Analogical classification in formal grammar

Matías Guzmán Naranjo

Empirically Oriented Theoretical Morphology and Syntax 5

./langsci/graphics/lan

Empirically Oriented Theoretical Morphology and Syntax

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. Bowern, 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

ISBN: 978-3-96110-186-3 (Digital)
978-3-96110-187-0 (Hardcover)

ISSN: 2366-3529
DOI: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: XeLaTeX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

./langsci/graphics/storagelogo.pdf

Contents

7 Morphological processes and analogy

So far we have seen how the analogical relations between nouns reflect the grammatical structuring and type system of the lexicon. A common trait in the previous cases is that the morphological markers have all been suffixes. We also saw that it was only the ending of the stems (and some additional phonological information like the number of syllables and stress placement) that helped as predictors. This kind of correlation is often found in the literature on phonologically conditioned morphology and analogy in general. There are only a handful of studies in which the beginning of words were found to have a conditioning effect on some morphological process (Bybee & Slobin 1982; Köpcke & Zubin 1984), and studies that examine prefixes are even rarer.

Some well-known phenomena in phono-syntax suggest that this relation might not be coincidental. The choice between *a* and *an* in English, or the choice between *la* and *el* in Spanish (in Spanish feminine nouns can use the masculine definite article *el* if they begin with a stressed /a/, see Harris 1987), are conditioned by the first segment of the following word. This makes intuitive sense, but it is not obvious why it should be the case. It would be perfectly possible that suffix selection depended on the first segment of the stem, or the second vowel, etc.

To explore this question I look at three different phenomena in this chapter: Swahili noun classes, Otomi verb classes and Hausa plurals. Swahili and Otomi are relevant to the overall question of this chapter because they use prefixes instead of suffixes, and Hausa has complex plural formations.

7.1 Prefixes and gender: Swahili noun classes

Swahili, like other Bantu languages, has a noun class system in which all nouns belong to a specific, partially conditioned, class. Traditional Swahili grammars list eleven main classes for Swahili nouns, which are presented in Table 7.1.[1] These classes are defined by a prefix on the noun and can mark either singular or plural.

[1] I have omitted classes 14 (abstractions), 15 (verbal infinitives) and 16-18 (locatives). For classes 9 and 10, *N* represents three possible markers: *n-*, *ny-* or *m-*.

For the most part, noun classes are lexically determined, with a few classes being determined by derivational morphemes (diminutives, etc.).

Table 7.1: Swahili noun classes.

class	form	number
1	m-	singular
2	wa-	plural
3	m-	singular
4	mi-	plural
5	∅ ~ ji-	singular
6	ma-	plural
7	ki-	singular
8	vi-	plural
9	N-	singular
10	N-	plural
11	u-	singular

Corbett (1991), however, suggests that Swahili noun classes should be treated as genders, not very differently from other gender systems. The reason is that all the properties of a gender system are present in the Swahili class system, like agreement with determiners and adjectives as shown in (1).[2]

(1) *ki*-kapy *ki*-kubwa *ki*-moja *ki*-lianguka
CL7-basket CL7-large CL7-one CL7-fell
'One large basket fell.'

The class marker *ki* agrees with the verb, noun, adjective and determiner, just like German adjectives agree with their nouns. The fact that these are genders can be seen more clearly from cases where the prefix on a noun is 'wrong', in the sense that it usually denotes some other class than what it is actually agreeing with. In (2b) (Corbett 1991: 45) we see for example (a) that *tu* ('person') takes a marker for class 1, while the agreement with the verb is the marker of class 2. A similar situation arises in example (b) where there seems to be a disagreement between the different markers. For this reason Corbett (1991) argues that there are two different system: inflection class and gender proper.

[2]The examples in this section are taken from Corbett (1991), who in turn takes them from Welmers (1973: 159-183).

(2) a. *m*-tu *wa*-mepotea
CL1-person CL2-is.missing
'A parson is missing.'

b. *ki*-faru *m*-dogo *wa*-likuwa hapa
CL7-rhinoceros CL1-small CL2-was here
'A small rhinoceros was here.'

Thus, grouping the singular and plural forms we get the six genders (the original proposal in Corbett (1991: 47) suggests seven) in Table 7.2.

Table 7.2: Swahili genders.

