em* nouns which lost their distinctive  $r^j$  which started taking *om*. That is, only when a certain set of *em* nouns started phonological shapes which would fit the *om* class, did this nouns started being overabundant. We have thus a system that went from being perfectly predictable (as already mentioned in simple cases of phonological conditioned allomorphy) to overabundance.

<sup>2</sup>This model had one hidden layer with 5 nodes and a decay rate of 0.01.

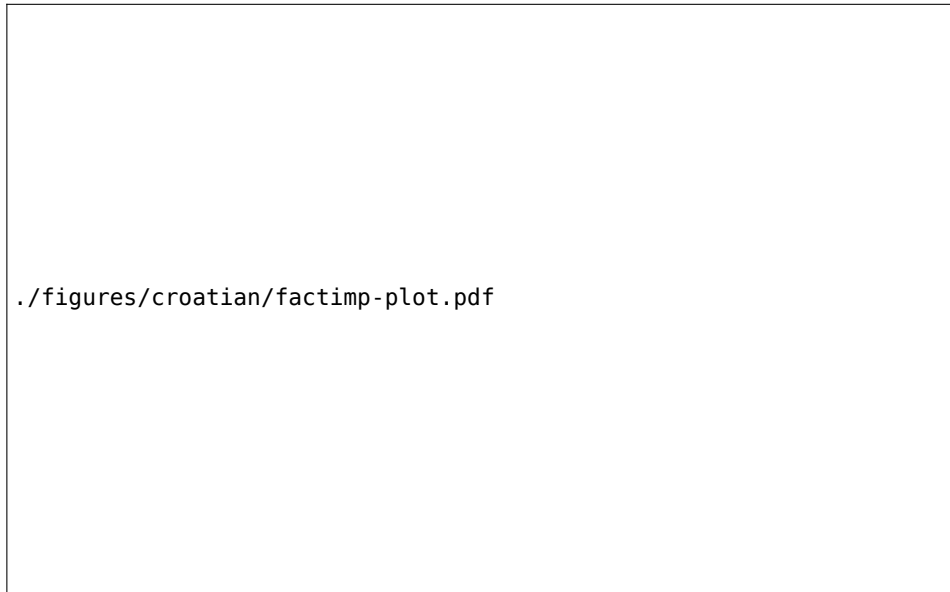


Figure 6.2: Additive (left) and subtractive (right) accuracy and kappa scores for for the model predicting Croatian instrumental.



Figure 6.3: Proportion of confusion between classes in Croatian.

## 6.2 Frequency and analogical similarity: Russian diminutives

### 6.2.1 Russian diminutives

Nouns in Russian<sup>3</sup> can form the diminutive with a wide range of different suffixes. Some examples of diminutive suffixes are shown in (4).<sup>4</sup>

- (4) -jets (-ец), -ik (-ек), -jok (-ёк), -ochik (-очек), -jechik (-ечек), -jochik (-ёчек), -itsa (-ица), -ichka (-ичка), -iko (-ико), -ko (-ко), -jetso (-есо), -tso (-цо), -tse (-це), **-ik** (-ик), **-ok** (-ок), **-chik** (-чик)

The choice of suffix is partly due to the gender of the noun (Kempe & Brooks 2001; Kempe et al. 2003), as described in (5), but not completely. There are several possible diminutive forms for each gender, and the explanation why some nouns chose one or the other is not completely clear. As the whole system is too complex to be addressed here, I will focus on masculine nouns that build the diminutive with **-ik**, **-ok**, or **-chik** exclusively.

- (5) a. -iko, -ko, -tso, -tse → neuter nouns  
 b. -itsa, -ichka → feminine nouns  
 c. -ik, -ok, -chik → masculine nouns

In the masculine subset: **-ik**, **-ok**, **-chik**, we find a particularly complex affix competition problem. Example (6) illustrates some nouns which can only appear with one of the three forms, while (7) shows the nouns which occur with two of the three different markers.

- (6) a. -ik, \*-chik, \*-ok: stol, kot, miač ('table', 'cat', 'ball')  
 b. \*-ik, -chik, \*-ok: zabor ('fence')  
 c. \*-ik, \*-chik, -ok: molot, vjehjer ('hammer', 'vening')
- (7) a. -ik, -chik, \*-ok: stul, shkaf ('chair', 'cabinet')  
 b. \*-ik, -chik, -ok: rukav ('sleeve')  
 c. -ik, \*-chik, -ok: rot, list, chas ('mouth', 'leaf', 'hour')  
 d. -ik, -chik, -ok: ?

<sup>3</sup>A previous version of the study in this section was presented in Olinco 2016 (Guzmán Naranjo & Pyatigorskaya 2016).

<sup>4</sup>For clarity I will use the Latin transliterations in the examples, but the models were trained using their Cyrillic forms.



A similar situation can be found in German, where there is competition between the diminutive forms *-chen* and *-lein*, with some degree of overlap between the two: *Häuschen~Häuslein* ('small house'). Similarly, Spanish has the forms *-illo*, *-ito*, *-cito*, *-ico*, among others, and some overlap between these forms in a substantial set of nouns: *casita* ('small house'), *pollito* ('chick'), *gatito~gatico* ('kitten').

For Russian, there seems to be no rule-based account of which nouns can take which markers (Gouskova et al. 2015). Some research on Russian diminutives has focused on the relation between the different forms and gender, as well as gender acquisition (Kempe et al. 2010; Protassova & Voeykova 2007; Voeykova 1998), but relatively little attention has been given to the actual conditions that help decide between the different forms. Gouskova et al. (2015) is the most recent approach to this problem. The authors propose what they label sublexical phonotactics, a model that is very similar to an analogical model. The basic idea is that each form a speaker encounters is stored in a sublexicon specific to that form, i.e. *-ik* forms are stored in an *ik* sublexicon, *-chik* forms are stored in an *chik* sublexicon and so on. Speakers find phonotactic regularities in each sublexicon, and new items are coined based on those regularities. Conceptually, there are only a few minor differences between traditional analogical models and Gouskova et al. (2015)'s, but in terms of implementation some issues with the latter exist. From a theoretical perspective, Gouskova et al. (2015) propose a flat model, where speakers simply have lists for each type, and there is no structuring of said types. This is quite common in analogical approaches.

Because the issues with Gouskova et al. (2015)'s approach are of secondary concern, I will only discuss them briefly. The essence of their implementation is as follows. Using the UCLA phonotactic learner (Hayes & Wilson 2008), a first instance of their model infers phonotactic regularities in a dataset of Russian nouns marked for their diminutive preference (the phonotactic regularities are inferred from the bases of the nouns). In a second instance, a mixed effects model is trained on the phonotactic generalizations to determine which are statistically significant.

There are several problems with this method. Of some real concern is that there is no cross-validation. Their results stem from testing the model on the same dataset it was trained on; however, this could be solved. It is also somewhat unclear what the purpose of the mixed effect model is, because the UCLA learner is already predicting classes based on the phonotactic patterns. Finding statistically significant patterns is problematic because these patterns are highly correlated with each other (because of phonotactics), and mixed effects models

are not robust against co-linearity, which means any statistical significance is in question.<sup>5</sup>

The results in Table 6.3 can be obtained by looking at the predictions made by the UCLA learner<sup>6</sup> vs observed diminutives in the dataset. The corresponding statistics are in Table 6.4. It is clear that the UCLA model is not learning to properly discriminate the diminutive classes, and the model performs at chance level.

Table 6.3: Actual results for Gouskova et al. (2015)’s model.

Reference	Predicted		
	chik	ik	ok
chik	135	54	76
ik	133	54	65
ok	122	56	75

Table 6.4: Overall statistics for Confusion Matrix Table 6.3.

Overall statistics:
Accuracy : 0.3429
95% CI : (0.3093, 0.3776)
No Information Rate : 0.5065
Kappa : 0.01

In one of the tests, the authors try to evaluate their model on a wug experiment with Russian speakers. They designed 300 nonce words and asked speakers to produce the corresponding diminutives. The authors claim high correlation between their model and speaker choices but it is not clear how this correlation was measured. First, their evaluation is made by using Kendall’s and Spearman’s correlation coefficients, which would lead to believe that they are testing predicted proportions vs produced, but this is not made clear. In the code the correlation

<sup>5</sup>I could not reproduce their results in order to check because the versions of the statistical software used by the authors are no longer supported.

<sup>6</sup>The original dataset and code used by Gouskova et al. (2015) were kindly provided to me by the lead author. I use the results from their code.

calculations are made on categorical variables, which is not advisable and makes the results hard to interpret.

