Analogical classification in formal grammar

Matías Guzmán Naranjo

Draft of May 23, 2019, 11:31

./langsci/graphics/la

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. Bowern, Claire, Laurence Horn & Raffaella Zanuttini (eds.). On looking into words (and beyond): Structures, Relations, Analyses

# Analogical classification in formal grammar

Matías Guzmán Naranjo

./langsci/graphics/storagelogo.pdf

Contents

Contents

4 Methodological notes

In this chapter I present, somewhat informally, the statistical and data visualization tools I will use in the rest of this book. Readers interested in the mathematical details should consult the cited sources, since detailed descriptions and explanations of these techniques would merit a short book and a strong statistical background. My intention is only to provide the reader with a good intuition of what these methods do.

4.1 On the general methodology

Many of the phenomena I discuss in the following case studies have been studied exhaustively before, and it is not my intention to develop complete analyses for any of these cases. Moreover, in several instances, I will use sub-optimal models that ignore semantics or other possible strong predictors, this should only make the main point stronger: formal analogy occurs even in unexpected cases, and it follows the grammatical hierarchy of the language. Similarly, I do not provide full formal linguistic analyses, but rather only sketches to motivate a plausible type hierarchy. It is my intention that the ideas proposed in this book can be formally implemented in different linguistic theories (Construction Grammar, Cognitive Grammar, Paradigm Function Morphology, HPSG or similar). This is why theoretical assumptions are kept to a minimum.

I make no strong claims concerning the psycho-linguistic reality to these models. The fact that we can predict, to a greater or lesser degree, word classes from formal properties of words, does not mean that speakers necessarily do the same. It is possible that the way speakers perform class assignment in some of the languages studied has some parallels to the models proposed here, but it is also possible that speakers do rely on different aspects of cognition. These are related but independent questions. The patterns I will present could be productive, or vestiges of previous systems, but not any less real. I will, however, make some connections with some ideas about cognitive aspects of language during the final discussion.

There are two main reasons for the choice of languages in this study: theoreti-

cal relevance and data availability. I have not tried to compile a representative typological sample. Far from it, most examples are from Indo-European languages of various subfamilies, and only a few are taken from African languages. The analogical models require an electronic, morphologically annotated dictionary, which are still very rare for languages spoken by smaller language communities. The theoretical relevance relates to the classes a language has and how they are organized.

4.2 Statistical models and methodology

For all the cases to follow I use the same general method for building the analogical models. From the stem of the words (nouns or verbs) I extract predictors which might be at play in the analogical relations and fit a neural network with the nnet package (Venables & Ripley 2002)[1] in R (R Development Core Team 2008). The use of neural networks has a long history in linguistics, and they are usually linked to connectionist models (Bechtel & Abrahamsen 2002; Churchland 1989; McClelland & Rumelhart 1986; Rumelhart & McClelland 1986a). However, I do not make any claims about the underlying linguistic system, or the rightness or wrongness of connectionism. The use of neural networks for the following analogical models is purely practical. Similar effects could probably be achieved using different algorithms like Random Forest (Breiman 2001) or Support Vector Machines (Smola & Schölkopf 1998; Scholkopf & Smola 2001). For the present book, the actual technology is not important, only the concept behind it[2]. My aim is to show that prediction is possible, not to find the best possible method.

The *stems* in the models are not theoretical objects, and the ideas in these models should be compatible with word-based models. The idea is that there is a distinction between the phonological material that expresses some property like MASCULINE, and the phonological material that expresses the meaning 'cat'. I take the *stem* to be the full word minus the phonological material that marks the category at hand. In a trivial Spanish example, the stem of *gato* ('male cat') is /gat/, since the /o/ segment is the gender correlate, and we have the opposing form *gata* ('female cat'). For many non-trivial cases some compromise had to

---

[1] For all models I used the softmax linking function and the maximum number of weights and iterations set in a way that the models converge. Whether a hidden layer was present or not, and the number of hidden units, varied from model to model.

[2] The next subsection provides a simple illustration of how the analogical models work, but the interested reader should consult Venables & Ripley (2002) for a rigorous mathematical explanation of neural networks and the nnet package.

be reached, and it will be described in detail. Crucially, this approach does not consider underlying representations, only surface forms. Of course, one could compare the results of a model based on some sort of underlying representation with the results of more surface oriented models.