Class	Prefix on noun	Verbal agreement
1/2	m-/wa-	a-/wa-
3/4	m-/mi-	u-/i-
5/6	∅ ~ ji-/ma-	li-/ya-
7/8	ki-/vi-	ki-/vi-
9/10	N-/N-	i-/zi-
11/10	u-/N-	u-/zi-

Swahili has received some attention with respect to how nouns are assigned to a given gender. Corbett (1991: 47) suggests that "for Swahili we require semantic and morphological assignment rules". The author lists (p. 47) the following rules (adapted) to account for how nouns are assigned to their gender class in Swahili. When in conflict, the semantic rules override the morphological rules:
Semantic assignment:

1. augmentatives belong to gender 5/6
2. diminutives belong to gender 7/8
3. remaining animates belong to gender 1/2

Morphological assignment:

1. morphological class 3/4 (m-/mi-) ⟶ gender 3/4
2. morphological class 5/6 (∅ ~ ji-/ma-) ⟶ gender 5/6

3. morphological class 7/8 (ki-/vi-) ⟶ gender 7/8
4. morphological class 9/10 (N-/N-) ⟶ gender 9/10
5. morphological class 11/10 (u-/N-) ⟶ gender 11/10

Corbett (1991: 48) also provides some additional semantic regularities: plants are often in gender 3/4, fruits in gender 5/6, animals in gender 9/10 and small objects in gender 7/8. This list is further expanded by Contini-Morava (1994), who provides strong additional semantic grounding for most of the six genders.

With all these rules combined, we have a system where we expect that phonological analogies will be rather weak. Because of its heavy semantic component, and because speakers are usually quite certain with regards to inflectional class assignment upon encountering a noun, the need for analogical relations is greatly reduced.

7.1.1 Materials

I compiled a list of Swahili nouns with their corresponding classes by combining the list given in the Wiktionary page for Swahili (Wikimedia Foundation 2016), and extracting all the nouns for which class information is available in the *Mgombato: Digo-English-Swahili Dictionary* (Mwalonya et al. 2004). Because the extraction from the Swahili dictionary relied on optical character recognition, there is some degree of noise in the data. I removed all clear errors of nouns containing punctuation marks. The result is 3081 nouns, distributed as shown in Figure 7.1. There were not enough *u-* marked nouns to properly work with the 11/10 gender.

Because the classes are uneven in terms of members, models including the whole data-set tended to under-perform[3]. To control for this, I randomly extracted 378 nouns for each class (the size of the smallest class in the original data-set). This produced a final data-set with 1890 nouns in total.

In terms of pre-processing, Swahili has a series of digraphs (e.g. *mb* ⟶ /^{m}b/), which I converted into single character representations to aid the analogical model. Otherwise, this is a relatively poor data-set in terms of features. We do not have any extra semantic or morphological information to aid the models.

[3] The reason is that the neural network models are sensitive to type frequency. This is not very important if the predictors are strong enough, but in cases where the predictors are weak, the model tries to optimize for general accuracy, and over-predicts the most frequent class.

Figure 7.1: Type frequency of Swahili genders.

7.1.2 Results

In our first model we investigate whether the first and second segments of the stem (that is, after removing the class prefixes) can predict to any degree the inflectional class of Swahili nouns with the model `class ~ first.1 + first.2`.[4] The results, shown in Table 7.3 and Table 7.4, are not very good in themselves. The accuracy is barely above chance, and the kappa score is very small. This basically means that there is very little information about inflection class just in the phonological shape of the stem. But this result is not really surprising. Swahili speakers encounter nouns with the prefix or some agreeing forms, and there is little ambiguity about their class.

Next, we compare this model to one where we use the endings of the nouns instead of the initial segments, as shown in Table 7.5. In this model we see performance at chance level.

Finally, we try a model that combines the first two segments of the noun, the last segment, and length in letters with the formula: `class ~ final.1 + first.1 + first.2 + length`. The results are presented in Table 7.7 and Table 7.8. This model shows a slight improvement from the model only using the first segments.

[4] With 0 hidden nodes and a decay rate of 0.1. A more complex model with interactions did not perform any better.

Table 7.3: Confusion Matrix for the model predicting inflection class of Swahili nouns.

	Reference				
Prediction	1-2	3-4	5-6	7-8	9-10
1-2	155	96	47	69	46
3-4	85	130	48	78	63
5-6	44	49	168	84	74
7-8	44	53	46	92	49
9-10	50	50	69	55	146

Table 7.4: Overall statistics for Confusion Matrix in Table 7.3.