Despite these potential issues, the basic idea that the affix competition in Russian diminutives is resolved analogically is on the right track. In the end, it is not of too much interest whether the phonotactics approach could outperform the neural networks I employ here, or the other way around. The important question that is left to be addressed is whether a flat list approach like that of Gouskova et al. (2015)'s is more appropriate than a structured model.

### 6.2.2 Modelling the system

Conceptually, if one ignores semantics and stress assignment (which also seem to have no straightforward solution according to Gouskova et al. 2015), it is possible to capture the system with a cross-classification approach similar to the Croatian system.

The hierarchy in Figure 6.4 shows a simple sketch of how the system can be captured. Figure 6.4 shows that all pairwise combinations are possible.<sup>7</sup>

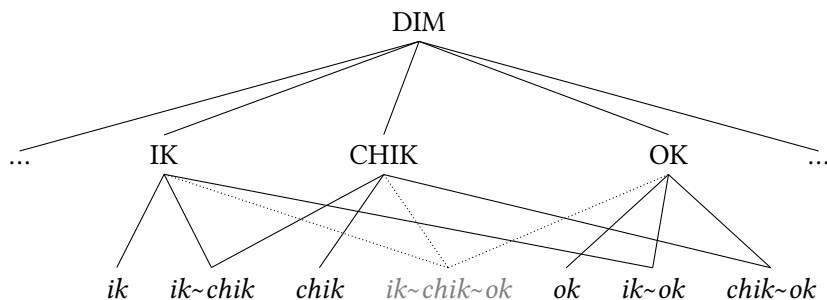


Figure 6.4: Hybridization in Russian nouns.

Since all combinations are possible, we expect confusion between all three types. However, the frequency at which the combinations occur is not uniform. Figure 6.5 shows that the most frequent classes are the non alternating classes and the *ik~ok* class (see the next section for details on the dataset used). This case is thus interesting because it shows the effects of type frequency. If type frequency plays no role, then we would expect that the confusion between *ik*, *ok* and *chik* should be more or less equal. If, on the other hand, type frequency does play a role, we expect confusion between all classes, with the highest confusion

<sup>7</sup>I am not aware of cases where all three suffixes are possible with one noun and could not elicit any from my informants.

between *ik*, *ok* and *ik~ok*, while the lowest confusion should be between *chik* and *ik~ok* (since this combination was not attested).

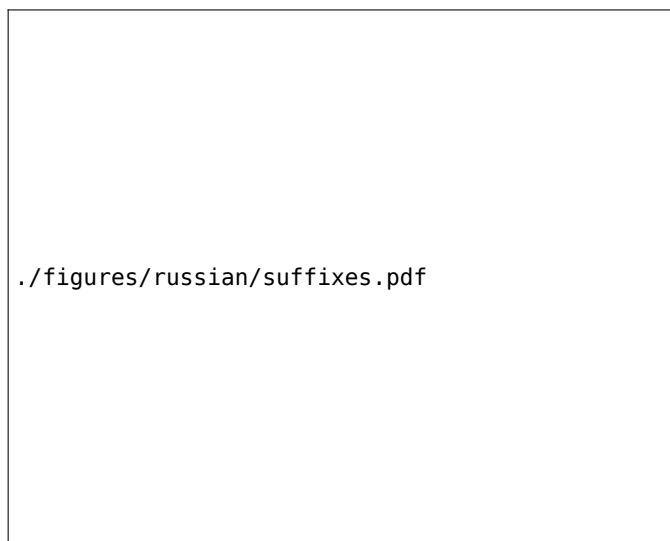


Figure 6.5: Type frequency of Russian suffix classes.

### 6.2.3 The dataset

I used the diminutive dataset collected by Gouskova et al. (2015), but hand checked them with a native speaker of Russian.<sup>8</sup> The original dataset was extracted from the google-ngram corpus (Michel et al. 2011) and contains 1367 forms. Since there are no Russian taggers which can identify diminutives, the authors relied exclusively on the endings of the words to find diminutive forms. This caused the dataset to have many problematic cases. To solve this, we removed errors (perceived to be ungrammatical by a few informants), non-diminutives (e.g. the word *alkogol-ik* ‘alcoholic’ has an *-ik* suffix but is not actually a diminutive) and non-words. This left us with 821 diminutives. My informant provided stress marks for the bases of the selected diminutives.

Because there are not enough cases for the classes *chik~ok* (f=7) and *chik~ik* (f=8), I removed them from the data set for fitting the models.<sup>9</sup> The final dataset had a total of 811 nouns.

<sup>8</sup>This part of the work would not have been possible without the invaluable help of Elena Pyatigorskaya, who manually checked and corrected the whole-data set.

<sup>9</sup>These two classes were not predictable at all if included. It is hard to tell whether a model with more examples would perform better in these two cases.

6.2.4 Results

To predict the diminutive forms, I fitted a model using the formula: `diminutive ~ final.1 + final.2 + length_letters * n_vowels + stress_position * stressed_vowel`.

Basically, this model looks at the final two segments of the base nouns (in the nominative), the interaction between the length of the base and the number of vowels of the base, the position of the stress in the word (counting from the right) and the stressed vowel. The results can be seen in Table 6.5 and the corresponding statistics in Table 6.6. The relative importance of the predictors is shown in Figure 6.6.

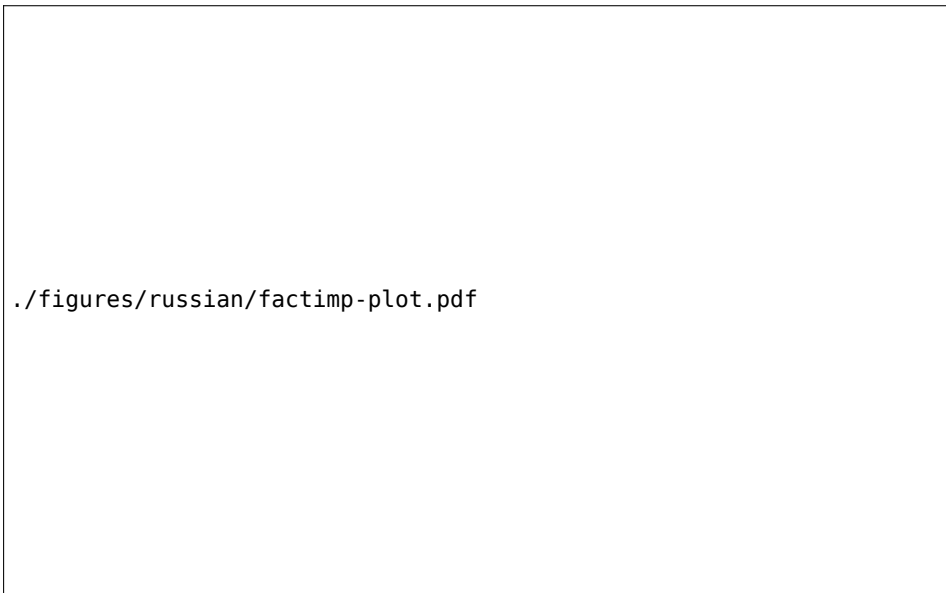
Table 6.5: Confusion Matrix for the model predicting Russian diminutives.

Reference				
Prediction	chik	ik	ik~ok	ok
chik	177	11	1	13
ik	8	175	28	33
ik~ok	1	22	34	18
ok	13	30	15	232

Table 6.6: Overall statistics for Confusion Matrix Table 6.5.

Overall statistics:
Accuracy : 0.762
95% CI : (0.7312, 0.791)
No Information Rate : 0.365
Kappa : 0.6654

First of all, the model is very accurate overall and well above random chance. All four classes can be distinguished to a certain degree. The most important factor is the last segment, but the other factors all seem to have an important effect. More importantly, we find the exact predicted result in the error distribution. The class *ik~ok* is less confused with the class *chik* than with any of the other classes. Class *chik* is rarely confused with classes *ik* and *ok*, while these two are



./figures/russian/factimp-plot.pdf

Figure 6.6: Additive (left) and sbustractive (right) accuracy and kappa scores for the model predicting Russian diminutives.

confused with each other with relatively high frequency. This is clearly shown in Figure 6.7.

There are two ways how to interpret these findings. Firstly, in one scenario, it could be postulated that there is a need for quantitative information to be hard coded into the hierarchy, i.e. we should assign stronger connections to *IK* and *OK*, than to the other combinations. I propose, however, that this addresses the problem backwards. The more straightforward alternative is to see the higher type frequency of *ik~ok* as a byproduct of the analogical system itself, and not as something one has to directly integrate into the model. The fact that we have more *ik~ok* nouns than *chik~ik* or *chik~ok* nouns is due to the constraints for *IK* and *OK* being more compatible with each other, and producing a more relaxed set of constraints than *CHIK~IK* or *CHIK~OK*.

