

Analogical classification in formal grammar

Matías Guzmán Naranjo

Draft
of May 23, 2019, 11:24

Empirically Oriented Theoretical Morphology and Syntax

Chief Editor: Stefan Müller

Consulting Editors: Berthold Crysmann, Laura Kallmeyer

In this series:

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

ISSN: 2366-3529

Analogical classification in formal grammar

Matías Guzmán Naranjo


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

This title can be downloaded at:

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

© 2019, Matías Guzmán Naranjo

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

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

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

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

ISSN: 2366-3529

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

Source code available from www.github.com/langsci/186

Collaborative reading: paperhive.org/documents/remote?type=langsci&id=186

Cover and concept of design: Ulrike Harbort

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

Typesetting software: Xe_{La}T_EX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

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

Contents

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

4	Methodological notes	53
4.1	On the general methodology	53
4.2	Statistical models and methodology	54
4.3	Analogical models using neural networks	57
4.4	Measuring variable importance	67
4.5	Clustering and distances between classes	68
4.6	Summing up	72
5	Gender systems	73
5.1	Masculine-feminine syncretism: Latin	75
5.1.1	The Latin third declension	75
5.1.2	Data	76
5.1.3	Results	77
5.2	Gender vs inflection class: Romanian	79
5.2.1	The Romanian gender and plural system	79
5.2.2	Modelling the system	85
5.2.3	Data	91
5.2.4	Results	93
5.3	Interim conclusion	102
6	Hybrid classes	105
6.1	Overabundant inflection: Croatian singular instrumental	105
6.1.1	Modelling the system	107
6.1.2	Materials	108
6.1.3	Results	108
6.2	Frequency and analogical similarity: Russian diminutives	111
6.2.1	Russian diminutives	111
6.2.2	Modelling the system	114
6.2.3	The dataset	115
6.2.4	Results	116
6.3	Interim conclusion	118
7	Morphological processes and analogy	119
7.1	Prefixes and gender: Swahili noun classes	119
7.1.1	Materials	122
7.1.2	Results	123
7.2	Prefixes and inflection classes: Eastern Highland Otomi	127
7.2.1	Verb classes in Eastern Highland Otomi	127
7.2.2	Materials	128

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

2 Remarks on analogy

Analogy can be defined in many ways, and it can be ascribed to various kinds of processes. The literature on analogy is vast and covers all sorts of phenomena and domains. Most work on it focuses on phenomena that are not directly relevant to the overall question of this book, but which are related in some way or another. In linguistics, the term *analogy* is usually employed whenever a process makes reference to direct comparison of surface items without making use of general rules, or when phonological or semantic similarities are involved, which are not easily captured as categorical generalizations. However, as a concept, analogy is rather fuzzy, and has no precise or unique definition. In the following subsections, I briefly mention some of the different phenomena for which the term analogy has been used, and in the final section of this chapter I focus on the actual kind of systems I will address in the present book.

Making justice to the history of analogy in linguistics would require a book (or several) of its own. Extensive discussions of the development of analogy as a concept in linguistics can be found in Anttila (1977), Rainer (2013) and, most extensively, Itkonen (2005).

2.1 The many meanings of analogy

2.1.1 Single case analogy

The simplest form of analogy is a similarity relation between two single items that plays a certain role in triggering or blocking a phonological or morphological process. An example of this type of analogy has been proposed to explain unpredictable new coinages and neologisms that make use of unproductive morphemes or non-morphemes (Motsch 1977: 195, see also Butterworth 1983). In such cases, a newly coined form does not make use of any derivational morphological process but is directly built on the basis of some existing form instead. Booij (2010: 89) cites the examples in (1):

- | | | | | |
|-----|----|--------------------|---|------------------|
| (1) | a. | angst-haas | → | paniek-haas |
| | | fear-hare | | panic-hare |
| | | ‘terrified person’ | → | ‘panicky person’ |

2 Remarks on analogy

- b. moeder-taal → vader-taal
mother-language father-language
'native language' → 'father's native language'
- c. hand-vaardig → muis-vaardig
hand-able mouse-able
'with manual skills' → 'with mouse-handling skills'

