

Analogical classification in formal grammar

Matías Guzmán Naranjo

Draft
of May 23, 2019, 11:41

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. Bowerman, 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

Collaborative reading: 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_{La}T_EX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

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

4	Methodological notes	53
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
5	Gender systems	73
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	102
6	Hybrid classes	105
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
7	Morphological processes and analogy	119
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
8	Complex inflectional classes	139
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
9	Concluding remarks	209
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
	References	213
	Index	235
	Name index	235
	Language index	235
	Subject index	237

5 Gender systems

In this chapter, I discuss two cases of gender assignment and inflection class interaction: Latin third declension nouns and Romanian nouns. The question of gender assignment is an old one, and there are many papers proposing analogical models to account for this phenomenon in different languages. Some early work on the matter concluded that “there seems to be no practical criterion by which the gender of a noun in German, French or Latin [can] be determined” (Bloomfield 1933: p. 280). But since Bloomfield, there has been great progress towards establishing the opposite conclusion: “French grammarians have been hasty in their conclusion that there are no regularities or only minimal ones to gender determination” (Tucker et al. 1968: 316), and “gender can be predicted for a large proportion of German nouns, and that there is a complex interplay of overlapping semantic, morphological and phonological factors” (Corbett 1991: 49).

Corbett (1991), for example, reports on a series of languages where he notes that the shape of nouns is a strong predictor of their inflectional class and gender:

Declensional type may in turn overlap with phonology; it may be possible to predict the declensional type from the phonological shape of the stem. Where this is systematically the case, we shall consider it to be phonological assignment; this is the simpler claim, since phonological information must in any case be stored in the lexicon (Corbett 1991: 34)

The most relevant work on gender prediction can be found in C. Matthews (2005; 2010, see also Lyster 2006) where the author looked at French¹ gender assignment. Matthews (2010: 879) found that “the results [of the model] show that not only does the final syllable prove a reliable indicator but that it is, in fact, more reliable than most other sequences” (see Marchal et al. (2007) and Seigneuric et al. (2007) for evidence that children use these cues when learning French nouns, but compare Boloh & Ibernnon (2010)). Similar to French, gender

¹Non-indoeuropean languages have received considerably less attention, exceptions being Navajo (Southern Athabaskan) (Eddington & Lachler 2006; McDonough 2013) and Swahili (Bantu), discussed in Chapter §7.

assignment in Spanish has received a lot of attention (Morin 2006; Sánchez 1995; Smead 2000), including some analogical computational models (Eddington 2002).

Similarly, for German, there is a vast amount of background on how speakers predict the gender of nouns (Hahn & Nakisa 2000; Köpcke & Zubin 1984; Köpcke et al. 2010; Zubin & Köpcke 1985; Salmons 1993; Schwichtenberg & Schiller 2004; Zubin & Köpcke 1986; 1984). In Köpcke & Zubin (1984), the authors propose a series of schemata for predicting the gender of German monosyllabic words, with 90% accuracy. These schemata are partly phonological and partly semantic. The authors also found several semantic factors underlying the system, like the fact that specific concepts tend to be feminine or masculine, while more abstract concepts tend to be neuter.

These studies have mostly focused on the properties of the system but others have also explored the cognitive underpinnings of gender assignment, and how analogical systems are actually responsible for how gender is assigned to new nouns (Holmes & Segui 2004; Caffarra et al. 2015; Caffarra & Barber 2015; Taylor 2012).

A key point worth emphasizing is the difference between gender and inflection class. Gender relates to agreement, inflection class is about the actual markers. The need to differentiate between both has been made explicit before (Aronoff 1994; Harris 1991) but it is not always made explicit. Although there is often correlation between gender and inflection class in nouns and adjectives, as the examples of this chapter will show, this correlation is only partial.

In the following two sections, I explore two languages which have received less attention from an analogical perspective: Latin and Romanian. These two showcases were chosen due to the shapes of their systems. In Latin, we have a very tree-like hierarchy, which allows us to explore what happens in simple configurations with few classes. The Romanian gender-inflection class interaction offers a more complex case, in which there are multiple proposals regarding the correct number of genders found in Romanian, and how they relate to inflection class.

5.1 Masculine-feminine syncretism: Latin

5.1.1 The Latin third declension

In the Latin third declension, we find syncretisms between the masculine and feminine nouns². Table 5.1 shows that the masculine noun *pater* ('father') and the feminine noun *vox* ('voice') have the same inflectional endings, while the neuter noun *nomen* ('name') presents a different set of endings. Some gender assignment rules have already been proposed for these nouns. Aronoff (1994) proposes a series of regularities but in the end does not pursue a completely formalized system.

Table 5.1: Paradigms for *pater* ('father'), *vox* ('voice') and *nomen* ('name').

	masculine		feminine		neuter	
	singular	plural	singular	plural	singular	plural
nom.	<i>pater</i>	<i>patr-ēs</i>	<i>vox</i>	<i>vōc-ēs</i>	<i>nōme-n</i>	<i>nōmin-a</i>
acc.	<i>patr-em</i>	<i>patr-ēs</i>	<i>vōc-em</i>	<i>vōc-ēs</i>	<i>nōm-en</i>	<i>nōmin-a</i>
gen.	<i>patr-is</i>	<i>patr-um</i>	<i>vōc-is</i>	<i>vōc-um</i>	<i>nōm-inis</i>	<i>nōmin-um</i>
dat.	<i>patr-ī</i>	<i>patr-ibus</i>	<i>vōc-ī</i>	<i>vōc-ibus</i>	<i>nōm-inī</i>	<i>nōmin-ibus</i>
abl.	<i>patr-e</i>	<i>patr-ibus</i>	<i>vōc-e</i>	<i>vōc-ibus</i>	<i>nōm-ine</i>	<i>nōmin-ibus</i>

This same third declension syncretism is also found in adjectives. Take for example the paradigm for *vetus* ('old') in Table 5.2. Again, masculine and feminine classes take the same endings.

From a declension class perspective, this system is fairly simple³. The hierarchy in (1) basically says that feminine and masculine form a class, which easily captures the syncretism in that one inflectional construction will apply to neuters and one to non-neuters for the third declension.

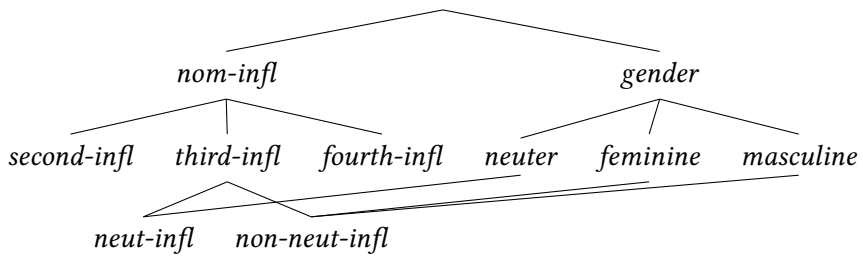
²The reason for only focusing on third declension nouns is precisely that this is the only declension class in Latin where we clearly find all three genders abundantly represented. Focusing only on one of the five declension classes also means that we are removing the effects of crossing trees like in Romanian, Spanish or Kasem.