The predictors used to fit the analogical models vary slightly from model to model, but they always contain phonological information about the shape of the word. The most straightforward way to do this is to simply take a set number of letters at the end (or beginning) of the stems and use them (together with their position) as predictors. What the positional information does is to make a distinction between, say, a *t* at the end of the stem and a *t* in the third to last position. This way of specifying phonological shape has advantages and disadvantages. A good aspect of this approach is that the model can, on its own, infer classes between phonemes as represented by letters. If *x* and *h* share some phonological feature which makes them into a natural class, and are thus predictive of the same inflection class or gender, the model will simply assign the corresponding weights to said inflection class. This means that we do not need a rich phonological representation to arrive at phonological analogies. Another possible issue, which might unfairly benefit or harm the models, is that in cases of low correspondence between the phonology and the orthography, certain spelling rules might contain some additional information not directly available to speakers, or some important information might be missing. There is no easy way to solve this problem, short of using detailed phonological transcriptions, which are unavailable for most of the languages under consideration. Any sort of phonological process like methathesis, which could be easily captured by a rule-based approach, will be invisible to the model, thus reducing the amount of information available. To reduce loss of information due to some spelling systems representing a single phoneme with a character sequence (e.g. Spanish and German), I simplified spelling assigning special characters to those regular sequences. Some phonological information is, however, non-recoverable from the orthography (e.g. some vowel length/quality information in German, or the difference between long and double vowels in Kasem).

To prevent overfitting[3] the models I apply ten-fold cross-validation to every model. This is done by splitting the dataset into ten groups. The general model is then fitted using nine of the groups as training data and testing the predictions of the model on the group not used for fitting it. The process is repeated for

---

[3]Overfitting happens when models predict the same items they learned from. This is a problem because if a model is overfitted, it does not really tell us much about how good the predictors are on novel items.

each of the ten subgroups. This way we can look at all the data while preventing overfitting (Kohavi 1995).

In section §2.2 I discussed four possible ways of implementing analogy, and argued that the difference is a gradient rather than a truly categorical one. The present models fall somewhere between a weighted multiple-rule-based model like those presented by Albright & Hayes (2003) and Albright (2009) and a purely stochastic model like NDL (Arppe et al. 2014; Baayen 2011; Baayen et al. 2011) or AM (Skousen 1989; Skousen et al. 2002; Skousen 1992; Arndt-Lappe 2011; 2014). The difference between the present model and a weighted rule-based model is that I consider all possible patterns within some structurally defined positions in the word (e.g. the last two segments, the last consonant, the number of syllables, etc.), and do not attempt to predefine the rules of the model, or decide to include or exclude some patterns. The difference to a completely stochastic system lies in the same property: the current model is sensitive to structural properties of the lexemes it sees, while NDL and AM are "blind" or completely amorphous. Like AM, and unlike some previous connectionist systems (Bechtel & Abrahamsen 2002; Churchland 1989; McClelland & Rumelhart 1986; Rumelhart & McClelland 1986a), the analogical model used here sees linguistically defined categories as the outputs. In traditional connectionist systems the networks directly paired semantics to sounds (Matthews 2005).

The similarities between these kinds of systems have been observed before:

> Connectionist networks themselves further illustrate the problem, in that they might be seen to fall in both camps. Back-propagation networks are often described as depending on similarity...However, they are also often described as using 'implicit rules' which can be extracted using appropriate analysis...Therefore, back-propagation networks appear rule- and similarity-based (Hahn & Chater 1998: p. 200)

In any case, it should be clear that I am not arguing for neural networks as a necessarily better implementation of analogical systems, or as a psychologically plausible system. Neural networks as used here are just one of the many alternatives we have to model analogy.

All this being said, a more clever and carefully designed model similar to the weighted multiple-rule-based model like those presented by Albright & Hayes (2003) and Albright (2009), or that of Beniamine & Bonami (2016), would probably perform better for any particular case and be more psychologically plausible. These models have some downsides, however. The most important one is that

they require much better structured datasets, with complete phonological information. This requirement is harder to fulfill than the rough, semi-phonemic transcriptions required for the neural network models. A second difficulty of these models is that they are extremely slow to fit because the rule inference step is computationally intensive. This makes it impossible in practice to fit many different models for each case and hence impractical to test various combinations of predictors.

4.3 Analogical models using neural networks

The easiest way to explain the intuition behind the models, and the tools I use for evaluating them, is with concrete examples. Suppose a language has two inflection classes, **A** and **B**. The dataset in (1) presents stems for lexemes belonging two said inflection classes.

(1)   a.   A: *lama, lara, lado, laso, pama, ra, dal, kar, tsar, sek, cess*
        b.   B: *egrr, liz, lo, loi, lu, lip, roop, oppe, toi, olor, gin, grip, wik*

There is no single (simple) rule which can predict to what class a given lexical item belongs. However, intuitively, the first vowel seems to be a strong indicator. All items for which the first vowel is *a* belong to class A, while items for which the first vowel is *i*, *o* or *u* belong to class B. Items with *e* are found in both groups. Because there are only a few lexical items, and the pattern is fairly simple, this generalization is fairly evident, but in a more complex system it would be much harder. These observations could also be inferred with a statistical model.