In these three cases, the item *haas* ('hare'), *taal* ('language') and *vaardig* ('able') are not derivational morphemes and cannot productively be used in other combinations. These are direct analogical formations because the new coinage is built from an existing compound. Various examples that follow similar processes can be found in other languages as well as can be seen in (2)-(4):

- (2) German
Früh-stück → Spät-stück
early-piece late-piece
'breakfast' → 'late breakfast'
- (3) English
handicaped & capable → handicappable
- (4) Spanish
perfumería + super → superfumería
perfume store very → 'large perfume store'

These are single case analogies because they are single formations based on the similarity to one or two words and not assumed to be a systematic (and predictable) mechanism of the language. This kind of process is not predictably productive, and there are no generalizations about when or where it can apply, but the process seems to be constantly available to speakers.

Within the rubric of single case analogies, there are multiple kinds of processes (Anderson 2015: 278). Some of these are: blending, where two words are joint together to form a new word *breakfast* + *lunch* → *brunch* (also the examples in (4)); back formation, where a new base is created for what appears to be a derived form, like the creation of the verb *edit* from the older noun *editor* (compare however van Marle 1985 and Becker 1993); folks etymology, where speakers infer the wrong etymology of a word based on analogy to another word. One such example is the word *vagabundo* ('homeless person') in Spanish which is often thought to come from *vagar* ('walk aimlessly') and *mundo* ('world') and has lead people to think it should be *vagamundo*; affix-based analogy (Kilani-Schoch & Dressler 2005), where an apparent base-affix is extended to new contexts like in

the French *atterrir* ('to land', from *terre* 'earth') → *amerrir* ('to land on the sea', from *mer* 'sea') → *alunir* ('to land on the moon', from *lune* 'moon').¹ Although there are clear differences between these processes, these cases of analogy are all based on individual specific items and do not really involve abstraction across categories.

In language change we also find examples of single case analogies, where the existence of a form prevents another form from following its expected path or, occasionally, leads to unexpected change (Bauer 2003). Anderson (2015: 276) describes this kind of phenomenon as: "where the regular continuation of some form would be expected to undergo some re-shaping by sound change, but instead it is found to have been re-made to conform to some structural pattern. This is what we usually mean by "Analogy" ". Rainer (2013) cites an example from the history of Spanish. A regular vowel change that happened between Latin and Spanish is the lowering of /i/ to /e/. Some examples of this change can be seen in (5):

- (5) a. *pīlum* → *pelo* 'hair'
- b. *īstum* → *esto* 'this'

According to this phonological rule, from the *lat. sinīstrum* ('left') the expected Spanish form would be *sinestro*, but because of analogy with the existing Spanish form *diestro* ('right (handed)'), it became *siniestro* ('sinister'). This is a single case analogical process at work. Because of semantic and phonological similarities to an existing word, some word fails to undergo a regular phonological change.

A related phenomenon is called contamination (Paul 1880: 160), which happens when two elements are so semantically similar that a new element with properties of both is created by speakers. As an example Paul mentions the German formation *Erdtoffel* ('potato') made out of *Kartoffel* and *Erdapple* (both also meaning 'potato'), and *Gemäldnis* ('painting') formed from *Bildnis* ('portrait') and *Gemälde* ('painting'). Some of these innovations are sporadic, but some can remain in the language.

Although most studies have almost exclusively focused on morphological and phonological phenomena, there has been some recent work on syntactic analogical change (De Smet & Fischer 2017). In syntax the idea is the same; a given syntactic construction changes or fails to change, by analogy to some other (usually more frequent) syntactic construction. In syntax, however, it is much harder

¹The same phenomenon is also found in Spanish with *atterrizar* ('to land on earth'), *alunizar* ('to land on the moon'), etc.

2 Remarks on analogy

to be certain that some change was due to analogical relations. A relatively recent (Colombian) Spanish innovation is [*lo más de X_{adj}*] (the most of X, meaning ‘quite X’) shown in (6):

- (6) [*lo más bonito*] + [*de lo más bonito*] → [*lo más de bonito*]
 the more pretty of the more pretty the more of pretty
 ‘the prettiest’ + ‘(one) of the prettiest’ → ‘quite pretty’

Here we see that the [*lo más de X_{adj}*] construction is a sort of blend between two different constructions, but has a unique and different meaning from the original constructions.