³It is simple because it only considers the third declension. The complete nominal declension system is much more complex.

Table 5.2: Paradigm for *vetus*, *veteris* ('old').

Case	masculine/feminine		neuter	
	singular	plural	singular	plural
nom.	vetus	veter-ēs	vetus	veter-a
acc.	veter-em	veter-ēs	vetus	veter-a
gen.	veter-is	veter-um	veter-is	veter-um
dat.	veter-ī	veter-ibus	veter-ī	veter-ibus
abl.	veter-e	veter-ibus	veter-e	veter-ibus
voc.	vetus	veter-ēs	vetus	veter-a
loc.	veter-ī	veter-ibus	veter-ī	veter-ibus

(1)



One set of constructions (or rules, etc.) would apply to the *neuter* type, while a different set would apply to the *non-neuter-infl* type, thus producing the observed syncretisms. The expectation would then be that masculine and feminine lexemes in the third declension should be more similar to each other than to neuter nouns.

5.1.2 Data

I extracted all third declension nouns from the digital Latin dictionary *Words* by Whitaker (2016). The totals by gender (after removing nouns marked with common gender, e.g. *celestis* ‘divinity’) are: 7773 feminine, 2993 masculine, 1499 neuter. We can see that there are many more feminine nouns than neuter or masculine nouns.⁴

⁴Note that we would expect such a disproportion to favour a model that grouped feminine nouns against neuter and masculine nouns. The reason for this being that when a model cannot reliably predict the class of some item, it tends to assign it to the most frequently observed class (since this is the most likely outcome). In other words, in a no information situation, it is more likely that a noun will be feminine than masculine or neuter.

As the basis for the analogy, I used the stems provided in the dictionary. This is likely to introduce an accuracy bias into the model, as it does not filter derivational morphology. Some Latin suffixes are gender assigning: third declension nouns ending in *-tor* are mostly masculine, with around four exceptions: *caritor* ('wool-carders', feminine), *litor* ('beach, landing place', neuter), *pector* ('breast, heart', neuter). It is clear that these cases do not really contain a derivational *-tor* suffix but rather happen to end in a *tor* sequence. Similar cases for feminine nouns are gender assigning suffixes like *-tat* and *-tas*: *absurditas* ('absurdity').

This particularity of the dataset, however, should not really represent a problem for the question at hand. It is true that the model will confound some morphological with phonological analogies, but these effects should have no impact either way on the similarity clustering over the three classes. If anything, the additional morphological information would reduce confusion between masculine and feminine classes. Nonetheless, I present models on two datasets, one which includes all derived nouns and a reduced dataset excluding clear cases of gender-assigning suffixes. The number of nouns by gender in the reduced dataset is: 6626 feminine, 2153 masculine and 1496 neuter.

5.1.3 Results

We fit an analogical model to the Latin data using the formula: `gender ~ final.1 + final.2 + final.3 + num_vowels`. This model looks at the three last segments of the stem and the number of vowels. The results can be seen in Table 5.3 and the corresponding statistics in Table 5.4.

Table 5.3: Confusion Matrix for the model predicting gender of Latin third declension nouns.

Prediction	Reference		
	Feminine	Masculine	Neuter
Feminine	7244	569	77
Masculine	432	2236	196
Neuter	97	188	1226

The equivalent model for the reduced dataset can be seen in Table 5.5 and the corresponding statistics in Table 5.6. For both datasets, the results are almost identical. As expected, the smaller dataset produces slightly worse results, be-

Table 5.4: Overall statistics for Confusion Matrix Table 5.3.

Overall statistics:			
Accuracy : 0.8729			
95% CI : (0.8669, 0.8787)			
No Information Rate : 0.6338			
Kappa : 0.7557			
Statistics by Class:			
	Feminine	Masculine	Neuter
Sensitivity	0.9319	0.7471	0.8178
Specificity	0.8562	0.9323	0.9735
Balanced Accuracy	0.8941	0.8397	0.8957

Table 5.5: Confusion Matrix for the model predicting gender of Latin third declension nouns.

	Reference		
Prediction	Feminine	Masculine	Neuter
Feminine	6114	577	70
Masculine	420	1391	182
Neuter	92	185	1244

cause the nouns removed were amongst the easily predicted ones⁵. A clustering analysis of both models can be seen side-by-side in Table 5.1.

Table 5.1 shows that the feminine and masculine nouns are closer to each other than they are to neuter nouns. This confirms the expectations of the ATC model and matches the inflectional system where we find syncretism between the masculine and feminine.

In conclusion, I have shown that in a very simple system like the one of Latin third declension nouns, the analogical model makes exactly the right predictions about how the three genders should cluster together based on formal properties of the stems. We see the same result for both datasets, with and without gender assigning suffixes.

⁵Because the derivational suffixes are identified by the model as sequences that reliably predict gender.

Table 5.6: Overall statistics for Confusion Matrix Table 5.5.

Overall statistics:			
Accuracy : 0.8515			
95% CI : (0.8445, 0.8583)			
No Information Rate : 0.6449			
Kappa : 0.7108			
Statistics by Class:			
	Feminine	Masculine	Neuter
Sensitivity	0.9227	0.6461	0.8316
Specificity	0.8227	0.9259	0.9684
Balanced Accuracy	0.8727	0.7860	0.9000

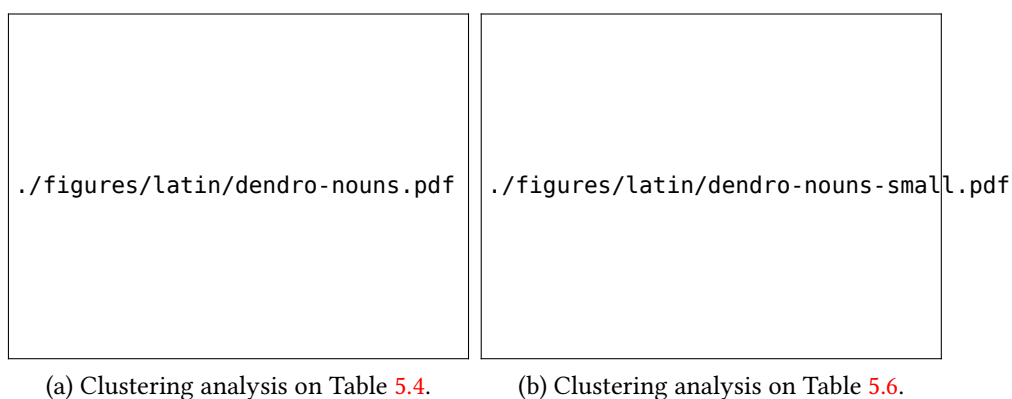


Figure 5.1: Clustering analysis for Latin gender assignment.