Put in simple terms, given a training dataset with items, and a series of predictors for each item, the neural network model learns from these items and assigns weights to the predictors. When presented with new data (the testing dataset), the network calculates from the weights the probability of each outcome for each item in the dataset. This is achieved the following way. The neural network sets a baseline for the prediction, based on one of the levels for each predictor. For each predictor, the first level (alphabetically) is chosen for the baseline node (in this case thus: `s1`=*c*, `s2`=*a*). This baseline has a weight for each outcome (the classes to be predicted). To each other level of each predictor, it assigns a weight for each outcome. The weights of the predictors are added to the baseline to calculate the probability of each outcome given some input.

We can apply this to our previous example. We split the data set into a training and a small testing dataset. For the testing dataset we randomly select the items:

lama, *lara*, *kar*, *egrr*, *liz*, *oppe*, *grip*, and we assign the rest of the items to the training dataset. For illustration, we can fit two different models. In Model 1, we set two predictors `s1` and `s2`, which correspond to the first and second letter in the items, respectively. For Model 2 we set `v1` and `c1`, which correspond to the first vowel and first consonant of the items, respectively. We can then train Model 1 and Model 2.

Figure 4.1 shows the structure of the neural network for Model 1. As can be seen, there is a direct connection between the predictors and the outcomes (a skip layer), and there are no intermediate steps (hidden layers). We can see how each letter predicts A or B. The thickness of the line represents the absolute value of the weight (thicker lines have larger absolute values), and the color represents whether the weight is positive (dark gray) and favors the outcome, or negative (light gray) and disfavours the outcome. In the node marked as `B1` we have the baseline (the bias node) made up of the combination of the levels *c* for `s1` and *a* for `s2`[4]. This combination clearly favours **A**, as we would expect. If any of these levels changes in the input, then the nodes in the skip layer activate and counteract the baseline. If, say, the input contains an *u*, *i* or *o* instead of an *a* in `s2`, then the corresponding node will strongly activate the outcome **B**, as we, again, would expect from the data set. The complete set of weights from the inputs to the outcomes for Model 1 is given in Table 4.1.

To calculate the actual class probabilities from the output weights, we use the softmax function. The intuition of this function is that, given a vector of weights, it will transform that vector into a vector of probabilities, where the element with the highest weight will receive the highest probability, and all probabilities add up to 1. The general form of this function is given in equation (4.2). In prose, we exponentiate each weight, and divide by the sum of all exponentiated weights.

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \tag{4.2}$$

As an example, assume the weight vector [2, 1, 0.1]. Exponentiating each member we get the vector [7.4, 2.7, 1.1], and their sum is 11.21. Dividing the exponentiated weights by the sum we get the probabilities [0.66, 0.24, 0.1].

To know how well the model performs, we predict the outcomes of the testing dataset, build a confusion matrix and calculate different accuracy scores. The corresponding confusion matrix for this model is shown in Table 4.3. Here we see the predictions that the model made for each testing item. There were two

---

[4]The models chooses the baseline levels purely on alphabetical order.

./figures/fake/model1.pdf
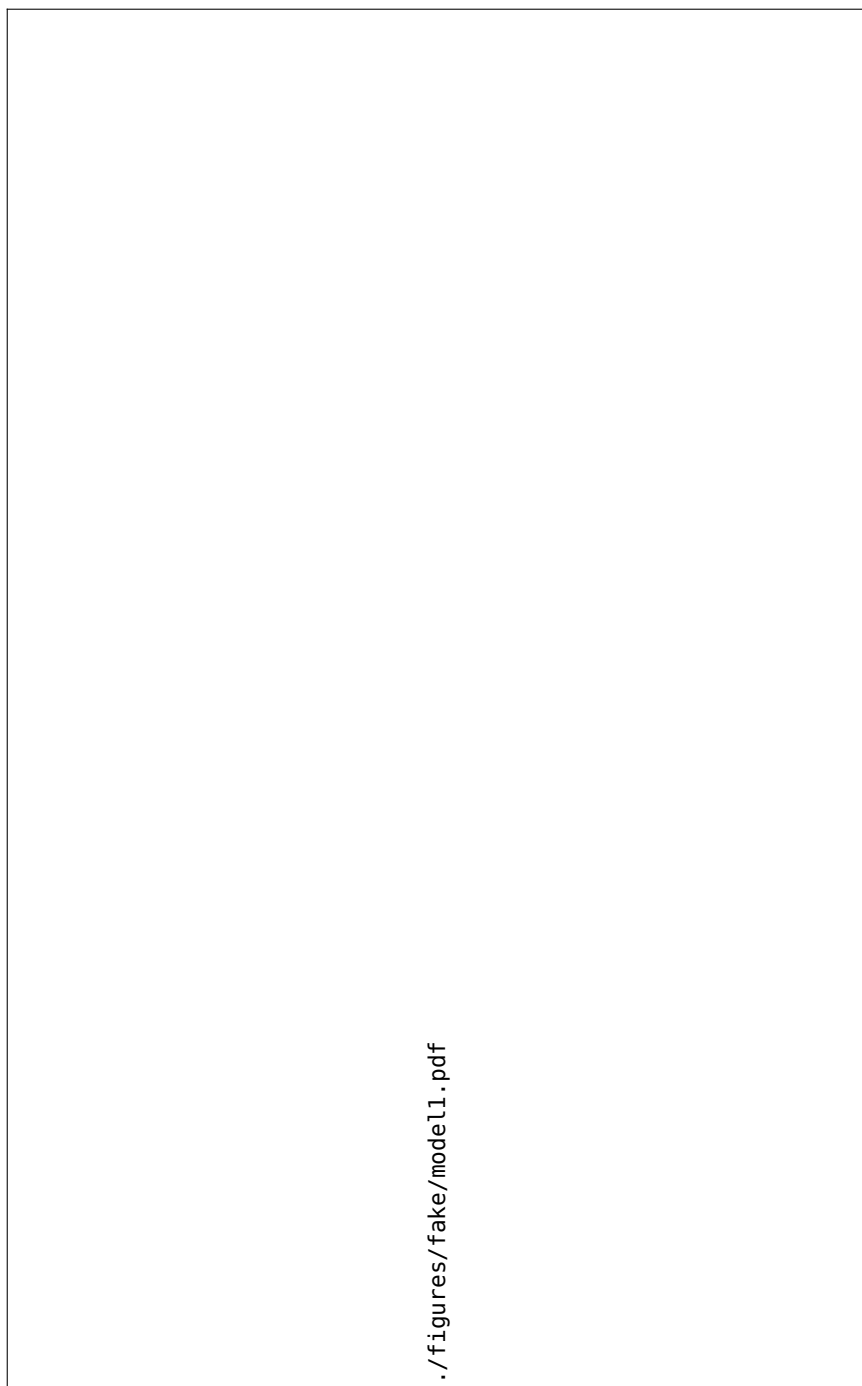
Figure 4.1: Representation of Model 1.