## 6.3 Interim conclusion

I have shown how the hybridization model can properly predict the distribution of both partially overlapping cases in Croatian, and overlapping diminutives with different type frequencies in Russian. These two examples show that the

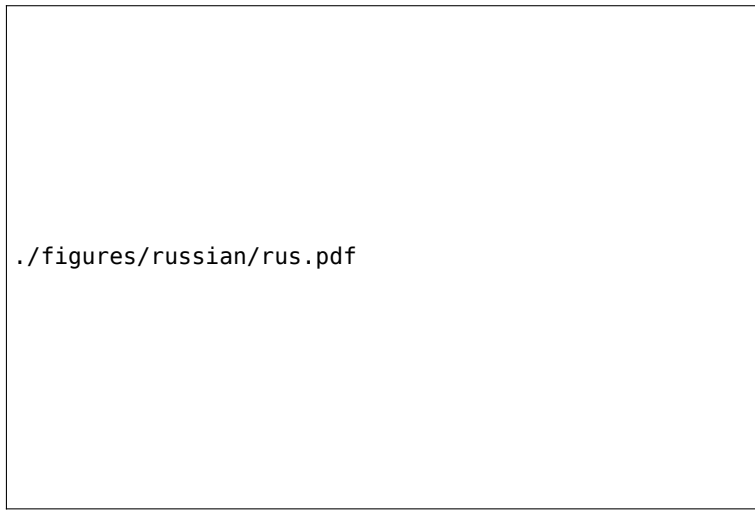


Figure 6.7: Proportion of confusion between classes in Russian.

predicted effects are not only present in simple trees, but can be also observed in more complex hierarchical structures. These results clearly reject the flat list approach and support a structured organization of these systems. It is also interesting to see that, despite the fact that one case is inflectional morphology and the other derivational morphology, the results for both studies are very similar in terms of the distribution of errors and the analogical relations. This result argues for an organization of the stems in classes independent of the type of morphological process, at least for morphological theories that make a distinction between inflection and derivation. So, even if overabundance in inflection and affix competition in derivation are treated as different kind of phenomena, the underlying structures would be equivalent.

# References

- Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89(3). 429–464.
- Ackerman, Farrell & Robert Malouf. 2016. Word and pattern morphology: An information-theoretic approach. *Word Structure* 9(2). 125–131.
- Afonso, Olivia, Alberto Domínguez, Carlos J. Álvarez & David Morales. 2014. Sublexical and lexico-syntactic factors in gender access in Spanish. *Journal of Psycholinguistic Research* 43(1). 13–25.
- Aguirre, Carmen & Wolfgang U. Dressler. 2008. On Spanish verb inflection. *Folia Linguistica* 40. 75–96.
- Alber, Birgit. 2009. Past participles in Mocheno: Allomorphy and alignment. In Michael T. Putnam (ed.), *Studies on German-language islands*, 33–64. Amsterdam, Philadelphia: Benjamins.
- Albright, Adam. 2008a. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Linguistic universals and language change*, 144–184. Oxford: Oxford University Press.
- Albright, Adam. 2008b. How many grammars am I holding up? Discovering phonological differences between word classes. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 1–20. Somerville, MA: Cascadia Proceedings Project.
- Albright, Adam. 2009. Modeling analogy as probabilistic grammar. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*, 200–228. Oxford, New York: Oxford University Press.
- Albright, Adam, Argelia Andrade & Bruce Hayes. 2001. Segmental environments of Spanish diphthongization. *UCLA Working Papers in Linguistics* 7. 117–151.
- Albright, Adam & Bruce Hayes. 1999. *An automated learner for phonology and morphology*. [https://www.researchgate.net/profile/Bruce\\_Hayes/publication/2876878\\_An\\_Automated\\_Learner\\_for\\_Phonology\\_and\\_Morphology/links/02bfe51194b1fab53a000000.pdf](https://www.researchgate.net/profile/Bruce_Hayes/publication/2876878_An_Automated_Learner_for_Phonology_and_Morphology/links/02bfe51194b1fab53a000000.pdf), accessed 2016-9-10.
- Albright, Adam & Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. In *Proceedings of the ACL-02 workshop on morpho-*



- logical and phonological learning*, vol. 6, 58–69. Association for Computational Linguistics.
- Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90(2). 119–161.
- Alexander, Ronelle. 2006. *Bosnian, Croatian, Serbian, a grammar: With sociolinguistic commentary*. Madison: University of Wisconsin Press.
- Anderson, Stephen R. 2008. Phonologically conditioned allomorphy in the morphology of Surmiran (Rumantsch). *Word Structure* 1(2). 109–134.
- Anderson, Stephen R. 2015. Morphological change. In Richard D. Janda & Brian D. Joseph (eds.), *The Routledge handbook of historical linguistics*, 264–285. Oxford: Blackwell Publishing.
- Anttila, Raimo. 1977. *Analogy*. Berlin: Mouton de Gruyter.
- Anttila, Raimo. 2003. Analogy: The warp and woof of cognition. In Richard D. Janda & Brian D. Joseph (eds.), *The Routledge handbook of historical linguistics*, 425–440. Malden, Mass.: Blackwell Publishing.
- Arndt-Lappe, Sabine. 2011. Towards an exemplar-based model of stress in English noun–noun compounds. *Journal of Linguistics* 47. 549–585.
- Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: The case of English *-ity* and *-ness*. *English Language and Linguistics* 18(3). 497–548.
- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Massachusetts: MIT Press.
- Arppe, Antti, Peter Hendrix, Petar Milin, R. Harald Baayen & Cyrus Shaoul. 2014. Ndl: Naive discriminative learning. *R package versions 0.1*.
- Arregi, Karlos. 2000. How the Spanish verb works. In *30th linguistic symposium on romance languages*. Gainesville: University of Florida. <http://home.uchicago.edu/karlos/Arregi-2000-how.pdf>.
- Awedoba, Albert K. 1980. Borrowed nouns in Kasem nominal classes. *Anthropological Linguistics* 22(6). 248–263.
- Awedoba, Albert K. 1996. Kasem nominal genders and names. *Research Review* 12(2). 8–24.
- Awedoba, Albert K. 2003. Criteria for noun classification in Kasem. In Manfred von Roncador, Kerstin Winkelmann & Ulrich Kleinewillinghöfer (eds.), *Cahiers Voltaïques / Gur Papers*, vol. 6, 3–15.
- Baayen, R. Harald. 2007. Storage and computation in the mental lexicon. In Gonia Jarema & Gary Libben (eds.), *Mental lexicon: Core perspectives*, 81–104. Amsterdam: Elsevier.
- Baayen, R. Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5(3). 436–461.

- Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Revista Brasileira de Linguística Aplicada* 12(2). 295–328.
- Baayen, R. Harald & Peter Hendrix. 2011. Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. In *Empirically examining parsimony and redundancy in usage-based models, lsa workshop*.
- Baayen, R. Harald, Petar Milin, Dusica Filipović Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438–81.
- Baerman, Matthew. 2007. Morphological reversals. *Journal of Linguistics* 43. 33–61.
- Baptista, Barbara O. & Jair L.A. Silva Filho. 2006. The influence of voicing and sonority relationships on the production of English final consonants. In Barbara O. Baptista & Michal Alan Watkins (eds.), *English with a Latin beat: Studies in Portuguese/Spanish-English interphonology*, 73–90. Amsterdam: John Benjamins.
- Bargery, George Percy & Diedrich Westermann. 1951. *A Hausa-English dictionary and English-Hausa vocabulary*. Oxford: Oxford University Press.
- Bateman, Nicoleta & Maria Polinsky. 2010. Romanian as a two-gender language. In Donna B. Gerdts, John C. Moore & Maria Polinsky (eds.), *Hypothesis A/hypothesis B: Linguistic explorations in honor of David M. Perlmutter* (Current Studies in Linguistics), 41–77. Cambridge, Massachusetts: MIT Press.
- Bauer, Laurie. 2003. *Introducing linguistic morphology*. Washington, D. C.: Georgetown University Press.
- Bechtel, William & Adele Abrahamsen. 2002. *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks*. Malden, Mass.: Blackwell Publishing.
- Becker, Thomas. 1990. *Analogie und morphologische Theorie* (Studien zur Theoretischen Linguistik 11). München: Fink.
- Becker, Thomas. 1993. Back-formation, cross-formation, and bracketing paradoxes in paradigmatic morphology. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 1–26. Dordrecht: Springer.
- Bellido, Paloma García. 1986. *Lexical diphthongization and high-mid alternations in Spanish: An autosegmental account*. Seattle, WA: Linguistic Analysis.
- Beniamine, Sacha. 2017. A computational approach to the abstraction of morphophonological alternations. In *Typologie et modélisation des systèmes morphologiques*. Paris. <http://www.llf.cnrs.fr/fr/node/5611>.