5.2 Gender vs inflection class: Romanian

5.2.1 The Romanian gender and plural system

A much more interesting gender and number system can be found in Romanian. Like Latin, Romanian is often analyzed as having three genders, which it inherited from Latin (Gönczöl 2007: 23). The interesting aspect of Romanian gender is that the neuter does not have a dedicated marker, but patterns with the masculine in the singular and with the feminine in the plural. As Cojocaru explains, this phenomenon can be observed on all elements that agree with a noun.

5 Gender systems

The distinctive part of the neuter gender in Romanian is that it does not have any formal particularities. The neuter nouns in the singular look like masculine nouns, while in the plural they look like feminine nouns. The same applies to adjectives, pronouns and pronominal adjectives. When they modify or replace a neuter noun in the singular they appear in their masculine singular form, and when they modify or substitute a neuter noun in the plural they appear in their feminine plural form. (Cojocaru 2003: 27)

One striking example of this situation is illustrated by the three inflection classes in Table 5.7, each of which is found in only one gender. In this part of the system, neuter nouns inflect like masculine nouns in the singular, and like feminine nouns in the plural. This means that, while there are no specific markers for neuter nouns, there is a three way split in the system.

Table 5.7: Three way gender system in Romanian.

	singular	plural	gloss
masculine	condr-u	condr-i	<i>forest</i>
neuter	muze-u	muze-e	<i>museum</i>
feminine	cas-ă	cas-e	<i>house</i>

In terms of agreement, we see the same phenomenon (adapted from Farkas 1990), as can be observed in examples (2)-(4):

(2) masculine

- a. Un trandafir alb e scump.
a.MASC.SG rose white.MASC.SG is expensive.MASC.SG
'A white rose is expensive.'
- b. Unii.MASC.PL trandafiri alb-i sunt scump-i.
some rose white-MASC.PL is expensive-MASC.PL
'Some white roses are expensive.'

(3) feminine

- a. O garoafa alb-a e scump-a.
 a.FEM.SG carnation white-FEM.SG is expensive-FEM.SG
 ‘A white carnation is expensive.’
- b. Unele garoafe alb-e sunt scump-e.
 some.FEM.PL carnation white-FEM.PL is expensive-FEM.PL
 ‘Some white carnation are expensive.’

(4) neuter

- a. Un scaun alb e scump.
 a.MASC.SG chair white.MASC.SG is expensive.MASC.SG
 ‘A white chair is expensive.’
- b. Unele scaune alb-e sunt scump-e.
 some.FEM.PL chairs white-FEM.PL are expensive-FEM.PL
 ‘Some white chairs are expensive.’

Here we have the identical type of distribution for agreement, as we saw for markers in Table 5.7. The word *alb* (‘white’) has the same marker (namely $-\emptyset$) when modifying a masculine or neuter noun in the singular, and it has the same marker (namely $-e$) when modifying a neuter or feminine noun in the plural. So, even though there are only two different agreement markers in the plural and in the singular⁶, the alignment pattern produces three genders. Additionally, Romanian has a relatively complex inflection class system for singular and plural. Table 5.8 presents the basic classes (Cojocaru 2003)⁷. Usually, the singular is taken to be a sort of simplex form, instead of being composed of a stem and a singular marker. I take a slightly different approach here and consider the singular to be composed of a stem and a singular marker.

Table 5.8 shows the problematic issue in the interaction between Romanian gender and number markers. Although gender correlates with inflection class, knowing the gender of a noun in Romanian is not enough for knowing its plural (or singular) form. Based on this fact, it has been argued that Romanian does not have three genders but rather a complex interaction between number markers. The most recent of these accounts is offered by Bateman & Polinsky (2010). The authors, partially following previous proposals by Hall (1965) and Farkas & Zec (1995), claim that there are only two genders in Romanian, masculine and

⁶In (2) one could argue either that the consonant is the marker, or that there is a $-\emptyset$ marker.

⁷Classes *iu-ie* and *u-ă* in the neuter are classified as exceptions by the author.

Table 5.8: Number inflection classes in Romanian.

Sg. marker	Pl. marker	Singular	Plural	Gloss
Masculine nouns				
-C	-C+i	elev	elevi	‘school student’
-u	-i	leu	lei	‘lion’
-e	-i	iepure	iepuri	‘rabbit’
-i	-i	ochi	ochi	‘eye’
-ă	-i	tată	tați	‘father’
Feminine nouns				
-ă	-e	casă	case	‘house’
-ă	-i	usă	usi	‘door’
-ă	-uri	marfă	mărfuri	‘merchandise’
-e	-i	lume	lumi	‘world’
-V+ie	-V+i	baie	băi	‘bathroom’
-C+ie	-C+ii	frecție	frecții	‘massage’
-a	-ale	basma	basmale	‘handkerchief’
-ea	-ele	cafea	cafele	‘coffee’
-i	-i	marți	marți	‘Tuesday’
Neuter nouns				
-C	-C+uri	tren	trenuri	‘train’
-C	-C+e	capac	capace	‘lid’
-u	-uri	lucru	lucruri	‘thing’
-u	-e	muzeu	muzee	‘museum’
-u	-ă	ou	ouă	‘egg’
-iu /ju/	-ii /ij/	exercițiu	exerciții	‘exercise’
-iu /iw/	-ie /i.e/	sicriu	sicrie	‘coffin’
-i /j/	-ie /je/	tramvai	tramvaie	‘tram’
-i /i/	-iuri	taxi	taxiuri	‘taxi’
-e	-e	nume	nume	‘name’

feminine, and that it is not gender that determines plural formation, but plural formation that helps determine gender in Romanian:

Our position is supported by the fact that in traditional three-gender analyses there is limited predictability of plural endings for nouns in the same class, clearly showing that gender specification alone does not predict plural form (Bateman & Polinsky 2010: 53)

Similarly, they claim that a two-gender system for Romanian is more parsimonious than a three gender system because “the same factors relevant for plural formation are indirectly relevant for predicting gender assignment and agreement in the plural” (Bateman & Polinsky 2010: 45).

To address the problem of agreement in Romanian nouns, Bateman & Polinsky (2010: 52) propose that “Romanian has two noun classes in the singular and in the plural” but clarify that “this categorization is not lexically specified”. These classes are in turn different for plural and singular. The first aspect of their system is easy to model with a type system, but the clarification that said noun classes are not lexically specified, less so. For this model, all lexemes in Romanian would be underspecified for a lexical class feature, which would only be specified after an inflectional process derives the singular or plural form of the noun.