Table 4.1: Weight table for Model 1.

	weight	predictor	response	variable
1	5.42	c-a	A	baseline
2	3.43	d	A	s1
3	-0.29	e	A	s1
4	-3.21	g	A	s1
5	0.58	k	A	s1
6	2.89	l	A	s1
7	-7.13	o	A	s1
8	4.37	p	A	s1
9	1.89	r	A	s1
10	2.24	s	A	s1
11	0.81	t	A	s1
12	-2.62	w	A	s1
13	4.67	e	A	s1
14	0.61	g	A	s1
15	-13.22	i	A	s1
16	-8.12	l	A	s1
17	-14.45	o	A	s1
18	0.03	p	A	s1
19	-0.33	r	A	s1
20	4.25	s	A	s1
21	-18.31	u	A	s1
22	-5.43	c-a	B	baseline
23	-3.88	d	B	s2
24	-0.35	e	B	s2
25	3.28	g	B	s2
26	0.49	k	B	s2
27	-2.19	l	B	s2
28	7.49	o	B	s2
29	-4.96	p	B	s2
30	-2.67	r	B	s2
31	-2.02	s	B	s2
32	-0.49	t	B	s2
33	2.95	w	B	s2
34	-5.20	e	B	s2
35	-0.48	g	B	s2
36	12.77	i	B	s2
37	8.35	l	B	s2
38	13.70	o	B	s2
39	-0.38	p	B	s2
40	-0.11	r	B	s2
41	-3.81	s	B	s2
42	17.85	u	B	s2

errors in total: *egrr* and *grip*. It is easy to see why these errors happen: there are no comparable items in the training dataset, *grip* starts with a *gr* sequence and *egrr* is the only item with an *e* as first letter and *g* as second letter.

Table 4.3 shows the confusion matrix for the predictions of Model 1, and Table 4.4 shows a matrix with the positions of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). The TP and TN are cases where the class predicted by the model match the real class of the items. FP and FN are the cases where the class predicted by the model does not match the real class of the items. The total population N is the sum of all these values: TP + TN + FP + FN.

Table 4.2: Predictions Model 1.

	Predicted	Observed	Word
1	A	A	lama
2	A	A	lara
3	A	A	kar
4	A	B	egrr
5	B	B	liz
6	B	B	oppe
7	A	B	grip

Table 4.3: Confusion matrix for Model 1.

	Reference	
Prediction	A	B
A	3	2
B	0	2

The accuracy is the number of correct predictions divided by the total number of items. Additionally, we can calculate the confidence interval (CI) of the accuracy by using a binomial test (Clopper & Pearson 1934; Newcombe 1998). The No Information Rate (or accuracy of a model under a no information situation) is calculated as the largest class percentage in the data. In this case, **A**'s class percentage is 0.4286 and **B**'s is 0.5714, thus the latter is taken to be the No Information Rate. In other words, the No Information Rate is the accuracy of a

Table 4.4: Diagram of True Positives, False Positives, True Negatives and False Negatives.