## References

- Beniamine, Sacha & Olivier Bonami. 2016. Generalizing patterns in Instrumented Item-and-Pattern Morphology. In *Structural complexity in natural language(s) (SCNL)*. Paris.
- Bergen, Benjamin & Nancy Chang. 2005. Embodied construction grammar in simulation-based language understanding. In Jan-Ola Östman & Mirjam Fried (eds.), *Construction grammars: Cognitive grounding and theoretical extensions*, 147–190. Amsterdam, Philadelphia: John Benjamins.
- Bermúdez-Otero, Ricardo. 2013. The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. *International Journal of Latin and Romance Linguistics* 25(1). 3–103.
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.
- Blevins, James P. 2008. Declension classes in Estonian. *Linguistica Uralica* 4. 241–267.
- Blevins, James P. 2013. The information-theoretic turn. *Psihologija* 46(3). 355–375.
- Blevins, James P. 2016. *Word and paradigm morphology*. Oxford, New York: Oxford University Press.
- Blevins, James P., Petar Milin & Michael Ramscar. 2016. The Zipfian paradigm cell filling problem. In Ferenc Kiefer, James Blevins & Huba Bartos (eds.), *Perspectives on morphological organization: Data and analyses*, 141–158.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Reinhart & Winston.
- Bodomo, Adams. 1994. The noun class system of Dagaare: A phonology-morphology interface. In *Working Papers in Linguistics, Norwegian University for Science and Technology*.
- Bodomo, Adams. 1997. *The structure of Dagaare*. Stanford: CSLI publications.
- Boersma, Paul P. G. 1997. How we learn variation, optionality, and probability. In *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 43–58.
- Boersma, Paul P. G. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Den Haag: Holland Academic Graphics/IFOTT.
- Boersma, Paul P. G. & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1). 45–86.
- Boloh, Yves & Laure Ibernou. 2010. Gender attribution and gender agreement in 4- to 10-year-old French children. *Cognitive Development* 25(1). 1–25.
- Bonami, Olivier & Sacha Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

- Bonami, Olivier & Gilles Boyé. 2003. Supplétion et classes flexionnelles. *Langages* 37(152). 102–126.
- Bonami, Olivier & Gilles Boyé. 2006. Deriving inflectional irregularity. In *Proceedings of the 13th International Conference on HPSG*, 39–59. Varna: CSLI Publications.
- Booij, Geert E. 1998. Phonological output constraints in morphology. In Wolfgang Kehrein & Richard Wiese (eds.), *Phonology and morphology of the Germanic languages*, 143–163. Tübingen: Niemeyer.
- Booij, Geert E. 2010. *Construction morphology*. Oxford, New York: Oxford University Press.
- Borg, Ingwer & Patrick J. F. Groenen. 2005. *Modern multidimensional scaling: Theory and applications*. New York: Springer.
- Boyé, Gilles & Patricia Cabredo Hofherr. 2010. The distribution of prethematic vowels in Spanish verbs. <http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=boye&subURL=BoCa-Probus-PrethematicVowelsSpanish.pdf>, accessed 2010-5-10.
- Boyé, Gilles & Patricia Cabredo Hofherr. 2004. Étude de la distribution des suffixes-er/-ir dans les infinitifs de l'espagnol à partir d'un corpus exhaustif. *Corpus* (3). 237–260.
- Boyé, Gilles & Patricia Cabredo Hofherr. 2006. The structure of allomorphy in Spanish verbal inflection. *Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset* 13. 9–24.
- Brame, Michael K. & Ivonne Bordelois. 1973. Vocalic alternations in Spanish. *Linguistic Inquiry* 4(2). 111–168.
- Braune, Wilhelm. 1895. *Gotische Grammatik: Mit einigen Lesestücken und Wortverzeichnis*. Halle (Saale): Niemeyer.
- Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-functional syntax*. Chichester: Wiley-Blackwell.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94.
- Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118. 245–259.
- Brindle, Jonathan Allen. 2009. On the identification of noun class and gender systems in Chakali. In *Proceedings of the 38th annual conference on african linguistics*, 84–94. Somerville, MA: Cascadilla Proceedings Project.

## References

- Brovetto, Claudia & Michael T. Ullman. 2005. The mental representation and processing of Spanish verbal morphology. In *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, 98–105. Somerville, MA: Cascadilla Proceedings Project.
- Brown, Dunstan & Andrew Hippisley. 2012. *Network morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press.
- Brown, Wayles. 1993. Serbo-Croat. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, 306–387. London, New York: Routledge.
- Butterworth, Brian. 1983. Lexical representation. In Brian Butterworth (ed.), *Language production: development, writing, and other language processes*, 257–294. London: Academic Press.
- Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425–455.
- Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan L. & Clayton Beckner. 2015. Language use, cognitive processes and linguistic change. In Claire Bower & Bethwyn Evans (eds.), *The Routledge handbook of historical linguistics*, 503–518. London: Routledge.
- Bybee, Joan L. & Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58(2). 265–289.
- Caffarra, Sendy & Horacio A. Barber. 2015. Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research* 1605. 83–92.
- Caffarra, Sendy, Anna Siyanova-Chanturia, Francesca Pesciarelli, Francesco Vespignani & Cristina Cacciari. 2015. Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology* 52(8). 1019–1030.
- Callow, John C. 1965. Kasem nominals: A study in analyses. *Journal of West African Languages* 2(1). 29–36.
- Carreira, Maria. 1991. The alternating diphthongs of Spanish: A paradox revisited. In *Current studies in Spanish linguistics*, 407–445. Washington, D. C.: Georgetown University Press.
- Carstairs, Andrew. 1990. Phonologically conditioned suppletion. *Contemporary Morphology*. 17–23.
- Carstairs, Andrew. 1998. Some implications of phonologically conditioned suppletion. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1998*, 67–94. Dordrecht: Springer.
- Casali, Roderic F. 2008. ATR harmony in African languages. *Language and Linguistics Compass* 2(3). 496–549.

- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
- Churchland, Paul M. 1989. *A neurocomputational perspective: The nature of mind and the structure of science*. Massachusetts: MIT press.
- Clahsen, Harald, Fraibet Aveledo & Iggy Roca. 2002. The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language* 29. 591–622.
- Clopper, C. J. & Egon S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26(4). 404–413.
- Cojocaru, Dana. 2003. *Romanian grammar*. Durham: The Slavic and East European Language Resource Center.
- Contini-Morava, Ellen. 1994. *Noun classification in Swahili*. <http://www2.iath.virginia.edu/swahili/swahili.html>, accessed 2016-11-9.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. & Norman M. Fraser. 1993. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29. 113–142.
- Costanzo, Angelo Roth. 2011. *Romance conjugational classes: Learning from the peripheries*. Columbus, OH: The Ohio State University dissertation.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William & Alan D. Cruse. 2004. *Cognitive linguistics*. Cambridge, MA: Cambridge University Press.
- Cucerzan, Silviu & David Yarowsky. 2003. Minimally supervised induction of grammatical gender. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, 40–47. Association for Computational Linguistics.
- Cuervo, Rufino José & Ignacio Ahumada. N.d. *Notas a la Gramática de la lengua castellana de don Andrés Bello*. Bogotá: Instituto Caro y Cuervo.
- Cysouw, Michael. 2007. New approaches to cluster analysis of typological indices. In Peter Grzybek & Reinhard Köhler (eds.), *Exact methods in the study of language and text*, 61–76. Berlin, Boston: De Gruyter Mouton.
- Czaplicki, Bartłomiej. 2013. Arbitrariness in grammar: Palatalization effects in Polish. *Lingua* 123. 31–57.
- Dakubu, Mary Esther Kropp. 1997. Oti-Volta vowel harmony and Dagbani. *Gur Papers* 2. 81–88.
- De Smet, Hendrik & Olga Fischer. 2017. The role of analogy in language change: Supporting constructions. In Marianne Hundt, Sandra Mollin & Simone E.