For the singular, Bateman & Polinsky (2010) propose classes A and B. Traditionally feminine nouns belong to class A, while traditionally masculine and neuter nouns belong to class B. Class assignment in the singular is driven by both formal and semantic features (since animate nouns can straightforwardly be categorized as feminine or masculine, as well as other minor semantic classes). For the plural, the authors propose classes C and D. Class D includes traditional masculine nouns, while class C includes all other nouns. Again, class membership is determined by semantic and formal cues, where the formal cues are the plural endings of the nouns. A graphic representation of the two gender model is shown in Table 5.2

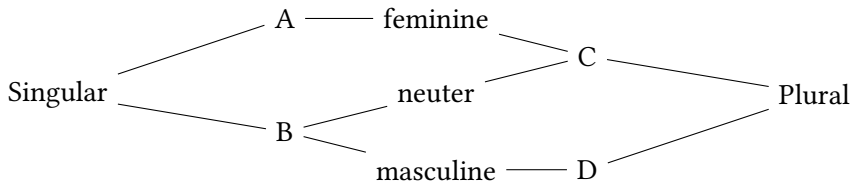


Figure 5.2: Two gender model for Romanian.

In the model by Bateman & Polinsky (2010), it is the plural class that determines gender:

In fact, with the exception of traditional masculines, all of which take the plural marker *-i*, there are very few feminine and neuter nouns for which gender classification alone can predict plural form. For example, feminine nouns ending in *-e* take the *-i* plural marker [...]. As we mentioned previously, there are also feminine nouns ending in stressed *-a* or *-ea* that take the *-le* plural marker, and there are neuter nouns ending in a stressed *-i* and borrowings from French ending in *-ow* that take *-uri* in the plural. Notice that in each of these cases the plural ending is determined by the noun's ending rather than its gender class, which supports our claim that the plural forms determine class membership in the plural, rather than the other way around (Bateman & Polinsky 2010: 54)

This approach has recently received some support from a computational model. Dinu et al. (2012) present systems based on two support vector machines, one trained on plurals and one trained on singulars, which manages to distinguish neuter nouns very well, at around 99% accuracy (for two previous computational approaches see Cucerzan & Yarowsky 2003 and Nastase & Popescu 2009). Dinu et al. (2012: 123) mention that their model supports Bateman and Polinsky's (2010) model, as plural class is in fact distinguishable for nouns from purely formal cues, and that gender is not needed. It is not completely clear from the study by Dinu et al. (2012) study, however, that gender would not provide extra information about plural formation. First, the authors looked at singular and plural words in the nominative, which means that their model had number information which is highly correlated with gender. A second issue is that the authors only considered the effect of formal cues for predicting gender, but did not fit a model that took into account the effect of gender.

Other solutions for modelling gender in Romanian have been proposed in different linguistic theories. Probably the most well known is Farkas & Zec (1995), who take an underspecification approach, where feminine nouns are specified as +FEM, masculine nouns as -FEM, and neuter nouns are not specified for gender (see also Farkas 1990, as well as Sadler 2006; Wechsler 2008; Kramer 2015). These approaches assume the existence of three genders, but diverge in how exactly their interrelations are implemented.

A different approach is pursued by Steriade (2008). Steriade identifies some phonological constraints on the plural choice of some nouns. Her approach, however, focuses on the different phonological processes that stems undergo with

certain markers, rather than on the actual choice of different markers. I will ignore stem processes in this study, but the approach developed in Chapter §8 for Spanish could be extended to the Romanian system.

5.2.2 Modelling the system

The two gender model as presented by Bateman & Polinsky (2010) has a conceptual problem. There are three types of nouns in Romanian based on their agreement behaviour. The discussion of how this originates and what features are responsible for this phenomenon is somewhat of a red herring. The fact is that we have three agreement patterns, and whether we need a lexically specified feature for this is a different question. Additionally, the argument that we do not need gender because inflection class is predictable from formal features, and because gender does not completely determine inflectional class, is not very convincing.

It is not really surprising that declension class is partially independent of gender, since this is not all that rare typologically speaking (Corbett (1991), and Chapter §8), and it is even present in other Romance languages. A simple example is Spanish, where exponents of the singular only partially correlate with gender (Harris 1991). Similarly, the fact that gender does not completely determine inflection class does not entail that gender has no information about the inflection class of a noun, independently of other formal features on the noun. Incomplete information does not mean no information.

As for the Romanian system, what does make sense in the two-gender proposed by Bateman & Polinsky (2010), is to have four agreement classes, two for singular and two for plural, and many different actual inflection classes according to the singular-plural marker combination. This idea is depicted in the hierarchy Table 5.3.

This hierarchy is exclusively about agreement because it indicates what the agreement would be with a given adjective. Notice that only listing the plural and singular for each noun is insufficient because adjectives do not agree with nouns in terms of markers, but in terms of gender:

- (5) a. tren-uri mic-i
 train.MASC.PL small.MASC.PL
 ‘small trains’
 b. lum-e mic-ă
 world.FEM.PL small.FEM.PL
 ‘small world’

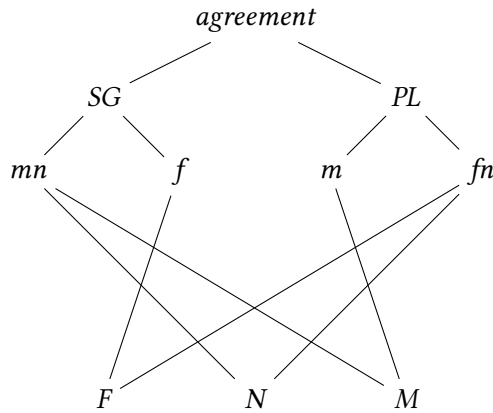


Figure 5.3: Romanian Gender-Number hierarchy.

- c. lucr-**u** mic
 thing.NEUT.PL small.NEUT.PL
 ‘small thing’

The adjective *mic* (‘small’) has three forms: *mic* (masculine and neuter singular), *mică* (feminine singular) and *mici* (plural). This adjective does not agree with the number markers on the nouns, but with their gender.

Because, as we saw, gender does not completely determine inflection class, this dimension needs to be modelled separately. For each gender, there are some markers, either singular or plural, which are unique to said gender. So, for example, the plural maker *-iuri* is only found with neuter nouns, the singular marker *-C+ie* only occurs with feminine nouns, and the plural marker *-C+i* is only found with masculine nouns. This is crucial because we cannot claim that masculine and neuter nouns share all singular markers, or that feminine and neuter nouns share all plural markers. Markers like *-ă* are found in the singular with feminine and masculine nouns, and in the plural with neuter nouns. Except for the classes where both the singular and plural use the same marker, markers are uniquely determined by the number they express. That is, even though the marker *-ă* can express singular or plural, knowing the gender of the noun immediately resolves the uncertainty. In the case of *-e* in feminine nouns, we need to know the other marker of the noun to be able to tell whether *-e* is a plural or singular marker.