	Reference	
Prediction	A	B
A	TP	FP
B	FN	TN

model that always predicts the most frequent outcome. In our example data **B** is the most frequent outcome. If the model predicted all outcomes to be **B**, then it would reach an accuracy of 4/7 = 0.5714. Models where all predictors have no information regarding the outcomes (i.e. they are poor predictors) tend to have an accuracy close to the No Information Rate, because always predicting the most frequent outcome guarantees the highest possible accuracy under a no information situation. The model is then said to perform above chance if the No Information Rate is less than the lower limit of the accuracy confidence interval.

There are three additional statistical values I will use in certain cases are: *Specificity*, *Sensitivity* and *Negative Predictive Value*. Specificity is the proportion of negatives that are identified as such (= TN/(TN + FP)), while sensitivity is the proportion of positives that are identified as such (= TP/(TP + FN)). The negative predictive value (= TN/(TN + FN)) will help us identify the class to which more items from other class are misclassified. These three statistics are not relevant for this particular example because we only have two classes here, but can be used, by class, in models with more than two outcomes.

Finally, the kappa statistic compares the observed accuracy with the expected accuracy (under random chance). The expected accuracy is calculated as follows. We multiply the observed frequency of **A** by the predicted frequency of **A**, and the observed frequency of **B** by the predicted frequency of **B**. We then divide these numbers by N, add them together and divide again by N. Thus, we get:

$$\text{Expected.Accuracy} = \frac{\frac{(TP+FN)*(TP+FP)}{N} + \frac{(TN+FP)*(TN+FN)}{N}}{N} \tag{4.3}$$

Finally to calculate the kappa statistic we use the following equation:

$$\text{Kappa} = \frac{\text{observed.accuracy} - \text{expected.accuracy}}{1 - \text{expected.accuracy}} \tag{4.4}$$

Kappa scores[5] go from 0 (in a perfectly random model) to 1 (in a perfectly accurate model), a kappa of 0.5 is halfway between the expected accuracy and 1. The advantage of using kappa is that it tells us how well above random chance the model is performing, and, to a degree, it allows to make model comparisons. The disadvantage is that there is no standarized interpretation and no objective cutoff point. A model with a kappa of 0.2 is not inherently bad, nor can it be said that it is at chance level. However, we can say that a model with a kappa of 0.7 is better than a model with a kappa of 0.5.

Table 4.5 shows the relevant statistics for Model 1. In this case, because our dataset is so small, the model's accuracy can not be said to be better than chance.

Table 4.5: Overall statistics for Model 1.

Overall statistics:	
Accuracy	0.7143
95% CI	(0.2904, 0.9633)
No Information Rate	0.5714
Kappa	0.3593

A second possible model for our dataset is to specify more linguistic information in the predictors. In Model 1 all we have is information about position of the segments, but not information about their nature. An alternative would be to set a model where the predictors are not selected by position only, but also by class. Instead of using the first and second letters in the pseudo words, we will now use the first consonant (c1) and the first vowel (v1). Figure 4.2 shows the structure of Model 2 as before. By selecting more structural predictors we have somewhat reduced the complexity of the model[6], but the same generalization remains: the main predictor is the first vowel of the word. The full set of weights for the model is given in Table 4.6.

---

[5]In our case: Expected.Accuracy = $\frac{\frac{3*(3+2)}{7} + \frac{2*(2+2)}{7}}{7}$ = $\frac{\frac{15}{7} + \frac{4}{7}}{7}$ = 0.4694. The accuracy is 5/7 = 0.7143. Thus, we have that kappa = $\frac{0.7143 - 0.4694}{1 - 0.4694}$ = 0.4615. Notice that the Expected Accuracy is different from the No Information Rate because the former is taken from a model that knows about the distribution of the outcomes in the traning dataset, while the latter is a completely random assignment of outcomes to inputs in the testing dataset.

[6]Notice this is only the case because of the characteristics of this dataset. In more complex datasets a more structured model will usually be more complex than a less structured one because it requires more information.

./figures/fake/model2.pdf

Figure 4.2: Representation of Model 2.

Table 4.6: Weight table for Model 2.

	weight	predictor	response	variable
1	7.43	a-c	A	baseline
2	11.87	e	A	v1
3	-21.81	i	A	v1
4	-24.18	o	A	v1
5	-34.05	u	A	v1
6	5.77	d	A	c1
7	-6.66	g	A	c1
8	0.34	k	A	c1
9	6.80	l	A	c1
10	5.68	p	A	c1
11	-2.33	r	A	c1
12	5.48	s	A	c1
13	-3.36	t	A	c1
14	-8.98	w	A	c1
15	-8.19	a-c	B	baseline
16	-12.23	e	B	v1
17	22.95	i	B	v1
18	25.17	o	B	v1
19	34.15	u	B	v1
20	-5.20	d	B	v1
21	7.28	g	B	c1
22	0.09	k	B	c1
23	-6.61	l	B	c1
24	-5.82	p	B	c1
25	2.02	r	B	c1
26	-6.57	s	B	c1
27	2.48	t	B	c1
28	9.22	w	B	c1