Overall statistics:
Accuracy: 0.3656
95% CI: (0.3439, 0.3878)
No Information Rate: 0.2
Kappa: 0.2

Table 7.5: Confusion Matrix for the model predicting inflection class of Swahili nouns.

	Reference				
Prediction	1-2	3-4	5-6	7-8	9-10
1-2	195	94	92	89	102
3-4	35	91	71	79	43
5-6	32	49	54	40	58
7-8	31	68	67	91	54
9-10	85	76	94	79	121

Table 7.6: Overall statistics for Confusion Matrix in Table 7.5.

Overall statistics:
Accuracy: 0.2921
95% CI: (0.2716, 0.3131)
No Information Rate: 0.2
Kappa: 0.1151

Table 7.7: Confusion Matrix for the model predicting inflection class of Swahili nouns.

	Reference				
Prediction	1-2	3-4	5-6	7-8	9-10
1-2	178	83	49	42	73
3-4	68	158	47	86	60
5-6	44	43	164	91	58
7-8	25	55	56	105	40
9-10	63	39	62	54	147

Table 7.8: Overall statistics for Confusion Matrix in Table 7.7.

Overall statistics:
Accuracy: 0.3979
95% CI: (0.3757, 0.4204)
No Information Rate: 0.2
Kappa: 0.2474

The overall evaluation of this final model can be seen in Figure 7.2. This figure basically shows that the main effect comes from the first segment, but that the other factors still play a minor role.

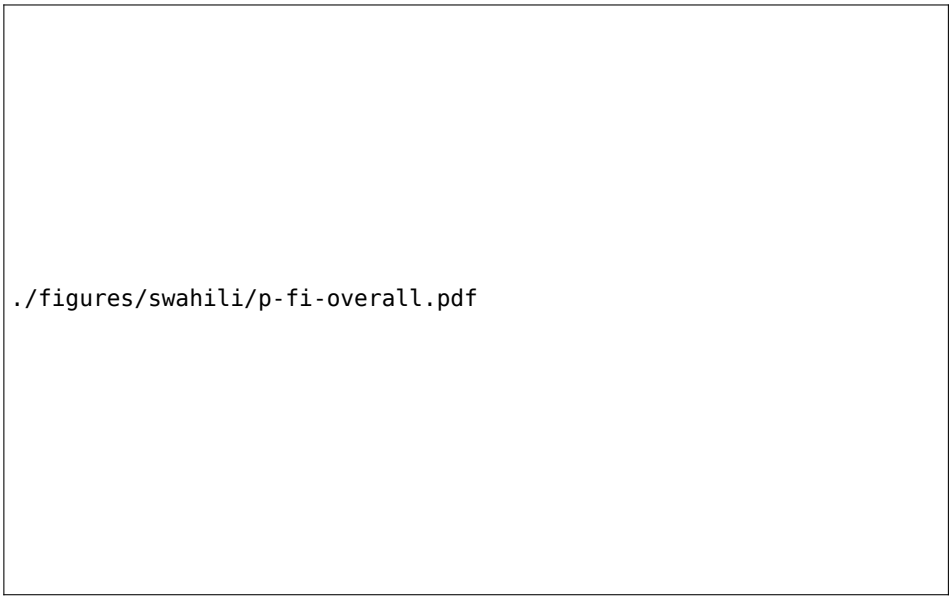

Figure 7.2: Additive (left) and subtractive (right) accuracy and kappa scores for the model predicting gender in Swahili.

The model including both beginning and ending of the nouns clearly performed better, and even though the main effect came from the beginning of the nouns, the ending did play a role.

It is possible that the current analogical relations of the Swahili noun classes are the product of some previous more regular system (Nurse & Hinnebusch 1993), and not of actual productive schemas speakers use. Because the analogical effects are so weak, the most likely explanation in this case is that the semantic component is much stronger, and thus phonological analogy is not as important for speakers. The important point here is that we do see a stronger effect of the beginning of the stem than of the ending of the stem.

7.2 Prefixes and inflection classes: Eastern Highland Otomi

7.2.1 Verb classes in Eastern Highland Otomi

Eastern Highland Otomi (Otomi from now on) is a Mesoamerican language of the Otomanguean family spoken in Mexico (Echegoyen & Voigtlander 1979). The Otomi verb system is relevant for the proposal in this book because, like in Swahili, it has inflection classes where the actual inflection is produced by a prefix instead of a suffix.