## References

- Pfenninger (eds.), *The changing English language*. Cambridge, New York: Cambridge University Press.
- De Vaan, Laura, Robert Schreuder & R. Harald Baayen. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2(1). 1–24.
- de Haas, Wim G. 1987. An autosegmental approach to vowel coalescence. *Lingua* 73(3). 167–199.
- de Haas, Wim G. 1988. *A formal theory of vowel coalescence: A case study of Ancient Greek*. Berlin: Walter de Gruyter.
- DeMello, George. 1993. -Ra vs. -se subjunctive: A new look at an old topic. *Hispania* 76(2). 235–244.
- Derwing, Bruce L. & Royal Skousen. 1994. Productivity and the English past tense. In Susan D. Lima, Roberta Corrigan & Gregory K. Iverson (eds.), *The reality of linguistic rules*, vol. 26, 193–218. Amsterdam, Philadelphia: John Benjamins.
- Di Sciullo, Anna-Maria & Edwin Williams. 1987. *On the definition of word*. Cambridge, Massachusetts: Springer.
- Dinu, Liviu P., Vlad Niculae & Octavia-Maria Sulea. 2012. Dealing with the grey sheep of the Romanian gender system, the neuter. In *COLING (Demos)*, 119–124.
- Echegoyen, Artemisa & Katherine Voigtlander. 1979. *Luces contemporáneas del otomí: Gramática del otomí de la sierra*. Mexico, D. F.: Instituto Lingüístico de Verano. <https://www.sil.org/resources/archives/2018>, accessed 2017-2-21.
- Echegoyen, Artemisa & Katherine Voigtlander. 2007. *Diccionario yuhú: Otomí de la Sierra Madre Oriental: Estados de Hidalgo, Puebla y Veracruz, México*. Estados de Hidalgo, Puebla y Veracruz, México: Instituto Lingüístico de Verano.
- Eddington, David. 1996. Diphthongization in Spanish derivational morphology: An empirical investigation. *Hispanic Linguistics* 8(1). 1–13.
- Eddington, David. 2000. Analogy and the dual-route model of morphology. *Lingua* 110(4). 281–298.
- Eddington, David. 2002. Spanish gender assignment in an analogical framework. *Journal of Quantitative Linguistics* 9(1). 49–75.
- Eddington, David. 2004. Issues in modeling language processing analogically. *Lingua* 114(7). 849–871.
- Eddington, David. 2009. Linguistic processing is exemplar-based. *Studies in Hispanic and Lusophone Linguistics* 2(2).
- Eddington, David & Jordan Lachler. 2006. A computational analysis of Navajo verb stems. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional research*, 143–161. CSLI Publications.

- Erelt, Mati, Tiiu Erelt & Kristiina Ross. 1997. *Eesti keele käsiraamat*. Tallinn: Eesti keele sihtasutus.
- Erelt, Mati, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael & Silvi Vare. 1995. *Eesti keele grammatika: Morfoloogia*. Tallinn: Eesti Teaduste Akadeemia Eesti Keele Instituut.
- Erelt, Tiiu, Tiina Leemets, Sirje Mäearu & Maire Raadik. 2001. *Eesti keele sõnaraamat*: ÕS. Tallinn: Eesti Keele Sihtasutus.
- Farkas, Donka F. 1990. Two cases of underspecification in morphology. *Linguistic Inquiry* 21(4). 539–550.
- Farkas, Donka F. & Draga Zec. 1995. Agreement and pronominal reference. In Guglielmo Cinque & Giuliana Giusti (eds.), *Advances in Romanian linguistics*, vol. 10, 83–101. Amsterdam, Philadelphia: John Benjamins.
- Federici, Stefano, Vito Pirrelli & François Yvon. 1995. A dynamic approach to paradigm-driven analogy. In *International Joint Conference on Artificial Intelligence*, 385–398.
- Feist, Timothy & Enrique L. Palancar. 2015. *Oto-Manguean inflectional class database*.
- Fertig, David L. 2013. *Analogy and morphological change*. Edinburgh: Edinburgh University Press.
- Fillmore, Charles J. & Paul Kay. 1995. *A Construction Grammar coursebook*. Berkeley: Unpublished ms, University of California.
- Foley, James Addison. 1965. *Spanish morphology*. Massachusetts: Massachusetts Institute of Technology dissertation.
- Fondow, Steven Richard. 2010. *Spanish velar insertion and analogy: A usage-based diachronic analysis*. Columbus, Ohio: Columbus, OH dissertation. [https://etd.ohiolink.edu/rws\\_etd/document/get/osu1290438177/inline](https://etd.ohiolink.edu/rws_etd/document/get/osu1290438177/inline), accessed 2017-6-3.
- Francis, Elaine J. & Laura A. Michaelis. 2014. Why move? How weight and discourse factors combine to predict relative clause extraposition in English. In Brian MacWhinney, Edith A. Moravcsik & Andrej L. Malchukov (eds.), *Competing motivations*, 70–87. Oxford, New York: Oxford University Press.
- Galván Torres, Adriana Rosalina. 2007. *Die Entwicklung der spanischen Diphthongierung anhand der Natürlichkeitstheorie*. Norderstedt: GRIN Verlag.
- Gerdts, Donna B., John C. Moore & Maria Polinsky (eds.). 2010. *Hypothesis A/hypothesis B: Linguistic explorations in honor of David M. Perlmutter* (Current Studies in Linguistics). Cambridge, Massachusetts: MIT Press.
- Ginzburg, Jonathan & Ivan A. Sag. 2000. *Interrogative investigations*. Stanford: CSLI publications.



## References

- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Univ. of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work*. Oxford: Oxford University Press.
- Goldsmith, John A. 2009. Morphological analogy: Only a beginning. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*, 138–164. Oxford, New York: Oxford University Press.
- Goldsmith, John A., Jason Riggle & C. L. Yu Alan. 2011. *The handbook of phonological theory*. Chichester: John Wiley & Sons.
- Gönczöl, Ramona. 2007. *Romanian: An essential grammar*. New York: Routledge.
- Gouskova, Maria, Luiza Newlin-Łukowicz & Sofya Kasyanenko. 2015. Selectional restrictions as phonotactics over sublexicons. *Lingua* 167. 41–81.
- Guzmán Naranjo, Matías & Olivier Bonami. 2016. *Overabundance as hybrid inflection: Quantitative evidence from Czech*. Grammar and Corpora. IDS. Mannheim.
- Guzmán Naranjo, Matías & Elena Pyatigorskaya. 2016. *Comparing naive discriminative learning, sublexicon phonotactics, and analogical learning*. Olinco. Olomouc.
- Hahn, Ulrike & Nick Chater. 1998. Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition* 65(2). 197–230.
- Hahn, Ulrike & Ramin Charles Nakisa. 2000. German inflection: Single route or dual route? *Cognitive Psychology* 41(4). 313–360.
- Hall, Robert A. 1965. The “neuter” in Romance: A pseudo-problem. *Word* 21(3). 421–427.
- Halle, Morris. 1978. Further thoughts on Kasem nominals. *Linguistic Analysis Seattle* 4(2). 167–185.
- Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Locke Hale & Samuel Jay Keyser (eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*, 111–176. Cambridge, MA: MIT Press.
- Hammond, Lila. 2005. *Serbian: An essential grammar*. New York: Routledge.
- Harris, James W. 1969. *Spanish phonology*. Cambridge, MA: MIT Press.
- Harris, James W. 1978. Two theories of non-automatic morphophonological alternations: Evidence from Spanish. *Language* 54(1). 41–60.
- Harris, James W. 1985. Spanish diphthongisation and stress: A paradox resolved. *Phonology* 2. 31–45.
- Harris, James W. 1987. Disagreement rules, referral rules, and the Spanish feminine article *el*. *Journal of Linguistics* 23. 177–183.

- Harris, James W. 1991. The exponence of gender in Spanish. *Linguistic Inquiry* 22(1). 27–62.
- Hay, Jennifer & Joan Bresnan. 2006. Spoken syntax: The phonetics of *giving a hand* in New Zealand English. *The Linguistic Review* 23(3). 321–349.
- Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry* 39(3). 379–440.
- Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Amsterdam, Philadelphia: Walter de Gruyter.
- Hock, Hans Henrich. 2003. Analogical change. In Richard D. Janda & Brian D. Joseph (eds.), *The handbook of historical linguistics*, 441–460. Malden, Mass.: Blackwell Publishing.
- Holmes, Virginia M. & B. Dejean de la Bâtie. 1999. Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics* 20. 479–506.
- Holmes, Virginia M. & Juan Segui. 2004. Sublexical and lexical influences on gender assignment in French. *Journal of Psycholinguistic Research* 33(6). 425–457.
- Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
- Itkonen, Esa. 2005. *Analogy as structure and process*. Amsterdam, Philadelphia: John Benjamins.
- Kapatsinski, Vsevolod. 2010. What is it I am writing? Lexical frequency effects in spelling Russian prefixes: Uncertainty and competition in an apparently regular system. *Corpus Linguistics and Linguistic Theory* 6(2). 157–215.
- Kapatsinski, Vsevolod. 2012. What statistics do learners track? Rules, constraints and schemas in (artificial) grammar learning. In Stefan Th. Gries & Dagmar Divjak (eds.), *Frequency effects in language learning and processing*, 53–82. Berlin, Boston: De Gruyter Mouton.
- Kapatsinski, Vsevolod. 2014. What is grammar like? A usage-based constructionist perspective. *Linguistic Issues in Language Technology* 11(1). 1–41.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell & Annie Zaenen (eds.), *Formal issues in lexical-functional grammar*, 29–130. Stanford: CSLI Publications.
- Kempas, Ilpo. 2011. Sobre la variación en el marco de la libre elección entre cantar y cantase en el español peninsular. *Moenia* (17). 243–264.
- Kempe, Vera & Patricia J. Brooks. 2001. The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? *Language Learning* 51(2). 221–256.