The issue becomes even clearer when we look at how the different number classes distribute across genders in Table 5.9 (see next section for an explanation

Table 5.9: Number classes by gender in Romanian.

	feminine	masculine	neuter
a-ale	172	0	0
a-e	178	0	0
ă-e	11647	0	0
a-i	56	0	0
ă-i	2855	51	0
ă-uri	1590	0	0
C-e	0	0	7746
C-i	0	7252	0
C-ii	0	0	25
C-uri	0	0	5586
ea-ale	1	0	0
ea-ele	384	0	0
e-e	807	0	155
e-i	13814	227	0
e-iuri	0	0	3
e-uri	17	0	90
i-e	0	0	31
ie-e	112	0	0
ie-ii	6771	0	0
ie-Vi	171	0	0
i-i	75	567	0
i-ie	0	0	189
i-iuri	0	0	237
iu-ie	0	0	19
iu-ii	0	0	348
u-ă	0	0	1
u-e	0	0	936
u-i	0	700	0
u-uri	0	0	456

of the dataset). What we clearly see here is that, with the exception of the classes⁸ $\check{a}-i$, $e-e$, $e-uri$ and $i-i$, declension class determines gender. We also see that the confusion is with the feminine, i.e. the masculine and neuter classes are never confused. Notice that this has the reverse structure of the agreement pattern, where neuter patterns with masculine and feminine, but these two do not pattern together.

There are some additional classes not listed in Cojocaru (2003). For example, *nutria* ('otter') forms its plural as *nutrii*. Similarly, *anaconda* ('anaconda') forms its plural as *anaconde*. I leave these classes in as they are, but recognize that they might be special cases of foreign words or particular exceptions⁹.

The distribution of singular and plural markers can be seen in Table 5.10 and Table 5.11. In these distributions, we find something similar to what we had in the distribution of classes. Although there are markers that are shared by all three genders, namely $-i$ and $-e$ in the singular, there are no markers that are only shared by the neuter and feminine in the singular, and the only marker shared by the masculine and feminine is $-\check{a}$, with a suspiciously low type frequency in the masculine. On the other hand, in the plural, except for $-i$, sharing of markers is only found between neuter and feminine.

With these facts in mind, there are three alternatives for a hierarchy of number markers in Romanian. If we wanted to keep the symmetry in the hierarchy between plural and singular, we could separate the markers that cross 'the wrong' classes into two. There are two potential justifications for this move, one theoretical and one empirical. Thinking in terms of simplicity, adding three additional singular markers, and one additional plural marker reduces the complexity of the hierarchy. The second reason has to do with the relative type frequencies of the problematic markers. If we look at their distributions, in the singular, $-\check{a}$ and $-e$, as shown in Table 5.10, are **much** more common with the feminine than with the masculine or neuter. In a similar way, $-i$ has more or less the same type frequency for the neuter and masculine, and it is less frequent with the feminine. In the plural, $-i$ is much more frequent with the feminine than the masculine.

Pursuing a symmetric approach, the system would have $-e_f$, $-i_f$, $-\check{a}_f$ and $-e_{mn}$, $-i_{mn}$ and $-\check{a}_m$ in the singular; and $-i_m$ and $-i_{nf}$ in the plural. In Figure 5.4 we see what a hierarchy under these assumptions would look like.

⁸I use the notation *sg-pl*.

⁹The markers $-i$ and $-Vi$ for the plural feminine could be collapsed into a single $-i$ marker. For consistency with Cojocaru (2003), I keep them as distinct markers, but in the end this decision should not really make much of a difference.

Table 5.10: Singular classes by gender in Romanian.

	Feminine	Masculine	Neuter
a	406	0	0
ă	16092	51	0
C	0	7252	13357
e	14638	227	248
ea	385	0	0
i	75	567	457
ie	7054	0	0
iu	0	0	367
u	0	700	1393

Table 5.11: Plural classes by gender in Romanian.

	Feminine	Masculine	Neuter
ă	0	0	1
ale	173	0	0
e	12744	0	8868
ele	384	0	0
i	16800	8797	0
ie	0	0	208
ii	6771	0	373
iuri	0	0	240
uri	1607	0	6132
Vi	171	0	0

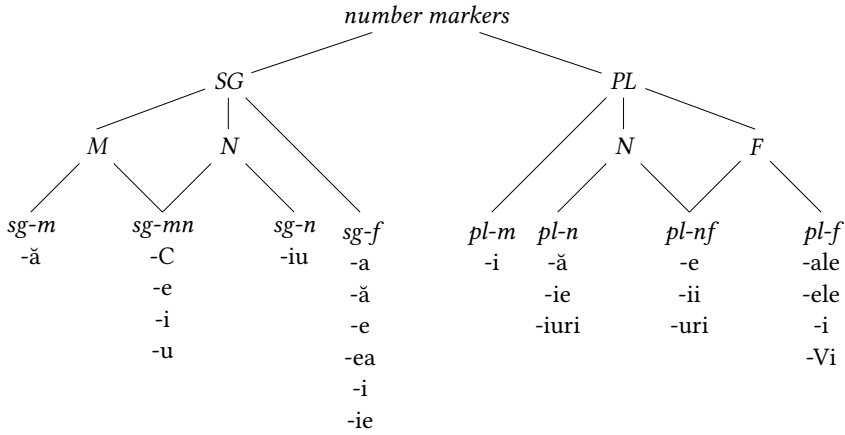


Figure 5.4: Romanian marker hierarchy.

An alternative is to have an asymmetric hierarchy, but fewer individual markers. A sketch of this hierarchy can be seen in Table 5.5. In this case, there is no real symmetry between the singular and the plural, nor is there any with the agreement patterns.

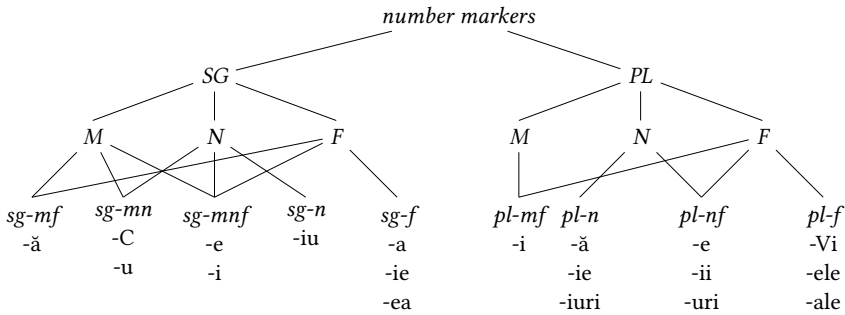


Figure 5.5: Symmetric Romanian marker hierarchy.

The final inflection classes result from specifying pairings between the singular and plural markers shown in Table 5.8. Since there is no free combination between singular and plural markers, each class must be specified directly.

The third alternative consists of two independent flat lists for singular and plural markers and then specify each inflection class as in Table 5.5. A simplified hierarchy for this approach is given in Table 5.6. The advantage of this model is that it is simpler than the previous two, in that it requires less complex interactions between types. The downside is that the partial correlations between gender and inflection class would be lost.

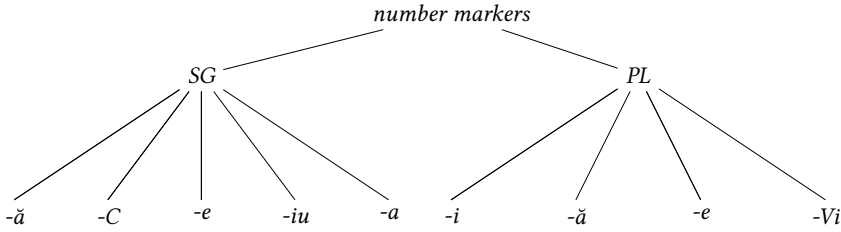


Figure 5.6: Simplified Romanian marker hierarchy.