The verbs are organized in four classes according to Echegoyen & Voigtlander (1979), and five classes according to Feist & Palancar (2015). Examples of these classes can be found in Table 7.9.

Table 7.9: Otomi inflection classes.

		Class I.a 'gather'	Class I.b 'save'	Class II 'walk'	Class III 'fix'	Class IV 'hurry'
Incompletive	1st	dí joni	dí -n yäni	dí 'yo	dí -dí hoki	dí -dí xøni
	2nd	gí joni	gí -n yäni	gí 'yo	gí -dí hoki	gí -dí xøni
	3rd	(i) joni	i -n yäni	(i) 'yo	(i) -di hoki	(i) -di xøni
Imperfect	1st	dmí joni	dmí -n yäni	dmí 'yo	dmí -dí hoki	dmí -dí xøni
	2nd	gmí joni	gmí -n yäni	gmí 'yo	gmí -dí hoki	gmí -dí xøni
	3rd	mí joni	mí -n yäni	mí 'yo	mí -dí hoki	mí -dí xøni
Completive	1st	dá joni	dá yäni	dá -n 'yo	dá hoki	dá -n xøni
	2nd	gá joni	gá yäni	gá -n 'yo	gá hoki	gá -n xøni
	3rd	bi goni	bi yäni	bi -n 'yo	bi hoki	bi -n xøni
Perfect	1st	xtá joni	xtá yäni	xtá -n 'yo	xtá hoki	xtá -n xøni
	2nd	xká joni	xká yäni	xká -n 'yo	xká hoki	xká -n xøni
	3rd	xø-n goni	xø -n yäni	xø -n 'yo	xø hoki	xø -n xøni
Pluperfect	1st	xtá joni	xtá yäni	xtá -n 'yo	xtá hoki	xtá -n xøni
	2nd	xkí joni	xkí yäni	xkí -n 'yo	xkí hoki	xkí -n xøni
	3rd	xí goni	xí yäni	xí -n 'yo	xí hoki	xí -n xøni
Irrealis	1st	ga joni	ga -n yäni	da -n 'yo	ga hoki	da -n xøni
	2nd	gi joni	gi -n yäni	ga -n 'yo	gi hoki	ga -n xøni
	3rd	da goni	da yäni	di -n 'yo	da hoki	di -n xøni

Capturing the class system in Otomi requires positing five independent types, but nonetheless there is a degree of organization between these types. The important thing to observe here is that classes *III* and *IV* share an extra *-di-* segment

in the incompletive and imperfect, while classes *I* and *II* do not have this feature. Meanwhile, class *II* and class *IV* share the use of an extra *-n* in the completive, perfect, pluperfect and irrealis. Class *I.a* can either be grouped with classes *I.b* and *III* or as a completely independent class, depending on the property involved.

7.2.2 Materials

For this case study I used the inflection class database by Feist & Palancar (2015) (based on Echegoyen & Voigtlander 1979, Echegoyen & Voigtlander 2007 and Voigtlander & Echegoyen 2007). This database contains 1998 verbs, all of which were analyzed and assigned to one of the five classes. It also contains information about whether the verb is transitive or not, its stem and citation form. I performed no extra processing on the data and used it as it was.

7.2.3 Results

In terms of complexity, the model for Otomi is probably the one with the most factors. As predictors, I included the first three segments (with an interaction between the first and second segment), the last two segments, the tone of the citation form, and whether the verb is transitive or not: `class ~ first.1 * first.2 + first.3 + Transitivity + last.1 + last.2 + tone`.[5] The confusion matrix for this model is shown in Table 7.10 and the accuracy measures in Table 7.11.

Table 7.10: Confusion matrix for the model predicting inflection class in Otomi.

	Reference				
Prediction	Ia	Ib	II	III	IV
Ia	609	6	46	141	56
Ib	6	29	2	8	0
II	50	2	284	27	85
III	82	15	10	249	14
IV	36	3	74	28	136

We see that classes are mostly predictable for Otomi, but there is some degree of confusion. The accuracy metrics show that *class-Ia* is receiving most of the miss-classifications, which is to be expected, this being the most frequent class.

[5] The model contained no hidden nodes and a decay rate of 0.1.

Interestingly, *class-Ib* is only mildly confused with *class-Ia*, and much more confused with *class-III*.

Table 7.11: Accuracy scores for Table 7.10.