Comprehensive discussions of the role of analogy in language change and historical linguistics can be found in Anttila (2003), Hock (1991; 2003), Trask (1996) and, of special historical relevance, Paul (1880).

Finally, it is important to mention that single case analogy is usually thought of as a cognitive process and not as a description of a system property. Single case analogy is about what speakers do when new forms are coined, single items regularize, or when some predictable phonological change fails to apply in some specific cases. This kind of analogy will not be discussed in this book.

2.1.2 Proportional analogies

A different kind of analogy is termed proportional analogy. In its simplest form, proportional analogy involves four elements, such that: $A : B = C : X$, A is to B as C is to X. The idea here is that we can find X by looking at the relation between A and B. The earliest mention of this kind of analogy is in Aristotle’s Poetics:

By ‘analogical’ I mean where the second term is related to the first as the fourth is to the third; for then the poet will use the fourth to mean the second and vice versa. And sometimes they add the term relative to the one replaced: I mean, for example, the cup is related to Dionysus as the shield is to Ares; so the poet will call the cup ‘Dionysus’ shield’ and the shield ‘Ares’ cup’; again old age is to life what evening is to day, and so he will call evening ‘the old age of the day’ or use Empedocles’ phrase, and call old age ‘the evening of life’ or ‘the sunset of life’. (Russell & Winterbottom 1989: chapter III)

This is a rather old concept, which has also been used in linguistics extensively, most notably in morphology but also in historical linguistics (Paul 1880). This

kind of analogy is often present in word-based theories of inflection and derivation, where fully inflected forms are related to each other by proportional analogies, instead of operations deriving inflected forms from stems (Blevins 2006; 2008; 2016). Blevins (2006: 543) gives an example from Russian, with the nouns *škola* ('school') and *muščina* ('man') in the nominative and accusative as in (7).

- (7) Analogical deduction
- a. *škola:školu* = *muščina:X*
 - b. *X=muščinu*

Example (7) illustrates that if we know that for the nominative form *škola* there is an accusative form *školu*, then we can infer that for the nominative form *muščina* there will be an accusative form *muščinu*. Word based and exemplar based theories of morphology usually assume that the whole inflectional (and sometimes derivational) system of a language works as a system of analogies between known forms. This also implies that proportional analogy can (and should) be extended to sets. For example, it is not just the relation *škola-školu* which determines the relation *muščina-muščinu*, it is rather the whole set of nominative-accusative pairs speakers know.

The use of proportional analogies has not been limited to inflectional morphology. There are several proposals for derivational morphology. Singh & Ford (2003) propose a model in which derived words and simplex forms are related to each other by proportional analogies and not through morphemes or rules (see Singh et al. 2003 for several related papers, also Neuvel 2001). In this approach, formations like: *Marx:Marxism=Lenin:Leninism*, are not related by a morpheme *-ism*, but by direct analogies as shown in (8):

- (8) $/X_{Name}/ \rightarrow /X_{izm}/$

However, it is not completely clear how this differs from theories like Booij's Construction Morphology (Booij 2010), where this exact kind of relation is expressed by a construction in a very similar manner as in (9):

- (9) $[X_{Name}\text{-ism}] \leftrightarrow [\text{pertaining to SEM}(X)]$

Booij (88) suggests the difference between analogy of this kind and constructions is a gradient one, but without a clear formalization it is hard to evaluate this claim. This is a common issue with the use of proportional analogies to model some (or all) of morphology. These proposals are rarely, if ever, properly formalized (a notable exception is Beniamine 2017), and it is not always clear how they differ from rules. From a purely non-cognitive perspective, it is not obvious what

it means to say that there are no morphemes or rules, but only analogies between whole forms. The real difference seems to be in the assumptions about mental representation and the need for rich storage of fully inflected forms.

One possible clear distinctive feature of proportional analogy approaches is the existence of bidirectional relations, not usually assumed in other kinds of approaches to morphology. Proportional analogies can usually go in any direction, from any cell in a paradigm to any other cell and from a member of a derivational family to any other of its members. This property also means that there is no need for an arbitrary partition of words into stems and markers/morphemes, but the rules can look at whole words.