In Table 4.7 we see now the results of the predictions. This time the model only made one mistake: *egrr*. The reason why *grip* is correctly classified this time is that the model finds it similar enough to *liz*, *lip* and *gin*, because it now knows what its vowel is. Trying to reconstruct the evaluation of *egrr* is instructive. For obtaining the weight for **A** we add to the baseline (7.43) the weight for c1=g (-6.66) and v1=e (11.87), which gives us 12.64, and we do the same for **B** (-8.19-12.23+7.28) and we get -13.14. This clearly makes **A** win, but the node for c1 was pulling, in both cases, for **B**. This means that even though the model made the wrong choice, it did see a similarity between *egrr* and other **B** items (namely having *g* as its first consonant). This can be seen in the probabilities in Table 4.7. Of those items classified as **A**, *eggr* had the highest (even if small) probability of belonging to class **B**. I will use this aspect of the analogical models in the next chapters to measure similarity between classes.

Table 4.7: Predictions Model 2, including the probabilities for class **A** and **B**.

	Predicted	Probability A	Probability B	Observed	Word
1	A	9.99e-01	3.25e-07	A	lama
2	A	9.99e-01	3.25e-07	A	lara
3	A	9.99e-01	5.20e-07	A	kar
4	A	9.99e-01	3.18e-06	B	egrr
5	B	5.14e-05	9.99e-01	B	liz
6	B	1.83e-04	9.99e-01	B	oppe
7	B	5.55e-09	1.00e+00	B	grip

The corresponding confusion matrix and statistics for Model 2 are given in Table 4.8 and Table 4.9, respectively.

Table 4.8: Confusion matrix for Model 2.

		Reference	
Prediction		A	B
	A	3	1
	B	0	3

Table 4.9: Overall statistics for Model 2.

Overall statistics:	
Accuracy	0.8571
95% CI	(0.4213, 0.9664)
No Information Rate	0.5714
Kappa	0.72

One final important point is that we see how the CI and the Kappa are partially independent of each other. In Model 2, we obtained a Kappa of 0.72, which is considerably higher than what we would get in a random model, but because this metric is not sensitive to sample size, it fails to take into account the fact that there are only seven observations. The CI information, on the other hand, does take this into account and rightly tells us that we cannot draw any conclusion from this tiny dataset. I will use both metrics together when evaluating models.

The models in the following chapters are too large and complex to either plot, or explore by hand. For this reason I will only make use of confusion matrices and accuracy scores to evaluate them, but in principle it would be possible for someone to inspect any of the analogical models presented here.

4.4 Measuring variable importance

An issue with neural networks is the fact that it is relatively difficult to interpret the exact importance that the different factors have on the overall model. Unlike linear or logistic regression, we cannot directly explore the coefficients. However, in some cases, it is important to understand which factor plays a more or less important role predicting some dataset. To address this question we can make use of additive and subtractive modelling. The idea is very simple. For subtractive modelling we start with the complete model (with all predictors), and we compare its accuracy and kappa scores to models leaving one predictor out. This technique allows us to compare the relative importance of each individual predictor in the context of the complete model, each individual predictor is. The additive variant of this idea consists of starting with a null model without predictors, and one by one, adding the original predictors back and comparing at each step the accuracy and kappa scores.

4.5 Clustering and distances between classes

The second important method I will use throughout this book is that of clustering and measuring class similarity. Imagine the new made up set of stems whose inflectional class we want to predict as shown in (5):

(5)   a. A: *lama, lara, lado, laso, tama, ga, gal, tar, tsar, tek, tess*

       b. B: *egrr, liz, lo, loi, le, lep, loop, olpe, toi, olor, gen, grap, tak*

       c. C: *yrro, yrto, yro, undo, ujo, jyr, juk, juz, ryk*

In this new example we have three classes (**A**, **B** and **C**) which are easily described in terms of their first vowels and consonants. Words in **A** have an *a*, or an *e* preceded by a *t*, *s* or *g*. Words in **B** have an *o*, *i*, *a*, or an *e* not preceded by a *t* (except for *tak*). Words in **C** have a *y* or *u*, usually with an *r* or *j*. Additionally, we can observe that there is a much greater similarity between **A** and **B**, than **C** to the other two. **A** and **B** can appear with an *l* or *e*, and to a lesser degree *t* or *g*, while **C** does not.

We can fit a new model using again the predictors c1 and v1 to this new dataset (Model 3), and because the system is much more regular now, it should predict perfectly the class of an item. What we really want to achieve now is measuring the similarity between the three classes based on the analogical model. This can be done in different ways. In a model with few classes and lots of errors between the classes, we could look at the degree of confusion between any two classes and set classes with more confusion as more similar. In models with many classes this is less practical because class size is Zipf distributed (Blevins et al. 2016), which means that many classes will have very few members. In highly accurate models with very few errors, the measured similarity for small classes will be much less reliable. An alternative I will use in this situation is to directly use the probabilities predicted by the model.