Overall Statistics					
Accuracy : 0.6542					
95% CI : (0.6328, 0.675)					
No Information Rate : 0.3919					
Kappa : 0.5211					
Statistics by Class:					
	Class: Ia	Class: Ib	Class: II	Class: III	Class: IV
Sensitivity	0.7778	0.52727	0.6827	0.5497	0.46735
Specificity	0.7951	0.99177	0.8963	0.9217	0.91740
Neg Pred Value	0.8474	0.98669	0.9148	0.8747	0.90994
Balanced Accuracy	0.7864	0.75952	0.7895	0.7357	0.69238

The important fact regarding Otomi is the relative effects of the different factors. In Swahili we saw that both the first segments and final segment of the nouns carried some information about gender. In this case, we have more or less the same situation. Figure 7.3 shows the additive and subtractive model evaluation plots. On the left, we see that all factors used provide small increases to model performance. Moreover, on the right, we see that the two most important factors were the interaction between the first two segments of the verb and the verb's transitivity. The interesting thing to note is that the first segments were much more important for predicting inflection class than the final segments.

Once more, classes that trigger prefixing processes are predictable from analogies based on the beginning of words, much more so than analogies based on the endings. The fact that the endings did play a minor role is interesting. It probably means that both Otomi and Swahili are susceptible to word size schemas, similar to how in German nouns gender is determined by both initial and final segments (Köpcke & Zubin 1984).

Figure 7.3: Additive (left) and subtractive (right) accuracy and kappa scores for the model predicting inflection class.

7.3 Stem changing processes: Hausa plural classes

7.3.1 The Hausa plural system

The Hausa plural system is too complex to be fully explored here, but some of its properties are relevant to the overall theme of this chapter. First, there seems to be little agreement with regards to how many plural classes there are in Hausa, and an analysis could go anywhere between "many" (Migeod 1914), around thirty (Schön 1862), to twenty macro-classes (Newman 2000), or the many more sub-classes Newman identifies. For this study I follow the macro-classes defined by Newman (2000), which are given in Table 7.12[6].

As we can observe in Table 7.12, some plural classes assign their own tonal pattern to the plural forms, independently of the tonal patterns of the singular, while others carry over the tonal pattern of the singular class (Newman 2000: 430). There are several reduplication patterns, and several 'broken' plurals, where there is a vocalic change before and after the final consonant of the singular. It is worth keeping in mind that these are macro-classes, and one could find an

[6]Because the dictionary I use for the data (Bargery & Westermann 1951) does not distinguish between the retroflex and rolled *r*, and between long and short vowels, I will not mark these features here. For tone representation I follow Newman (2000), with high tone unmarked, low tone marked with a grave accent, and falling tone with a circumflex accent.

Table 7.12: Hausa plural macro-classes.

Class	Singular	Plural	Gloss
a-a	sirdì	siràd**a**	'saddle'
a-e	gulbi	gulàb**e**	'stream'
a-u	kurmì	kuràm**u**	'grove'
-aCe	wuri	wur**àre**	'place'
-ai	malàm	malàm**ai**	'teacher'
-anni	watà	wàt**ànni**	'moon'
-awa	talàkà	talak**awa**	'commoner'
-aye	zomo	zom**àye**	'hare'
-Ca	tabò	tab**ba**	'scar'
-Cai	tudù	tù**ddai**	'high ground'
-ce2	ciwò	cìwà**ce-cìwàce**	'illness'
-Cuna	cikì	cik**kunà**	'belly'
-e2	camfì	càmf**e-càmfe**	'superstition'
-i	tàurarò	tàuràr**i**	'star'
-oCi	tagà	tag**ogi**	'window'
-u	kujèra	kùjèr**u**	'chair'
u-a	cokàli	cok**uà**	'spoon'
-uka	layì	lay**ukà**	'lane'
-una	rìga	rig**unà**	'grown'
X2	àkàwu	àkàwu-**àkàwu**	'clerk'

even more fine-grained division, with many subdivisions within each of these classes. Because of this fact, there are no good arguments in favor of a specific hierarchical organization of these classes.

Newman (2000: chapter 56) observes several regularities in the formation of plurals. He mentions, for example, that *-aCe* plurals only occur with CVCVV nouns, while *a-a* plurals tend to appear with CVCCVV nouns (p. 431). Newman gives similar patterns for other macro-classes, but states that ultimately Hausa plurals are not fully predictable.