The lack of computational implementations of these proposals means that we cannot really evaluate how well word-based models perform at a larger scale. Although very appealing for their simplicity, it is possible that models solely based on proportional analogies cannot capture certain parts of morphology. In the end, we require a precise system that produces the *X* in the analogical equations, and this usually boils down to some sort of phonological rule set. This is not to say that there has been no work on computational implementations of proportional analogies. On the contrary, there is extensive literature on how proportional analogies can be modelled computationally (Federici et al. 1995; Fertig 2013; Goldsmith 2009; Lepage 1998; Pirrelli & Federici 1994b,a; Yvon 1997). An extensive discussion of this work is not possible, but two issues are worth mentioning. First, most work on computational implementations of analogy focuses on languages like English, Italian or Spanish. This means that it is unclear how well these systems generalize to phenomena not found in Indo-European languages (e.g. phenomena like non-concatenative morphology, tonal processes found in African languages, etc.). Second, well formalized, computational implementations of proportional analogies tend to only cover some part of a language or address some specific task. I am not aware of a computational model of proportional analogies which covers all of derivation and inflection of some language.

A different kind of phenomenon also modeled with proportional analogies is paradigm leveling. Paradigm leveling is the process by which irregular or alternating forms in the paradigm of a verb become homogeneous. A simple recent example is the superlative of *fuerte* ('strong') in Spanish. The original form in 19th century Spanish was *fortísimo* ('very strong'), but it eventually turned into *fuertísimo* during the 20th century. The idea is that proportional analogies with *bueno:buenísimo* ('good'),² *puerco:puerquísimo* ('dirty'), etc., would cause the change.

²The form *bonísimo* existed until around the 19th century. The assumption is that this form also regularized on the basis of other analogies at the time.

A generalization of this kind of process can be seen in the development of paradigm uniformity in language change (see Albright (2008a) for a review). Albright (2008a: 144) gives the example of the *eu* ~ *ie* alternations in New High German in Table (2.1):³

Table 2.1: Middle High German to Early New High German.

'to fly'	Middle High German		Early New High German		New High German
1SG	vliuge	>	fleuge	⇒	fliege
2SG	vliugest	>	fleugst	⇒	fliegst
3SG	vliuget	>	fleugt	⇒	fliegt
1PL	vliegen	>	fliegen	>	fliegen
2PL	vlieget	>	fliegt	>	fliegt
3PL	vliegen	>	fliegen	>	fliegen

The singular and plural forms had different diphthongs in MHG and ENHG, but in the change to NHG the singular and plural stems became identical. The claim is that because of an analogical process with the rest of the paradigm, the *eu* forms for the singular cells of the paradigm were replaced by *ie* forms to make the paradigm more uniform. This goes beyond single case analogies, but it can still be seen as regularization product of proportional analogies in the sense that the leveling increases the scope of a proportional analogy, making it more useful for speakers.

Proportional analogies are not really a process. Unlike the kinds of analogies discussed in the previous subsection, proportional analogies hold independently of speakers and cognitive processes. Proportional analogies hold, for example, for morphological paradigms of dead languages no longer spoken. But proportional analogies can motivate a leveling process in a paradigm, as with the examples in (2.1).

2.1.3 Analogical classifiers

A superficially similar, but distinct type of analogy is what I will call analogical classifiers. Analogical classifiers are assumed to be responsible for disambiguating between two alternatives for some lexical item. Languages often exhibit instances where a given lexeme has to be assigned to a certain category or class,

³As marked by Albright (2008a: 144), in the example > represents a regular sound change while ⇒ represents a form that has been replaced by an analogical process.

or must receive some feature, but this assignment does not directly follow from other morphosyntactic properties of said lexemes. In such cases, speakers are faced with a choice between two or more categories (or processes or features or classes, etc.) that could apply to this item and they must choose from several alternatives. Since speakers do make a choice, and usually there is agreement about what the right choice is, there must be a mechanism in place that disambiguates between the alternatives. This mechanism is *analogical* if it is based on similarity relations between the item that needs to be classified and other items for which class assignment is known. This is the type of analogy I will focus on in the remainder of this book.