## References

- Kempe, Vera, Patricia J. Brooks & Anatolij Kharkhurin. 2010. Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning* 60(1). 127–153.
- Kempe, Vera, Patricia J. Brooks, Natalija Mironova & Olga Fedorova. 2003. Diminutivization supports gender acquisition in Russian children. *Journal of Child Language* 30. 471–485.
- Kikuchi, Seiichiro. 1997. A correspondence-theoretic approach to alternating diphthongs in Spanish. *Journal of Linguistic Science Tohoku University* 1. 39–50.
- Kilani-Schoch, Marianne & Wolfgang U. Dressler. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.
- Koenig, Jean-Pierre. 1999. *Lexical relations*. Stanford: CSLI publications.
- Kohavi, Ron. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on artificial intelligence*, 1137–1143. Morgan Kaufmann.
- Köpcke, Klaus-Michael. 1988. Schemas in German plural formation. *Lingua* 74(4). 303–335.
- Köpcke, Klaus-Michael. 1998a. Prototypisch starke und schwache Verben der deutschen Gegenwartssprache. *Germanistische Linguistik* 141(142). 45–60.
- Köpcke, Klaus-Michael. 1998b. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language* 25. 293–319.
- Köpcke, Klaus-Michael, Klaus-Uwe Panther & David A. Zubin. 2010. Motivating grammatical and conceptual gender agreement in German. In Hans-Jörg Schmid & Susanne Handl (eds.), *Cognitive foundations of linguistic usage patterns*, 171–194. Berlin, Boston: De Gruyter Mouton.
- Köpcke, Klaus-Michael & David A. Zubin. 1984. Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*. 26–50.
- KordiĆ, Snježana. 1997. *Serbo-Croatian*. München: Lincom Europa.
- Kramer, Ruth. 2015. Impoverishment, gender and number: Predicting the patterns of syncretism. In *Roots IV*. Georgetown University.
- Krott, Andrea, R. Harald Baayen & Robert Schreuder. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39. 51–94.
- Kuryłowicz, Jerzy. 1945. La nature des procès dits «analogiques». *Acta Linguistica* 5(1). 15–37.

- Lečić, Dario. 2015. Morphological doublets in Croatian: The case of the instrumental singular. *Russian Linguistics* 39(3). 375–393.
- Lee, Hansol H. B. 1989. *Korean grammar*. Oxford, New York: Oxford University Press.
- Lepage, Yves. 1998. Solving analogies on words: An algorithm. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, 728–734.
- Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10(8). 707–710.
- Ljubešić, Nikola & Filip Klubička. 2014. {bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian. In *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg, Sweden: Association for Computational Linguistics.
- Lyster, Roy. 2006. Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies* 16. 69–92.
- Maiden, Martin. 2001. A strange affinity: ‘Perfecto y tiempos afines’. *Bulletin of Hispanic Studies* 78(4). 441–464.
- Maiden, Martin. 2005. Morphological autonomy and diachrony. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2004*, 137–175. The Netherlands: Springer.
- Malkiel, Yakov. 1966. Diphthongization, monophthongization, metaphony: Studies in their interaction in the paradigm of the Old Spanish-ir verbs. *Language* 42(2). 430–472.
- Malkiel, Yakov. 1988. A Cluster of (Old) Portuguese derivational suffixes:–” ece,–ice, ez (a)”, viewed in relation to their Spanish counterparts. *Bulletin of Hispanic studies* 65(1). 1–19.
- Marchal, Harmony, Maryse Bianco, Philippe Dessus & Benoît Lemaire. 2007. The development of lexical knowledge: Toward a model of the acquisition of lexical gender in French. In *Proceedings of the european cognitive science conference 2007*, 268–273. Taylor and Francis.
- Mateo, Francis & Antonio J. Rojo Sastre. 1995. *El arte de conjugar en español: Diccionario de 12000 verbos*. Paris: Hatier.
- Matthews, Clive A. 2005. French gender attribution on the basis of similarity: A comparison between AM and connectionist models. *Journal of Quantitative Linguistics* 12. 262–296.
- Matthews, Clive A. 2010. On the nature of phonological cues in the acquisition of French gender categories: Evidence from instance-based learning models. *Lingua* 120(4). 879–900.

## References

- McClelland, James L. & David E. Rumelhart. 1986. A distributed model of human learning and memory. In James L. McClelland & David E. Rumelhart (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Psychological and biological models*, 170–2015. Cambridge: Mit Press.
- McDonough, Joyce M. 2013. The Dene verb: How phonetics supports morphology. In *Proceedings of 18th Workshop on Structure and Constituency in the Languages of the Americas*. University of California, Berkeley.
- Meyniel, Jean-Philippe, Paul H. Cottu, Charles Decraene, Marc-Henri Stern, Jérôme Couturier, Ingrid Lebigot, André Nicolas, Nina Weber, Virginie Fourchotte & Séverine Alran. 2010. A genomic and transcriptomic approach for a differential diagnosis between primary and secondary ovarian carcinomas in patients with a previous history of breast cancer. *BMC Cancer* 10(1). 1–10.
- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig & Jon Orwant. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182.
- Migeod, Frederick William Hugh. 1914. *A grammar of the Hausa language*. London: K. Paul, Trench, Trübner & co., ltd.
- Mladenović, A. 1977. Neka pitanja varijantnosti norme u savremenom srpskohrvatskom književnom jeziku. In Stanisław Urbáńczyk (ed.), *Wariancja normy we współczesnych słowiańskich językach literackich*, vol. 38 (Prace Komisji Słowianoznawstwa), 51–56. Kraków.
- Morin, Regina. 2006. Spanish gender assignment in computer and Internet related loanwords. *Rivista di Linguistica* 18. 325–54.
- Moscoso del Prado Martín, Fermín, Aleksandar Kostić & R. Harald Baayen. 2004. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94. 1–18.
- Motsch, Wolfgang. 1977. Ein Plädoyer für die Beschreibung von Wortbildungen auf der Grundlage des Lexikons. In Herbert Ernst Brekle & Kastovsky Dieter (eds.), *Perspektiven der Wortbildungsforschung*, 180–202.
- Müller, Stefan & Stephen Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40. 1–76.
- Mürk, Harri William. 1997. *A handbook of Estonian: Nouns, adjectives and verbs*. Bloomington: Indiana University, Research Institute for Inner Asia Studies.
- Murtagh, Fionn & Pierre Legendre. 2014. Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification* 31(3). 274–295.

- Mwalonya, Joseph, Alison Nicolle, Steve Nicolle & Juma Zimbu. 2004. *Mgombato: Digo-English-Swahili Dictionary*. Kwale: Digo Language and Literacy Project.
- Naden, Tony. 1988. The Gur languages. *The languages of Ghana* 2. 12–49.
- Naden, Tony. 1989. Gur. In John Bendor-Samuel (ed.), *The Niger-Congo languages*, 141–168. New York: University Press of America.
- Nastase, Vivi & Marius Popescu. 2009. What’s in a name? In some languages, grammatical gender. In *Conference on empirical methods in natural language processing*, 1368–1377. Singapore: ACL and AFNLP.
- Năvălici, Cristian. 2013. PyDEX. <https://github.com/cristianav/PyDEX>, accessed 2016-6-15.
- Neuvel, Sylvain. 2001. Pattern analogy vs. word-internal syntactic structure in West-Greenlandic: Towards a functional definition of morphology. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*, 253–278. Amsterdam: Springer.
- Nevins, Andrew. 2011. Phonologically conditioned allomorph selection. In Colin Ewen, Elizabeth Hume, Marc Van Oostendorp & Keren Rice (eds.), *The companion to phonology*, 2357–2382. London & New York: Continuum.
- Newcombe, Robert G. 1998. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med. Statistics in Medicine*. 857–72.
- Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.
- Niggli, Idda & Urs Niggli. 2007. *Dictionnaire bilingue Kasim-Français Français-Kassem*. SIL International. [kassem-bf.webonary.org/](http://kassem-bf.webonary.org/), accessed 2016-10-11.
- Nosofsky, Robert M. 1990. Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology* 34(4). 393–418.
- Nosofsky, Robert M., Steven E. Clark & Hyun Jung Shin. 1989. Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(2). 282–304.
- Nurse, Derek & Thomas J. Hinnebusch. 1993. *Swahili and Sabaki: A linguistic history*. Berkeley: University of California Press.
- O’Bryan, Margie. 1974. The role of analogy in non-derived formations in Zulu. *Studies in the Linguistic Sciences* 4. 144–178.
- Paul, Hermann. 1880. *Prinzipien der Sprachgeschichte*. Tübingen: Walter de Gruyter.
- Phelps, Elaine. 1975. Simplicity criteria in generative phonology - Kasem nominals. *Linguistic Analysis* 1(4). 297–332.