7.3.2 Materials

I extracted all nouns from *A Hausa-English Dictionary and English-Hausa Vocabulary* by Bargery & Westermann (1951). The dictionary contains around 3000

nouns, of which only some 1450 have a plural. Of these, quite a few have indications about multiple alternatives. Some of the alternatives are marked as *rare*, or for regional preferences. It is not really possible to work with these overabundant variants (Migeod 1914; Salim 1981; Newman 2000) because there are just not enough of them (around 150). As a practical solution, I simply took the first variant offered in the dictionary and ignored the rest. Similarly, in cases where the dictionary offered multiple possible singulars for a noun, I only used the first singular form listed.

Identifying plural classes automatically in Hausa is not a trivial task, and it is not completely clear how many examples fit into Newman (2000)'s macro-classes. I followed the definitions as given in Table 7.12. Although this approach is likely to produce some errors, it should mostly give us the right classification. The main difference in the classes I use is that I take four reduplication classes instead of the three listed in Table 7.12: *class-RED-e* and *class-RED-comp* correspond to the *class-X2* and *class-e2* classes identified by Newman (2000). I included *class-RED-id* which consists of cases where the plural is the reduplication of the singular form without additional changes, and a general *class-RED* class with all the cases that do not quite fit into any of the other classes. The *class-ce2* did not have enough members to be usable. Finally, an extra class I include is *class-oi*, which is not explicitly mentioned by Newman (2000), but which had enough members to be distinguished as an independent macro-class. We can see the frequencies of the classes in the data-set in Figure 7.4.

As expected, some classes are considerably more frequent than others, and the general distribution is roughly zipfian. However, it is hard to tell which of these classes are productive, which are irregulars, and which misanalyses.

A serious shortcoming of this database is the lack of information about vowel length. According to Newman (p.c.), several of the macro-classes are strongly correlated with vowel length of the singular, which means there is an important factor missing.

7.3.3 Results

First we look at a model predicting the plural class from structurally defined predictors. Since most of the macro-classes presented in Figure 7.4 are defined by two vowels and a potential consonant between them, I defined the predictors as follows: `plural class ~ V.1*T.1 + C.1 + V.2 + CVCV.4 + length`.[7] Here,

[7] The model included one hidden layer with five nodes and a decay rate of 0.1. Gender did not play a significant role in any of the models.

Figure 7.4: Type frequency of macro-classes in Hausa.

`V.1` and `V.2` are the final and prefinal vowels, respectively, `C.1` is the final consonant, `T.1` is the final tone of the singular, `length` the length in letters, and `CVCV.4` is the CV structure of the final four segments of the singular. In this case, we are specifying an interaction between the final vowel and the tone of that vowel. Newman (2000: chapter 56) makes reference to all these factors, in some way or another, in his analysis of the Hausa plurals. It is therefor no surprise that they play a role in the analogical model.

The results of this model can be seen in Figure 7.5 and the corresponding statistics are presented in Table 7.13. We see that most classes can be predicted to a relatively high degree of accuracy. There is a clear darker trace along the main diagonal in Figure 7.5, but with some noise for most classes.[8] In the table there are errors across most classes with no clear structure to them, besides some apparent foci (*class–a-a*, *class–a-e*, *class–ai*, *class–Cai* and *class–oCi*). The accuracy statistics do reveal that the model is performing well above chance, and that there is a significant analogical relation between these classes.

For comparison, a model that does not specify structural analogy: `plural class ~ final.1*T.1 + final.2 + final.3 + CVCV.4 + length`[9], can be seen

[8] Because the numbers used for shading are log scaled from the actual confusion matrix, the error rates appear slightly higher than they actually are.

[9] The model included no hidden nodes and a decay rate of 0.1.

in Table 7.14. It is not surprising that this model also performs relatively well, after all, the predictor `final.1` captures the same information as the predictor `V.1`.

Figure 7.5: Heatmap for the model predicting plural forms in Hausa.

Table 7.13: Accuracy scores for Figure 7.5.

Overall Statistics
Accuracy : 0.5425
95% CI : (0.5161, 0.5686)
No Information Rate : 0.2082
Kappa : 0.488

Table 7.14: Accuracy scores for the non-structurally defined model.