The previous sections showed that analogy is sometimes understood as a process speakers use, which is different in the case of analogical classifiers. Here, we do not deal with a process, but a system of relations. As we will see, analogical classifiers can be implemented with the help of various techniques, but this does not mean that the techniques we use to build analogical classifiers have a direct relation to what speakers do. There is so far no answer to this question, and I will not attempt to answer it here.

Analogical classifiers are a relatively popular area of research among both formal and cognitive linguists. The role of phonological conditions on morphological processes and allomorphs has been acknowledged for quite some time (Kuryłowicz 1945; Bybee & Slobin 1982; Carstairs 1990) as well as the role of semantic factors (Malkiel 1988) on similar processes. This is usually known in generative grammar as phonologically conditioned allomorphy (Nevins 2011) and in usage-based and cognitive linguistics as analogy (Bybee & Slobin 1982). Despite some apparent terminological disagreements, and despite the fact both communities tend to ignore each other, phonologically conditioned allomorphy and analogy (in the sense of analogical classifiers) are not different kinds of phenomena. In both cases, we are dealing with alternations between multiple alternatives, which are resolved on the basis of phonological and semantic factors.

Analogy as a classifier lies in strong opposition to proportional analogies, however. As explained in the previous subsection, according to a model of proportional analogies, given some form C for which we want to find a corresponding X , we infer X by looking at items A similar to C for which we know B . This approach tries to avoid an abstraction step, namely the use of classes.

Given the basic proportional analogy formula $A:B=C:X$, the association between A and B is direct and thus the association between C and X must also be direct. But this does not need to be the case, the association between A and B can be mediated by an intermediate abstract feature. To make things more clear

we look at some concrete examples. Tables 2.2–2.5 show the inflection classes *-a*, *-ja*, *-o* and *-jo* for Gothic nouns (Braune 1895).⁴

Table 2.2: Gothic *-a* declension class.

	‘day’				‘bread’			
	Singular		Plural		Singular		Plural	
NOM	dags	-s	dagōs	-ōs	hlaifs	-s	hlaibōs	-ōs
ACC	dag	-∅	dagans	-ans	hlaif	-∅	hlaibans	-ans
GEN	dagis	-is	dagē	-ē	hlaibis	-is	hlaibē	-ē
DAT	daga	-a	dagam	-am	hlaiba	-a	hlaibam	-am

Table 2.3: Gothic *-ja* declension class.

	‘army’				‘herdsman’			
	Singular		Plural		Singular		Plural	
NOM	harjis	-jis	harjōs	-jōs	haírdeis	-eis	haírdjōs	-jōs
ACC	hari	-i	harjans	-jans	haírdi	-i	haírdjans	-jans
GEN	harjis	-jis	harjē	-jē	haírdeis	-eis	haírdjē	-jē
DAT	harja	-ja	harjam	-jam	haírdja	-ja	haírdjam	-jam

Table 2.4: Gothic *-o* declension class.

	‘gift’			
	Singular		Plural	
NOM	giba	-a	gibōs	-ōs
ACC	giba	-a	gibōs	-ōs
GEN	gibōs	-ōs	gibō	-ō
DAT	gibái	-ái	gibōm	-ōm

⁴In class *-ja* /ei/ can contract to /ji/ on long stems.

Table 2.5: Gothic *-jo* declension class.

	'band'			
	Singular		Plural	
NOM	bandi	-i	bandjōs	-jōs
ACC	bandja	-ja	bandjōs	-jōs
GEN	bandjōs	-jōs	bandjō	-jō
DAT	bandjái	-jái	bandjōm	-jōm

If we only consider these four classes, we can find proportional analogies that help predict most cells. For example, knowing the dative plural form *haírdjam* ('herdsman') is enough to know that its genitive plural form must be *haírdjē*. However, some cells are not fully determined. Knowing that *gibōs* ('gift') is a nominative plural is not enough for us to determine that the nominative singular should be *giba* and not *gibs*, by analogy with *dagōs* ('day').^{5, 6}

From the perspective of analogical classifiers, the alternative is that the inflectional class completely determines all cells of the paradigm of any lexeme. The individual cells, in turn, carry information about the inflection class. The distinction might seem trivial, but it requires an important abstraction step. From the analogical classifier perspective, the form *haírdjam* uniquely determines that *haírd* belongs to class *-ja* and, similarly, the form *gibōs* should uniquely determine that *gib* belongs to class *-o*. Examples (10) and (11) schematically represent how each approach works.