The probabilities for Model 3 can be seen in Table 4.10[7]. In this table, each line shows the probabilities a stem has of belonging to either of the three classes. So, for *lama*, the probabilities are 0.8496 for class **A**, 0.1503 for class **B**, and 6.225e-09 for class **C**. From these probabilities we can build a (negative) correlation distance matrix[8] and from this, a distance matrix as shown in Table 4.11.

---

[7]For the purposes of this example I am not splitting the dataset into training and testing sets. For the actual case studies the probabilities used come from the same cross-validation process.

[8]When using errors instead of probabilities the process is the same, but we take the negative correlation measures for the confusion matrix instead.

Draft of May 23, 2019, 11:31

Table 4.10: Predicted probabilities for Model 3.

	Probability A	Probability B	Probability C	Word
1	8.496e-01	1.503e-01	6.225e-09	lama
2	8.496e-01	1.503e-01	6.225e-09	lara
3	8.496e-01	1.503e-01	6.225e-09	lado
4	8.496e-01	1.503e-01	6.225e-09	laso
5	9.512e-01	4.873e-02	5.017e-12	tama
6	5.987e-01	4.012e-01	3.328e-11	ga
7	5.987e-01	4.012e-01	3.328e-11	gal
8	9.512e-01	4.873e-02	5.017e-12	tar
9	9.512e-01	4.873e-02	5.017e-12	tsar
10	5.974e-01	4.025e-01	1.745e-11	tek
11	5.974e-01	4.025e-01	1.745e-11	tess
12	1.018e-01	8.981e-01	3.138e-11	egrr
13	5.833e-23	1.000e+00	1.999e-19	liz
14	2.353e-17	1.000e+00	4.455e-15	lo
15	2.353e-17	1.000e+00	4.455e-15	loi
16	3.006e-01	6.993e-01	1.220e-08	le
17	3.006e-01	6.993e-01	1.220e-08	lep
18	2.353e-17	1.000e+00	4.455e-15	loop
19	2.353e-17	1.000e+00	4.455e-15	olpe
20	8.127e-17	1.000e+00	1.107e-17	toi
21	2.353e-17	1.000e+00	4.455e-15	olor
22	1.018e-01	8.981e-01	3.138e-11	gen
23	5.987e-01	4.012e-01	3.328e-11	grap
24	9.512e-01	4.873e-02	5.017e-12	tak
25	6.060e-11	7.044e-14	1.000e+00	yrro
26	6.060e-11	7.044e-14	1.000e+00	yrto
27	6.060e-11	7.044e-14	1.000e+00	yro
28	1.308e-11	1.154e-13	1.000e+00	undo
29	5.052e-12	7.686e-15	1.000e+00	ujo
30	1.162e-09	1.756e-12	1.000e+00	jyr
31	5.052e-12	7.686e-15	1.000e+00	juk
32	5.052e-12	7.686e-15	1.000e+00	juz
33	6.060e-11	7.044e-14	1.000e+00	ryk

Table 4.11: Correlation distances for Model 3.

	Correlation matrix		
	A	B	C
A	0.000	-1.359	-1.533
B	-1.359	0.000	-1.598
C	-1.533	-1.597	0.000
	As distances		
	A	B	
B	0.359		
C	0.533	0.598	

From the distance matrix we can see that **A** is closest (has the smaller distance) to **B**, and that the greater distance is between **B** and **C**. Using the distance matrix we can then build a dendrogram using hierarchical clustering[9] (Rokach & Maimon 2005) as in Figure 4.3. Similarly, we can compress the information given in the correlation matrix from three to two dimensions using multidimensional scaling (MDS) (Borg & Groenen 2005; Cysouw 2007). Informally, MDS is a way of visualizing highly dimensional data in a two-dimensional plot. It tries to preserve as much of the original distance between two objects as possible. There is an inherent data loss when using MDS, which means the plots are an approximation, and there is dimensional data in the original distance matrix being lost. Using this two-dimensional representation of the data we can plot the categories on a two-dimensional plane as in Figure 4.4.

In this case, both representations agree with the observation from before: **A** and **B** are closer to each other than to **C**. Additionally, Figure 4.3 shows that **A** is somewhat closer to **C** than **B** is. For simple cases with only three groups I will only make use of dendrograms, but for cases with many classes I will also use MDS.

---

[9]For the clustering I use the Ward's linkage method. Although Ward's method (Murtagh & Legendre 2014) is designed to be applied to Euclidean distances, some recent studies have shown it performs formidably with other distance metrics (Meyniel et al. 2010; Strauss & von Maltitz 2017).