Overall Statistics
Accuracy : 0.5057
95% CI : (0.5792, 0.5321)
No Information Rate : 0.2082
Kappa : 0.4516

We can compare model performance for both models (Figure 7.6 and Figure 7.7). These evaluations reveal that indeed `final.1` and `V.1` have more or less the same impact on the model, but for the non-structurally defined model all other predictors become rather insignificant in the subtractive evaluation. The segments captured by both models are the same, but the additional structure does clearly play a role.

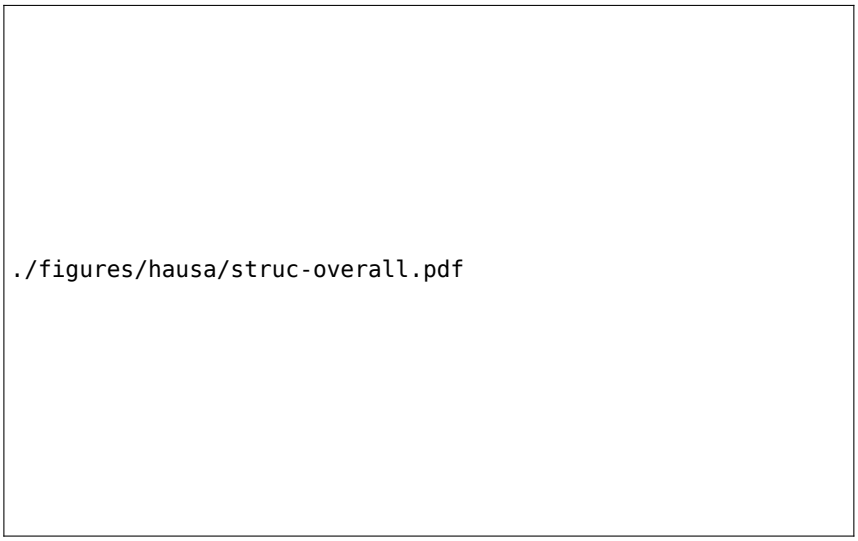

Figure 7.6: Additive (left) and subtractive (right) accuracy and kappa scores for the structurally defined model.

We can also see that the more structural predictors not only achieve a higher accuracy, but also have more independent weights (higher in accuracy in the subtractive evaluations). The main factors are clearly the vowels (and their interaction with tone), while the consonant has less influence. This strongly matches the broken plurals we see in Hausa, where the consonant remains stable and the vowels before and after it are changed.

7.4 Interim conclusion

In this chapter I have provided some evidence for a different aspect of analogical models, namely the fact that the analogical specifications, or the points where the analogy takes place, can be related to the actual morphological process. In Swahili and Otomi we see that a prefixing system triggers analogy mostly at the

Figure 7.7: Additive (left) and subtractive (right) accuracy and kappa scores for the non-structurally defined model.

beginning of words, and in Hausa we see how the analogical relation requires a specification that is similar to the actual structure of some plural classes. The results of this chapter should be taken only as a starting point. Two languages for prefixes is too small a sample to draw any definitive conclusions. As mentioned in Part I, this problem had already been raised before:

> The problem faced in the full elaboration of such models, however, is in specifying the relevant features upon which similarity is measured. This is a pressing empirical problem. We need to ask, why are the final consonants of the strong verbs more important than the initial ones? (Bybee 2010: 62)

This observation is very difficult to explain from a formal perspective. Assuming the model introduced in Part I is right, there is no way for the hierarchy to 'know' what kind of morphological process is being carried out on the different types, and to link that to the inheritance of analogical constraints. From a usage-based perspective, however, these results make more sense. A potential explanation is that speakers are more focused on finding similarities between words where the important changes happen, i.e., the segments before a suffix or after a prefix. This would also explain why there seems to be a distance effect from the edge in most of the other languages, that is, the very last segment tends to be more important than the second to last and so on (though not always).

A possible advantage of this explanation is that it also helps reduce the search space for speakers. Unless there was some innate constraint that specified where to look for analogies, speakers would have to analogize over all segments of all stems. The fact that analogies seem to be mostly constrained to the edge of the stem where the morphological process happens, helps reduce the amount of information that has to be considered. This variability of the 'where' of the analogy is an advantage for speakers of the language and not a drawback.

References

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89(3). 429–464.

Ackerman, Farrell & Robert Malouf. 2016. Word and pattern morphology: An information-theoretic approach. *Word Structure* 9(2). 125–131.

Afonso, Olivia, Alberto Domínguez, Carlos J. Álvarez & David Morales. 2014. Sublexical and lexico-syntactic factors in gender access in Spanish. *Journal of Psycholinguistic Research* 43(1). 13–25.