- (10) Proportional analogy
- $harjam:harjē = haírdjam:X$
 - $X=haírdē$
- (11) Analogical classifier:
- $harjam \in class\text{-}ja$

⁵Arguably, in a completely word-based approach there would also be confounding analogies with *bandjōs*.

⁶This situation where a cell in a paradigm only partially helps to predict another cell has been approached from an information theoretic perspective (Moscoso del Prado Martín et al. 2004; Ackerman & Malouf 2013; Blevins 2013; Ackerman & Malouf 2016; Bonami & Beniamine 2016). This approach measures the conditional entropy between cells in a paradigm, and thus quantify how informative different cells are about each other. In this book I pursue a different approach using accuracy measures.

- b. *haírdjam* ∈ *class-ja*
- c. GEN.PL, *class-ja*, *haírd* = *haírdjē*

While proportional analogies link forms to forms, analogical classifiers link forms to classes. Nevertheless, both analogical classifiers and proportional analogy models share the core idea that new forms can be generated by making reference to stored forms.

For simple cases like the Gothic examples above, there is empirically no difference between the approaches, and from a complexity perspective the analogical classifier requires extra components. On the other hand, analogical classifiers have certain advantages. The first one is that analogical classifiers are compatible with most, if not all, morphological theories. Meanwhile, models that make use of proportional analogy are usually their own theories of morphology. This means that accepting insights from analogical classifiers does not require giving up on other theoretical concepts (e.g. stems, rules of impoverishment or constructions). Additionally, from a historical perspective, analogical classifiers have been argued to be more accurate in describing linguistic change. According to Bybee & Beckner (2015: 506), constructions are responsible for licensing actual inflected forms, while analogies are responsible for licensing the combination of the aforementioned schemata with new lexical items: “given the productive schema [[VERB] + ed]_{past}, a new verb is added to the schematic category and that verb thereby becomes regular”, and it is an analogical classifier which assigns a new verb to this schema. Bybee & Beckner (2015) argue that class assignment (‘categorization’) is more important than pure proportional analogies in many cases of historical development. As an example the authors propose the verbs *strike* and *dig*, which ended up in the class of verbs like *cling*, *swing*, *hang*, etc. even though they do not actually match the schemas that describe this class (see next section for a discussion of this case). The argument is that proportional analogies did not actually take place, but speakers simply assigned these verbs to the *V~u* class: *swing~swung* (compare however De Smet & Fischer (2017) and Fertig (2013) for alternative views on the matter of analogical regularization).

This sort of change is relatively common. Single regular items might be recategorized as belonging to some irregular class, or irregular items might become regularized. Whenever there is a change in markers it tends to happen across the board, applying to all items of a class. This behaviour of inflection classes seems more compatible with a categorization system where class assignment and morphological realization are independent from each other, than with a system where they are handled by the same process.

All this being said, I will not focus on the distinction between analogical classifiers and proportional analogy models, and although I exclusively focus on analogical classifiers, some of the results from the case studies might also apply to models of proportional analogy.

2.1.4 **Summing up**

I have discussed three types of analogies that have been proposed in the linguistic literature: single case analogies, proportional analogies and analogical classifiers. Although being very different from each other, these three types of analogy all share the property of being processes or relations which: (i) focus on similarities between groups of items and (ii) allow for very fine-grained generalizations. As already mentioned, I will only discuss analogical classifiers in this book. Integrating single case analogy with theories of formal grammar will remain an open problem.