There is no way of deciding apriori which of these three approaches is better. The choice between the three will depend on considerations pertaining to what a theory of morphology should look like. They do, however, make slightly different predictions in terms of what we should find in the analogical model. In the first hierarchy, we would expect there to be little to no confusion between feminine, masculine and neuter nouns, and there should be a separation between the classes with the *-e*, *-ă* and *-i* markers in the singular, as well as those with the *-i* markers in the plural. That is, these dimensions should not be available for the analogical model.

The hierarchy in Table 5.5 predicts that those three markers should be available for nouns to cluster together, and we should thus see classes clustering around these markers. Similarly, these classes should allow for some limited confusion between masculine and feminine, especially in the classes with the shared markers.

Finally, the hierarchy in Table 5.6 predicts that clustering should be exclusively about markers and not around gender. Therefore, we would expect classes with shared markers to cluster together, but not classes forming clusters around the genders they correlate with. The hierarchies in Table 5.4 and Table 5.5 do predict some clustering around genders and some clustering around markers.

5.2.3 Data

For this study, I used the Romanian dictionary DEX Online (<https://dexonline.ro>), taking the data base from the python api (Năvălici 2013). From the dictionary¹⁰, I extracted all nouns in the nominative form for which a plural form was specified. From these, I removed all nouns with a plural form ending in *s* because these are

¹⁰More specifically, from the *dex_lexemes*, *dex_lexemes_inflections* and *dex_inflections* data bases provided. The search targeted entries with the fields: *plural*, *nearticulat*, *Substantive* and *Nominative*.

clear borrowings from Spanish and other languages. Finally, I removed all nouns with common gender. This process gives us 63646 nouns. For each noun, I extracted the plural and singular markers according to the description in Cojocaru (2003) and added the extra classes not listed there.

The distribution of nouns by gender in the extracted corpus is, for a total of 63501 nouns: 38737 feminine nouns, 8891 masculine nouns and 15873 neuter nouns.

5.2.3.1 Methodology and hypothesis

There are basically two claims at stake. On the one hand, it has been argued that gender information is not helpful when figuring out the plural form of nouns in Romanian. On the other hand, we want to test which of the three inflection class configurations makes more accurate predictions regarding the analogical relations between the stems of the nouns.

To test the first claim, we can fit two different analogical models: (i) one that only looks at phonological information, which would approximate proposals for gender assignment based on the ending of nouns in Romanian like those of Vrabie (1989) and Vrabie (2000); (ii) and then a similar model that also looks at gender information. If gender carries no useful information about the plural form of nouns in Romanian, as Bateman & Polinsky (2010) claim, then we should see no difference in the performance of each of the models. If adding gender clearly increases accuracy, we can say that there is a high probability that gender does in fact play a role in predicting inflection class¹¹.

To test the second claim, we can look at the overall gender+inflection class distribution. For this second part, I used a reduced and more balanced dataset. For each noun, I extracted its class as a tuple: *gender+singular+plural*. This produces a total of 57 classes, 17 feminines, 14 masculines and 26 neuters. The distribution of classes by type frequency can be seen in Table 5.7. From these, I removed the three lowest frequency classes (marked in red in Table 5.7), and took a random sample of up to 3000 nouns for the more frequent classes. This produces a somewhat more balanced dataset.

The basic prediction is that neuter classes should be confused with feminine and masculine classes, but these two should not be confused with each other.

¹¹We cannot have complete certainty because it is always possible that a different model solely based on formal cues could outperform the model including gender.

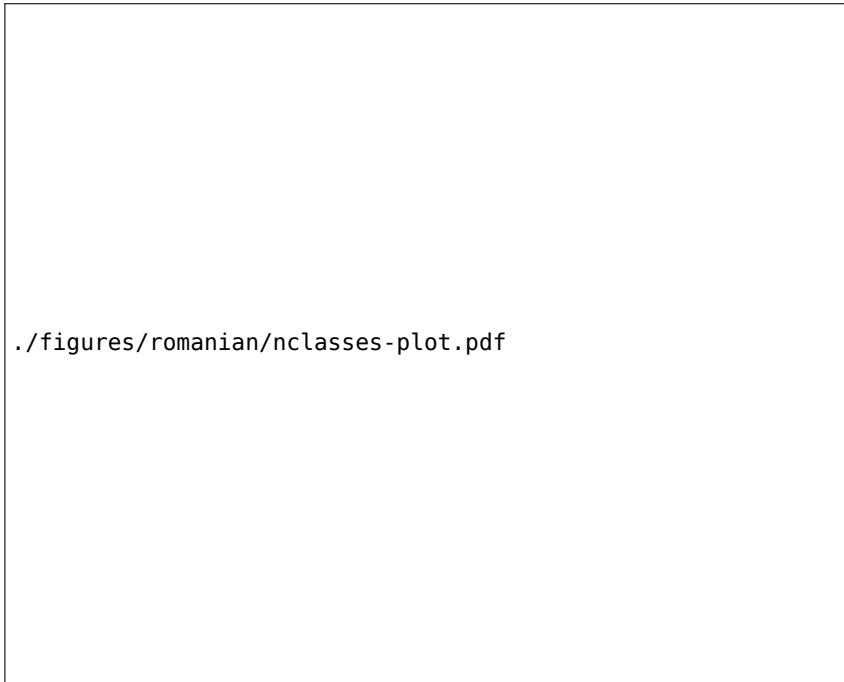


Figure 5.7: Class frequency.

5.2.4 Results

5.2.4.1 Predicting gender

First, we predict gender from the shape of the stems with the formula: $\text{gender} \sim \text{final.1} + \text{final.2} + \text{final.3} + \text{n_vowels}$.¹² Here we are simply looking at the final three segments and the number of vowels in the stem. The results can be seen in Table 5.12 and Table 5.13. The tables show that gender is in fact predictable without any information about number markers. However, in Table 5.12 we do see that there is a relatively large confusion between masculine and neuter, and between neuter and feminine, but not so much between feminine and masculine. This is again confirmed in Table 5.14 (larger numbers mean more distance between the classes). This observation matches the hierarchy in Table 5.3, where feminine and masculine genders do not share any set of common nodes, but masculine and neuter, and feminine and neuter do.

¹²There were no hidden nodes and a decay rate of 0.

Table 5.12: Confusion Matrix for the model predicting Gender of Romanian nouns.

Reference			
Prediction	Feminine	Masculine	Neuter
Feminine	14314	783	1750
Masculine	80	665	494
Neuter	987	2926	6246

Table 5.13: Overall statistics for the Confusion Matrix in Table 5.12.