Aguirre, Carmen & Wolfgang U. Dressler. 2008. On Spanish verb inflection. *Folia Linguistica* 40. 75–96.

Alber, Birgit. 2009. Past participles in Mocheno: Allomorphy and alignment. In Michael T. Putnam (ed.), *Studies on German-language islands*, 33–64. Amsterdam, Philadelphia: Benjamins.

Albright, Adam. 2008a. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Linguistic universals and language change*, 144–184. Oxford: Oxford University Press.

Albright, Adam. 2008b. How many grammars am I holding up? Discovering phonological differences between word classes. In *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 1–20. Somerville, MA: Cascadilla 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 morphopholonogical alternations. In *Typologie et modélisation des systèmes morphologiques*. Paris. http://www.llf.cnrs.fr/fr/node/5611.

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

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 Bowern & 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, María. 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.

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.
Farkaş, Donka F. 1990. Two cases of underspecification in morphology. *Linguistic Inquiry* 21(4). 539–550.
Farkaş, 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.

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 cantara 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.

Kempe, Vera, Patricia J. Brooks & Anatoliy 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*. Standford: 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.

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 *P roceedings 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.

Mlađenović, A. 1977. Neka pitanja varijantnosti norme u savremenom srpskohrvatskom književnom jeziku. In Stanisław Urbánczyk (ed.), *Wariancja normy we współczesnych słowiańskich językach literackich*, vol. 38 (Prace Komisji Słowianoznawstwa), 51–56. Kraków.

Morin, Regina. 2006. Spanish gender assignment in computer and Internet related loanwords. *Rivista di Linguistica* 18. 325–54.

Moscoso del Prado Martín, Fermín, Aleksandar Kostić & R. Harald Baayen. 2004. Putting the bits together: An information theoretical perspective on morphological processing. *Cognition* 94. 1–18.

Motsch, Wolfgang. 1977. Ein Plädoyer für die Beschreibung von Wortbildungen auf der Grundlage des Lexikons. In Herbert Ernst Brekle & Kastovsky Dieter (eds.), *Perspektiven der Wortbildungsforschung*, 180–202.

Müller, Stefan & Stephen Wechsler. 2014. Lexical approaches to argument structure. *Theoretical Linguistics* 40. 1–76.

Mürk, Harri William. 1997. *A handbook of Estonian: Nouns, adjectives and verbs*. Bloomington: Indiana University, Research Institute for Inner Asia Studies.

Murtagh, Fionn & Pierre Legendre. 2014. Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification* 31(3). 274–295.

Mwalonya, Joseph, Alison Nicolle, Steve Nicolle & Juma Zimbu. 2004. *Mgombato: Digo-English-Swahili Dictionary*. Kwale: Digo Language and Literacy Project.
Naden, Tony. 1988. The Gur languages. *The languages of Ghana* 2. 12–49.
Naden, Tony. 1989. Gur. In John Bendor-Samuel (ed.), *The Niger-Congo languages*, 141–168. New York: University Press of America.
Nastase, Vivi & Marius Popescu. 2009. What's in a name? In some languages, grammatical gender. In *Conference on empirical methods in natural language processing*, 1368–1377. Singapore: ACL and AFNLP.
Năvălici, Cristian. 2013. *PyDEX*. https://github.com/cristianav/PyDEX, accessed 2016-6-15.
Neuvel, Sylvain. 2001. Pattern analogy vs. word-internal syntactic structure in West-Greenlandic: Towards a functional definition of morphology. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2000*, 253–278. Amsterdam: Springer.
Nevins, Andrew. 2011. Phonologically conditioned allomorph selection. In Colin Ewen, Elizabeth Hume, Marc Van Oostendorp & Keren Rice (eds.), *The companion to phonology*, 2357–2382. London & New York: Continuum.
Newcombe, Robert G. 1998. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Stat. Med. *Statistics in Medicine*. 857–72.
Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar*. New Haven: Yale University Press.
Niggli, Idda & Urs Niggli. 2007. *Dictionaire bilingue Kasım-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.

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-*

gewälhlte Tropfen aus verschiedenen Lagen der spanischen Sprach-und Übersetzungswissenfschaft Fetschrift, 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.

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 Śākaṭāyana: 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 Hage: 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.

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

Subject index

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 commonis 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 regularties, 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