Particularly within morphology and phonology, analogical classifiers (computational and non-computational) have been proposed for a variety of languages: Dutch (Krott et al. 2001), English (Bybee & Slobin 1982; Arndt-Lappe 2011; 2014), German (Hahn & Nakisa 2000; Motsch 1977; Köpcke 1988; 1998b; Schlücker & Plag 2011), Catalan (Vallès 2004; Saldanya & Vallès 2005), French (Holmes & Segui 2004; Lyster 2006; Matthews 2005; 2010), Polish (Czaplicki 2013), Romanian (Dinu et al. 2012; Vrabie 1989; 2000) Russian (Kapatsinski 2010; Gouskova et al. 2015), Spanish (Afonso et al. 2014; Eddington 2002; 2004; 2009; Pountain 2006; Rainer 1993; 2013; Smead 2000), Navajo (Eddington & Lachler 2006), Zulu (O'Bryan 1974), as well as more theoretically oriented work (Skousen 1989; Skousen et al. 2002; Skousen 1992) among many others. It is not possible to discuss all, or even the majority, of these works here. In the following sections, I will address some of the most relevant studies. In addition, the case studies in Part II discuss some of the previous models that have tackled the phenomena in question.

2.2 **The mechanism for analogy**

So far I have not discussed what the mechanism for implementing the similarity relations in analogical classifiers actually is. As this is not the most crucial issue for the topic at hand, I will not be concerned with the question of the advantages and disadvantages of the different techniques. I will also not address the question of psycho-linguistic plausibility or mental representation. These are, no doubt,

important empirical issues, but they are ultimately tangential to the aim of this book. In this section I will present a brief overview of different systems that have been previously proposed and argue for the method I have chosen for the case studies in Part II.

In the literature there are four types of proposals for what the process behind analogy (understood as analogical classifiers) could be. These are listed in (12):

- (12) a. simple, contextual rules;
- b. schemata;
- c. multiple-rule systems; and
- d. computational statistical models

Many of the studies that have used one or the other also argued for why the alternatives are inferior or not to be preferred (Albright & Hayes 2003; Yaden 2003; Eddington 2000; Gouskova et al. 2015). I will argue instead that, leaving the point about cognitive representation aside, the systems in (12) are all more or less the same. The small differences we find between these four approaches are rather minor, and, in principle, one can almost always translate from one to the other.

2.2.1 Simple rules

Contextual rules are probably the oldest implementation of analogical classifiers, but they are also not associated with the word *analogy* very often. Contextual rules are commonly found in phonology (Chomsky & Halle 1968 and Goldsmith et al. 2011 among many others), but can be used for pretty much any domain. The format of contextual rules is usually P / C , where P stands for some process and C stands for a given context. Of course, not all uses of contextual rules count as analogical classifiers, but this does not prevent the implementation of an analogical classifier by using contextual rules. We can easily convert the format above into C / F , where C stands for a class and F for a feature, meaning that if an item has some feature F it then belongs to class C .

Phenomena that can be described in this manner are usually very small (in number of classes) and the generalizations tend to be rather straightforward. One well known example in the literature is the nominative marker in Korean (Lee 1989; Song 2006; Suh 2008).⁷ Korean nouns take the nominative marker *-i* after consonants and *-ka* after vowels as seen in (13):

⁷The actual distribution of this particle is more complex than just a nominative marker. See Song (2006) for a thorough description of its morphosyntactic properties.

2 Remarks on analogy

- (13) a. mom-i ‘body.NOM’
b. kanhowen-i ‘nurse.NOM’
c. nay-ka ‘I.NOM’
d. k^ho-ka ‘nose.NOM’

Based on grammatical descriptions, there do not seem to be any exceptions to this rule. One could model this behaviour in terms of rules as illustrated in (14).⁸

- (14) a. -i / ...C#
b. -ka / ...V#

But this is not a classifier. This is rather a morphological process that takes into account the phonological context under which it can apply. To model this phenomenon with an analogical classifier we simply propose two noun inflection classes for Korean: *class-i* and *class-ka*. Nouns belonging to *class-i* take the marker -i in the nominative, while nouns that belong to *class-ka* take -ka in the nominative. Then, the rules in (15) assign nouns to either class:

- (15) a. *class-i* / ...C#
b. *class-ka* / ...V#

This might look like we have simply rewritten same statement a different way, but it shows that analogical classifiers can easily handle simple regular cases of phonologically determined allomorphy. It also shows that simple contextual rules can be used to implement analogical classifiers without difficulty.