Overall statistics:			
Accuracy : 0.7515			
95% CI : (0.7464, 0.7565)			
No Information Rate : 0.5446			
Kappa : 0.5564			
Statistics by Class:			
	Feminine	Masculine	Neuter
Sensitivity	0.9306	0.1520	0.7357
Specificity	0.8031	0.9759	0.8019
Neg Pred Value	0.9064	0.8626	0.8759
Balanced Accuracy	0.8669	0.5639	0.768

Table 5.14: Distance matrix for the Confusion Matrix in Table 5.12.

	Feminine	Masculine
Masculine	2.346527	
Neuter	2.256934	0.154508

5.2.4.2 Predicting singular

Next, we turn to the number markers. In this case, we have several dimensions that need to be predicted. On the one hand, there are individual number markers, and on the other hand there are complete inflection classes with and without gender distinctions. Because there are some inflection classes which can appear with two genders, it is interesting to ask how well we can distinguish these cases. Additionally, because we are mostly interested in seeing how the clusters work, we can compare whether predicting inflection class without gender produces similar clusters to those we get when predicting inflection class with gender.

We start with the singular markers with the formula: `singular ~ final.1 + final.2 + final.3 + n_vowels`¹³. The results are given in Table 5.15 and Table 5.16. We see that the singular marker, as defined here, is relatively predictable, but not perfectly.

Table 5.15: Confusion Matrix for the model predicting the singular marker of Romanian nouns.

Prediction	Reference								
	a	ă	C	e	ea	i	ie	iu	u
a	15	17	15	3	5	2	7	0	2
ă	156	6181	361	813	205	315	677	156	37
C	117	232	8104	159	37	530	441	24	648
e	51	604	218	2993	25	63	186	68	25
ea	0	10	1	1	6	3	17	4	0
i	2	25	20	12	13	80	37	3	2
ie	45	379	181	72	76	87	1784	94	5
iu	1	9	0	2	7	1	6	12	3
u	19	54	126	16	10	19	128	6	1375

5.2.4.3 Predicting plural

We now turn to plural markers. The model used is the same as for singular markers, with the formula: `plural ~ final.1 + final.2 + final.3 + n_vowels`. The results for predicting plural markers are shown in Table 5.17 and Table 5.18. What we find is that the model can predict plural markers somewhat less accu-

¹³The model had no hidden nodes and a decay rate of 0.

Table 5.16: Overall statistics for the Confusion Matrix in Table 5.15.

Overall statistics:
Accuracy : 0.7276
95% CI : (0.7223, 0.7327)
No Information Rate : 0.3196
Kappa : 0.6425

Table 5.17: Confusion Matrix for the model predicting the plural marker of Romanian nouns.

	Reference								
Prediction	ale	e	ele	i	ie	ii	iuri	uri	Vi
ale	17	12	5	13	0	7	1	21	0
e	36	4570	106	1876	57	780	35	1053	56
ele	3	6	8	10	1	15	4	2	0
i	86	2415	121	7392	85	471	102	1219	61
ie	0	6	0	6	6	0	0	10	0
ii	25	530	85	187	3	1961	30	150	0
iuri	0	6	4	18	1	23	21	3	0
uri	5	665	55	796	55	117	43	2689	38
Vi	0	9	0	22	1	0	1	12	16

Table 5.18: Overall statistics for the confusion matrix in Table 5.17.

Overall statistics:
Accuracy : 0.5905
95% CI : (0.5848, 0.5963)
No Information Rate : 0.3654
Kappa : 0.4278

rately than singular markers. Nevertheless, the accuracy and kappa scores are quite far above random chance.

Now we address the claims by Bateman & Polinsky (2010) that gender does not really help to determine the plural marker a noun will take, and that plural class assignment is solely based on phonological features (including the singular marker). Properly testing this claim is not possible because the authors do not provide a full model for plural assignment. However, one can compare a model that only includes phonological features (and the singular marker) to one which also includes gender.¹⁴ We fit a model with the formula: `plural ~ final.1 + final.2 + final.3 + n_vowels + singular + gender`. The results can be seen in Table 5.19 below.

Table 5.19: Confusion Matrix for the model predicting the plural marker of Rumanian nouns with additional gender information.

	Reference								
Prediction	ale	e	ele	i	ie	ii	iuri	uri	Vi
ale	126	51	1	4	0	0	0	0	0
e	41	6308	0	473	0	21	16	1334	25
ele	0	0	383	0	0	0	0	0	0
i	5	712	0	9807	0	0	0	30	0
ie	0	0	0	0	170	9	42	0	0
ii	0	3	0	0	11	3333	0	5	4
iuri	0	13	0	0	28	1	179	0	0
uri	0	1093	0	36	0	9	0	3790	0
Vi	0	39	0	0	0	1	0	0	142

Predicting the plural marker with all predictors (gender and singular marker) gives us the results presented in Table 5.19, and the corresponding statistics in Table 5.20. The model evaluation is given in Table 5.8. Table 5.8 shows that removing gender from the model causes a very steep drop in accuracy, i.e. gender does help in the analogical model. These results clearly speak against Bateman & Polinsky (2010)'s claim that gender does not help distinguish plural classes. If this result is correct, then there are no strong arguments for a two gender system for Romanian.

¹⁴Of course, there is always the possibility that a better model, solely based on phonological features, would outperform the model presented here.

Table 5.20: Overall statistics for the confusion matrix in Table 5.19.

Overall statistics:
Accuracy : 0.8581
95% CI : (0.854, 0.8622)
No Information Rate : 0.3654
Kappa : 0.8063

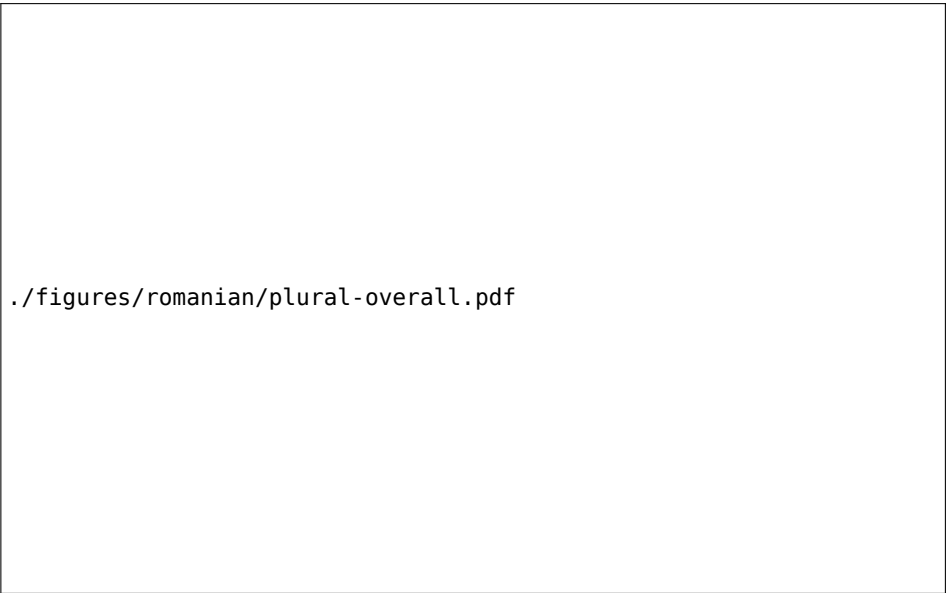


Figure 5.8: Additive (left) and subtractive (right) accuracy and kappa scores for the model predicting plural in Romanian.