## References

- Phelps, Elaine. 1979. Abstractness and rule ordering in Kasem: A refutation of Halle's maximizing principle. *Linguistic Analysis* 5(1). 29–69.
- Pinker, Steven & Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6(11). 456–463.
- Pirrelli, Vito & Stefano Federici. 1994a. Derivational paradigms in morphonology. In *Proceedings of the 15th conference on Computational linguistics-Volume 1*, 234–240.
- Pirrelli, Vito & Stefano Federici. 1994b. On the pronunciation of unknown words by analogy in text-to-speech systems. In *Proceedings of the Second Onomastica Research Colloquium*.
- Pollard, Carl & Ivan A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
- Port, Robert. 2010a. Forget about phonemes: Language processing with rich memory. In *Proceedings of the interspeech 2010*.
- Port, Robert F. 2010b. Rich memory and distributed phonology. *Language Sciences* 32(1). 43–55.
- Pothos, Emmanuel M. 2005. The rules versus similarity distinction. *Behavioral and Brain Sciences* 28. 1–14.
- Pountain, Christopher J. 2006. Gender and Spanish agentive suffixes: Where the motivated meets the arbitrary. *Bulletin of Spanish Studies* 83(1). 19–42.
- Protassova, Ekaterina & Maria D. Voeikova. 2007. Diminutives in Russian at the early stages of acquisition. *Language Acquisition and Language Disorders* 43. 43–72.
- R Development Core Team. 2008. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rainer, Franz. 1993. *Spanische Wortbildungslehre*. Berlin, New York: Walter de Gruyter.
- Rainer, Franz. 2013. Formación de palabras y analogías: Aspectos diacrónicos. In Isabel Pujol Payet (ed.), *Formación de palabras y diacronía* (Anexos Revista de Lexicografía 19), 141–172. A Coruña: Servicio de Publicaciones.
- Roca, Iggy. 2010. Theme vowel allomorphy in Spanish verb inflection: An autosegmental optimality account. *Lingua. Verb First, Verb Second* 120(2). 408–434.
- Roelofs, Ardi & R. Harald Baayen. 2002. Morphology by itself in planning the production of spoken words. *Psychonomic Bulletin & Review* 9(1). 132–138.
- Rojo, Guillermo. 2008. De nuevo sobre la frecuencia de las formas llegara y llegase. In Jörn Albrecht & Frank Harslem (eds.), *Heidelberger Spätlese. Aus-*

- gewählte Tropfen aus verschiedenen Lagen der spanischen Sprach-und Übersetzungswissenschaft Festschrift, vol. 70, 161–182.
- Rokach, Lior & Oded Maimon. 2005. Clustering methods. In *Data mining and knowledge discovery handbook*, 321–352. Springer.
- Rubach, Jerzy. 2007. Feature geometry from the perspective of Polish, Russian, and Ukrainian. *Linguistic Inquiry* 38(1). 85–138.
- Rubach, Jerzy & Geert E. Booij. 2001. Allomorphy in optimality theory: Polish iotation. *Language*. 26–60.
- Rumelhart, David E. & James L. McClelland. 1986a. On learning the past tenses of English verbs. In *Parallel distributed processing: Explorations in the microstructure of cognition: Psychological and biological models*. Cambridge: MIT Press.
- Rumelhart, David E. & James L. McClelland (eds.). 1986b. *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2 Psychological and biological models*. Cambridge: MIT Press.
- Russell, Donald Andrew & Michael Winterbottom (eds.). 1989. *Classical literary criticism*. Oxford, New York: Oxford University Press.
- Sadler, Louisa. 2006. Gender resolution in Rumanian. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*. Stanford, CA: CSLI Publications.
- Sag, Ivan A., Hans C. Boas & Paul Kay. 2012. Introducing sign-based construction grammar. In Ivan A. Sag Hans C. Boas (ed.), *Sign-based construction grammar*, 69–202. Wiley.
- Saldanya, Manuel Pérez & Teresa Vallès. 2005. Catalan morphology and low-level patterns in a network model. *Catalan Journal of Linguistics* 4. 199–223.
- Salim, Bello Ahmad. 1981. *Linguistic borrowing as external evidence in phonology: The assimilation of English loanwords in Hausa*. York: University of York dissertation.
- Salmons, Joseph C. 1993. The structure of the lexicon: Evidence from German gender assignment. *Studies in Language* 17(2). 411–435.
- Sánchez, María F. 1995. *Clasificación y análisis de préstamos del inglés en la prensa de España y México*. Lewiston: Edwin Mellen Press.
- Schlücker, Barbara & Ingo Plag. 2011. Compound or phrase? Analogy in naming. *Lingua* 121(9). 1539–1551.
- Schmid, Helmut. 1995. Treetagger: A language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart* 43. 28.
- Scholkopf, Bernhard & Alexander J. Smola. 2001. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA.: MIT Press.



## References

- Schön, James Frederick. 1862. *Grammar of the Hausa language*. London: Church missionary house.
- Schwenter, Scott. 2013. Strength of priming and the maintenance of variation in the Spanish past subjunctive. [https://www.academia.edu/4857119/\\_Strength\\_of\\_Priming\\_and\\_the\\_-Maintenance\\_of\\_Variation\\_in\\_the\\_Spanish\\_Past\\_Subjunctive-NWAV\\_42\\_2013](https://www.academia.edu/4857119/_Strength_of_Priming_and_the_-Maintenance_of_Variation_in_the_Spanish_Past_Subjunctive-NWAV_42_2013), accessed 2015-10-10.
- Schwichtenberg, Beate & Niels O. Schiller. 2004. Semantic gender assignment regularities in German. *Brain and Language* 90(1). 326–337.
- Seigneuric, Alix, Daniel Zagar, Fanny Meunier & Elsa Spinelli. 2007. The relation between language and cognition in 3- to 9-year-olds: The acquisition of grammatical gender in French. *Journal of Experimental Child Psychology* 96(3). 229–246.
- Singh, Rajendra & Alan Ford. 2003. In praise of Śakaṭāyana: Some remarks on whole word morphology. In Rajendra Singh, Stanley Starosta & Sylvain Neuvel (eds.), *Explorations in seamless morphology*, 66–76. New Delhi, Thousand Oaks, London: Sage.
- Singh, Rajendra, Stanley Starosta & Sylvain Neuvel (eds.). 2003. *Explorations in seamless morphology*. New Delhi, Thousand Oaks, London: Sage.
- Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Springer Science & Business Media.
- Skousen, Royal. 1992. *Analogy and structure*. Dordrecht: Springer.
- Skousen, Royal, Deryle Lonsdale & Dilworth B. Parkinson. 2002. *Analogical modeling: An exemplar-based approach to language*. Amsterdam: John Benjamins.
- Smead, Robert N. 2000. On the assignment of gender to Chicano anglicisms: Processes and results. *Bilingual Review/La Revista Bilingüe*. 277–297.
- Smola, Alex J. & Bernhard Schölkopf. 1998. *Learning with kernels*. Sankt Augustin: GMD Forschungszentrum Informationstechnik.
- Song, Jae Jung. 2006. *The Korean language: Structure, use and context*. London & New York: Routledge.
- Steels, Luc. 2011. *Design patterns in fluid construction grammar*. Amsterdam, Philadelphia: John Benjamins Publishing.
- Steriade, Donca. 2008. A pseudo-cyclic effect in Romanian morphophonology. *Inflectional Identity* 18.
- Strauss, Trudie & Michael Johan von Maltitz. 2017. Generalising Ward’s method for use with Manhattan distances. *PLoS ONE* 12(1).
- Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*. Cambridge, New York: Cambridge University Press.