Draft of May 23, 2019, 11:31

./figures/fake/dendro3.pdf
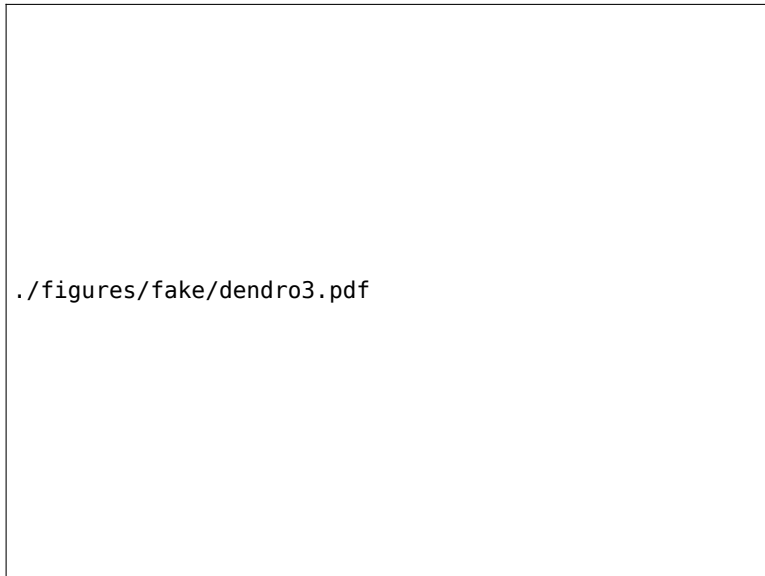
Figure 4.3: Dendrogram based on correlation distances for Model 3.
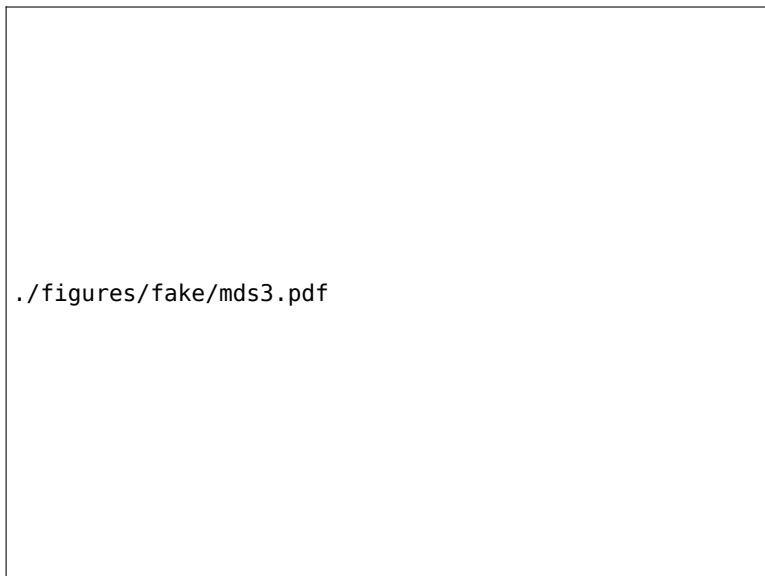
./figures/fake/mds3.pdf

Figure 4.4: MDS based on correlation distances for Model 3.

4.6 Summing up

I have shown in this chapter that building analogical models with neural networks is not conceptually different from finding analogical relations by hand. The statistical models are, to a great extent, a notational variant of informal descriptions or schemas. They have the advantage that they require less manual work and can be easily applied to very large datasets. The clustering analysis with dendrograms and the MDS analysis for finding similar classes does not produce substantially different results from what a linguist would arrive at by inspecting the items manually. As stated before, there is no claim about the cognitive reality or psycholinguistic plausibility of the neural networks themselves. Neural networks are simply tools. The claim is that the analogical relations are present in the data, and speakers can thus make use of these relations.

In the next chapters I will use these tools to explore different analogical systems in various languages. Part II contains four chapters besides this one, each corresponding to a general topic and containing at least two case studies. Chapter §5 deals with some general gender issues in Latin and Romanian. This chapter introduces the basic claim, and shows how analogical relations that predict gender in nouns have a correlate with the hierarchy. Chapter §6 shows what happens in systems where simple trees are not enough and we need hybrid types in the hierarchy. This chapter deals with the topic of overabundance and affix competition in Russian and Croatian. Chapter §7 explores the claim that we need structural information in the analogical models. I present examples from prefixing languages (Swahili and Otomí de la Sierra), where the analogical process takes place on the first segments of the items, and Hausa, were the analogical specification requires more structure than for other languages. Chapter §8 presents three cases of complex inflectional systems: Spanish verb classes and Kasem number classes. This chapter provides the strongest evidence for the interaction between type hierarchies and analogical processes. Finally, Chapter §9 sums up the results and their implications for both usage-based and formal linguistics.

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: 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.

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 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.

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 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.

Draft of May 23, 2019, 11:31

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.

Draft of May 23, 2019, 11:31

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.

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 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.

References

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.

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 Four-chotte & 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 srp-skohrvatskom 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.

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 morphology. 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-*

Draft of May 23, 2019, 11:31

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.

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 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.

Draft of May 23, 2019, 11:31

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.

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.

Draft of May 23, 2019, 11:31

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.