5.2.4.4 Inflection class

Finally, we turn to the prediction of inflection class. Again, there are two possibilities we want to look at. First, we predict inflection class without making gender distinctions, i.e. if class $e-e$ is found in feminine and neuter, we assume this is a single class and not two different classes. We use the formula as before: $\text{class} \sim \text{final.1} + \text{final.2} + \text{final.3} + \text{n_vowels}$. The results of this model can be seen in Table 5.9 and the corresponding statistics in Table 5.21.

./figures/romanian/class-1-cm.pdf

Figure 5.9: Heat map for the model predicting inflection class in Romanian.

Table 5.21: Overall statistics for the heat map in Table 5.9.

Overall statistics:
Accuracy : 0.5577
95% CI : (0.5518, 0.5635)
No Information Rate : 0.1062
Kappa : 0.5121

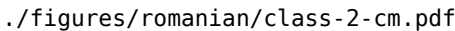
Table 5.21 shows that the model performs worse than the model predicting singular, but better (according to kappa score) than the model predicting only plurals. From the heat map in Table 5.9 it is clear that there is a high degree of confusion between the different classes, but it also looks like this confusion is not entirely random. The two strongest clusters of confusion are between classes with a *-C* singular marker, and between classes with a *-u* singular marker. If we perform cluster analysis on the corresponding similarity model, we get the results in Table 5.10. In this figure, I have additionally indicated the gender information for the inflection class for convenience, but the model itself had no information about gender. What can be seen from the clustering is that, although there is organization along marker lines, the strongest clustering effect is that of gender. Additionally, whenever masculine classes cluster together with neuter classes, these share the same singular marker, and masculine only classes seem to only cluster with neuter classes.



Figure 5.10: Clustering analysis of singular-plural class in Romanian.

Next, inflection classes are divided by gender, so that the five classes in the dataset which are ambiguous for gender are split into individual classes (one for

each gender). For this model, the results are presented in Table 5.11 and the corresponding statistics in Table 5.22. There is practically no difference in accuracy between both models. The clustering for this model is shown in Table 5.12. This clustering reflects almost exactly the hierarchy in Table 5.5 on page 90. Most clusters are found within a single same gender exclusively (clusters in light brown, light yellow, blue and dark gray), feminine and neuter, or masculine and neuter. Particularly clear are clusters where neuter and masculine share the same singular marker (clusters in pink, dark brown and light grey), or the feminine and neuter share the same plural marker (the cluster in green). The only cluster including the three genders has the classes with singular markers *-e* and *-ă*. Marker *-e* is the only one connected to all three genders in Table 5.5, while *-ă* is the only marker shared by feminine and masculine genders.



./figures/romanian/class-2-cm.pdf

Figure 5.11: Heat map for the model predicting inflection class by gender in Romanian.

Romanian plural markers are strongly predictable from phonological features of nouns. The models presented here are a strong computational validation of Vrabie (1989) and Vrabie (2000). The model by Bateman & Polinsky (2010) is also partially supported in the sense that we see strong evidence for four agreement classes. But the model presented in this section also refutes Bateman & Polinsky

Table 5.22: Overall statistics for the heat map in Table 5.11.

Overall statistics:
Accuracy : 0.5546
95% CI : (0.5488, 0.5605)
No Information Rate : 0.1062
Kappa : 0.5092

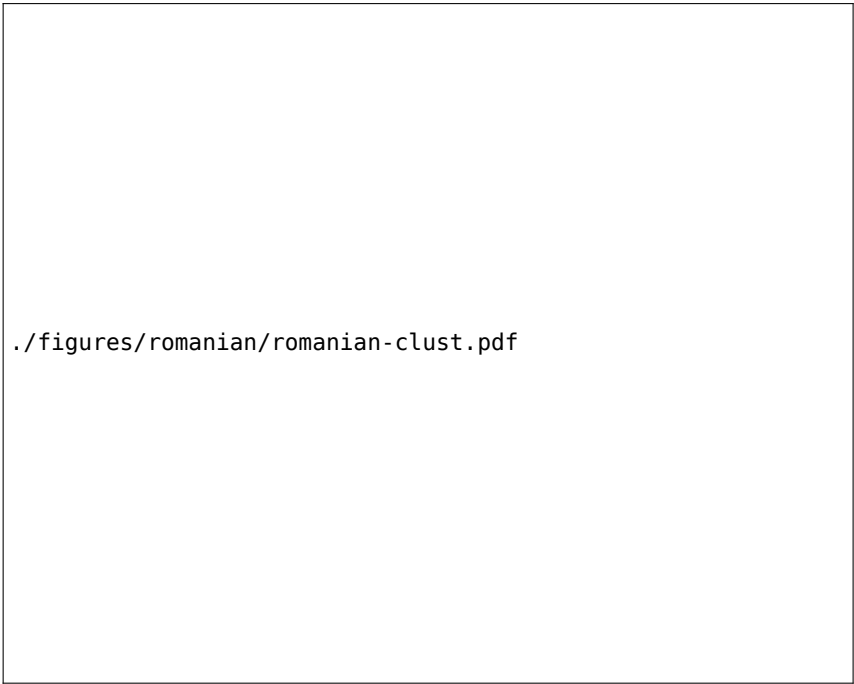


Figure 5.12: Clustering analysis of gender and class in Romanian.

in that there is evidence for a gender-number interaction. More precisely, there is evidence that inflection classes are partially dependent on gender, and that gender is predictive of plural, even when phonological features are considered. Most importantly, we do not see any evidence for a flat inflection class hierarchy, nor for the more symmetric hierarchy in Table 5.4 or the simpler hierarchy in Table 5.6, but we do see evidence for the hierarchy in Table 5.5, where inflection classes are partially conditioned by their gender alignment.

5.3 Interim conclusion

In this chapter I have presented two cases of gender-inflection class interactions, namely nouns from the Latin third declension and Romanian nouns. In Latin, we saw a relatively simple system where syncretisms in the inflection of nouns are conditioned by their gender. The Latin case could be modeled with a very simple tree clearly reflected in the analogical system. The nouns in Romanian presented a much more complex interaction between gender and inflection class. Therefore, a much more elaborate hierarchy had to be postulated. Still, I showed that the analogical model was helpful in distinguishing between the three alternatives.

With regards to the overall question of this book, in both cases we clearly saw that there are reflexes of the hierarchical structure in the analogical relations between the different classes.

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. Sub-lexical 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, 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-Probis-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, 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.), *Wariacja 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/, 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, 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, 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