- Suh, Yunju. 2008. Korean suffix allomorphy in OT. In *The Proceedings of SUNY/CUNY/NYU Mini-Conference: Linguistics in the Big Apple*.
- Taylor, John R. 2012. *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.
- Thornton, Anna M. 2010a. Diachronic paths to reduction and maintenance of overabundance in Italian verb paradigms'. In *14th IMM, Budapest*. Budapest.
- Thornton, Anna M. 2010b. Towards a typology of overabundance. In *Décembrettes 7: International Conference on Morphology, University of Toulouse*, 2–3. Toulouse.
- Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): A non-canonical phenomenon in Italian verb morphology. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from romance inflectional morphology*. Oxford: Oxford University Press.
- Trask, Robert Lawrence. 1996. *Historical linguistics*. Oxford, New York: Oxford University Press.
- Tucker, G. Richard, Wallace E. Lambert & André Rigault. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: De Gruyter.
- Tucker, G. Richard, Wallace E. Lambert, André Rigault & Norman Segalowitz. 1968. A psychological investigation of French speakers' skill with grammatical gender. *Journal of Verbal Learning and Verbal Behavior* 7(2). 312–316.
- Ullman, Michael T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30(1). 37–69.
- Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92(1). 231–270.
- Vallès, Teresa. 2004. *La creativitat lèxica en un model basat en l'ús: Una aproximació cognitiva a la neologia i la productivitat*. Barcelona: L'Abadia de Montserrat.
- van Marle, Jaap. 1985. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.
- Venables, William N. & Brian D. Ripley. 2002. *Modern applied statistics with S*. Fourth. New York: Springer.
- Viks, Ülle. 1992. *A concise morphological dictionary of Estonian: Introduction & grammar*. Vol. 1. Tallinn: Estonian Academy of sciences, Institute of language and literature.

## References

- Viks, Ülle. 1994. A morphological analyzer for the Estonian language: The possibilities and impossibilities of automatic analysis. *Automatic Morphology of Estonian* 1. 7–28.
- Viks, Ülle. 1995. *Rules for recognition of inflection types*. <http://www.eki.ee/teemad/morfoloogia/viks2.html>, accessed 2016-6-10.
- Voeykova, Maria D. 1998. Acquisition of diminutives by a Russian child: Preliminary observations in connection with the early adjectives. *Studies in the acquisition of number and diminutive marking*. 97–113.
- Voigtlander, Katherine & Artemisa Echegoyen. 2007. *Gramática del yuhú: Otomí de la Sierra Madre Oriental*. Mexico, D. F.: Instituto Lingüístico de Verano. <https://www.sil.org/resources/archives/2018>, accessed 2017-2-21.
- Vrabie, Emil. 1989. On the distribution of the neuter plural endings in Modern Standard Romanian (MSR). *The Slavic and East European Journal* 33(3). 400–410.
- Vrabie, Emil. 2000. Feminine noun plurals in Standard Romanian. *The Slavic and East European Journal* 44(4). 537–552.
- Wanner, Dieter. 2006. An analogical solution for Spanish Soy, Doy, Voy, and Estoy. *Probus* 18(2). 267–308.
- Wechsler, Stephen. 2008. Elsewhere in gender resolution. In Kristin Hanson & Sharon Inkelas (eds.), *The nature of the word: Essays in honor of Paul Kiparsky*, 567–586. MIT Press.
- Welmers, William E. 1973. *African language structures*. Berkeley: University of California Press.
- Whitaker, William. 2016. *William Whitaker's words*. <http://mk270.github.io/whitakers-words/>, accessed 2016-6-6.
- Whitney, William Dwight. 1986. *Sanskrit grammar: Including both, the classical language and the older dialects of Veda and Brāhmaṇa*. London: Kegan Paul, Trench, Trübner & Co.
- Wikimedia Foundation. 2016. *Wiktionary*. [https://en.wiktionary.org/wiki/Appendix:Swahili\\_noun\\_classes](https://en.wiktionary.org/wiki/Appendix:Swahili_noun_classes), accessed 2016-9-9.
- Wilkinson, Hugh E. 1971. Vowel alternation in the Spanish -ir verbs. *Ronshu* 12. 1–21.
- Wills, Andy J. & Emmanuel M. Pothos. 2012. On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin* 138(1). 102–125.
- Yaden, Bridet. 2003. Mental representations of Spanish morphology: Rules or analogy? In Paula Kempchinsky & Carlos-Eduardo Piñeros (eds.), *Theory, practice, and acquisition*, 299–312. Somerville, MA: Cascadilla Press.

- Yvon, François. 1997. Paradigmatic cascades: A linguistically sound model of pronunciation by analogy. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, 428–435.
- Zaleska, Joanna. 2017. *Coalescence without Coalescence*. Leipzig: Universität Leipzig dissertation.
- Zubin, David A. & Klaus-Michael Köpcke. 1984. Affect classification in the German gender system. *Lingua* 63(1). 41–96.
- Zubin, David A. & Klaus-Michael Köpcke. 1985. Natural classification in language. A study of the German gender system. *Buffalo Cognitive Science Report* 2.
- Zubin, David A. & Klaus-Michael Köpcke. 1986. Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. In Colette G. Craig (ed.), *Noun classes and categorization*, 139–180. Amsterdam: John Benjamins.
- Zwicky, Arnold M. 1986. The general case: Basic form versus default form. In *Annual Meeting of the Berkeley Linguistics Society*, vol. 12, 305–314.



# Language index

Bosnian, Croatian and Serbian, 105–  
111

Derivation, 111–117  
Dutch, 5

English, 6, 24  
Estonian, 22

German, 6, 7, 11  
Gothic, 13–15

Hausa, 130–135  
Highland Otomi, 127–129

Kasem, 164–207

Latin, 7, 77–81

Romanian, 81–103  
Russian, 111–117

Sanskrit, 48  
some language, *see* some other lan-  
guage  
*see also* some other lect also  
of interest

Spanish, 6, 8, 10, 139–163  
Swahili, 119–126



# Subject index

- Accuracy, 64
- Analogy and rules, 28–30
- ATC, 42–53
- Class distance, 69
- Clustering, 69
- Expected Accuracy, 64
- False Negative, 64
- False Positive, 64
- Feature structures, 37–38
- Inflection, 139–207
- Kappa score, 64
- Multiple rule systems, 22–26
- Neural networks, 26
- Phonologically conditioned allomorphy, 20
- Prefixation, 119–126
- Proportional analogy, 8–11
- Rule systems, 17–20
- Schemata, 20
- Softmax, 60
- some term, *see* some other term
  - see also* some other term
  - also of interest
- True Negative, 64
- True Positive, 64
- Type hierarchies, 38–42
- Variable importance, 69







# Analogical classification in formal grammar

The organization of the lexicon, and especially the relations between groups of lexemes is a strongly debated topic in linguistics. Some authors have insisted on the lack of any structure of the lexicon. In this vein, Di Sciullo Williams (1987: 3) claim that “[t]he lexicon is like a prison – it contains only the lawless, and the only thing that its inmates have in common is lawlessness”. In the alternative view, the lexicon is assumed to have a rich structure that captures all regularities and partial regularities that exist between lexical entries. Two very different schools of linguistics have insisted on the organization of the lexicon.

On the one hand, for theories like HPSG (Pollard Sag 1994), but also some versions of construction grammar (Fillmore Kay 1995), the lexicon is assumed to have a very rich structure which captures common grammatical properties between its members. In this approach, a type hierarchy organizes the lexicon according to common properties between items. For example, Koenig (1999: 4, among others), working from an HPSG perspective, claims that the lexicon “provides a unified model for partial regularities, medium-size generalizations, and truly productive processes”.

On the other hand, from the perspective of usage-based linguistics, several authors have drawn attention to the fact that lexemes which share morphological or syntactic properties, tend to be organized in clusters of surface (phonological or semantic) similarity (Bybee Slobin 1982; Skousen 1989; Eddington 1996). This approach, often called analogical, has developed highly accurate computational and non-computational models that can predict the classes to which lexemes belong. Like the organization of lexemes in type hierarchies, analogical relations between items help speakers to make sense of intricate systems, and reduce apparent complexity (Köpcke Zubin 1984).

Despite this core commonality, and despite the fact that most linguists seem to agree that analogy plays an important role in language, there has been remarkably little work on bringing together these two approaches. Formal grammar traditions have been very successful in capturing grammatical behaviour, but, in the process, have downplayed the role analogy plays in linguistics (Anderson 2015). In this work, I aim to change this state of affairs. First, by providing an explicit formalization of how analogy interacts with grammar, and second, by showing that analogical effects and relations closely mirror the structures in the lexicon. I will show that both formal grammar approaches, and usage-based analogical models, capture mutually compatible relations in the lexicon.

ISBN 978-3-96110-186-3