Although the Korean example is completely regular, this is rarely the case in phonologically conditioned allomorphy. The seemingly simple plural system in Spanish is a good example to illustrate this. Spanish nouns can end in vowels (*gato* ‘cat.MASC’) or consonants (*baúl* ‘trunk’), but not glides. The plural morpheme in Spanish has two main allomorphs: -s and -es, which are almost always predictable from the final segment of the singular form of the noun, as can be seen in (16):

- (16) a. *class-s* / ...V#
b. *class-es* / ...C#
(17) a. gatos
b. baúles

⁸Alternatively one could define only: -i / _C# as contextual, and -ka as default or the other way around.

However, it is easy to find systematic exceptions to this simple rule. One kind of exception is found in relatively recent English loanwords: *(e)sticker* - *(e)stickers* ('sticker'),⁹ *snicker* *snickers*, as well as with older French loanwords: *cabaret* - *cabarets* ('cabaret'), *carnet* - *carnets* ('ID card'). Less systematic exceptions occur in words with atypical phonotactic patterns such as *ají* ('chili peper') or *colibrí* ('hummingbird') which can take several different plural forms: *ajís/ajíes/ajises* and *colibrís/colibríes*. These are atypical because Spanish words do not usually end in a stressed /i/, but they are systematic in the sense that other words with this same ending would also allow for at least two different allomorphs (e.g. *manatí* - *manatís/manatíes* 'manatee'). This set of additional contexts could also be captured by additional rules:¹⁰

- (18) a. *class-es* / ...í#
 b. *class-s* / ...et#
 c. *class-s* / ...ker#

Additional (exception) classes would also be needed for markers like *-ses*: *ajises* ('chili peppers'), *doceses* ('twelves'). What this Spanish example shows is that even apparently simple cases might have some hidden complexity. In the end, however, contextual rules can be used to build a classifier that captures the system.

Phonologically conditioned allomorphy is a well known problem and there are many examples in the literature (Alber 2009; Anderson 2008; Baptista & Silva Filho 2006; Booij 1998; Carstairs 1998; Malkiel 1988; Rubach & Booij 2001), a recent review is given by Nevins (2011). However, the generative literature has almost exclusively focused on cases where the phonological conditioning is straightforward and can be written as a set of rules or constraints, ignoring those cases where there are no simple rules that can account for the phenomenon.

There are several reasons why phonologically conditioned allomorphy presents difficulties for traditional grammar theories. The main one is that this is a phenomenon which seems to be completely unmotivated and which adds unnecessary complexity to the grammar. The second reason is that many cases do not seem to follow any sort of clear rule pattern (although as we will see, if one looks closely enough, this is not the case). The lack of clear patterns means that the rules in the grammar must make reference to arbitrary features or adhoc constraints.

⁹Since this word is still in its early stages of borrowing there is no established orthography, but the pronunciation is /estiker/.

¹⁰One clarification would have to be added regarding additional exceptions like *caset* plural *casetes/casets*, where the system seems to have added a more regular plural.

2.2.2 Schemata

The previous subsection showed that Spanish plural formation, although relatively simple, is not uniquely determined by one single rule, but rather by several rules that make reference to the different endings of nouns. With this example in mind, one might ask how specific the phonological environment can be, and how many different possible environments there can be that determine a given alternation. There is no theory-internal or theoretically motivated answer to this question. In principle, the context of a rule could make reference to many segments, and one could have a system with dozens of different contexts. While the formal literature talks about rules, the usage-based literature talks about schemata.

To illustrate this, we will look at the phenomenon probably most often discussed in the literature: irregular verb formation in English. Regular verbs in English build their simple past form adding a *-t/d* marker to the stem. Additionally, there are groups of irregulars which do not follow this pattern. Bybee & Slobin (1982) showed that forms in (19) are not arbitrarily irregular (see also Köpcke (1998a) for a comparable analysis of German strong verbs) but that there are schematic properties they all share and that nonce words can be assigned to this conjugation pattern if they are formally similar enough to other existing items. Bybee & Slobin (1982) call these similarity relations a schema. For (19) they propose: */...ow#/~...uw#/,* and for (20): */...ɪ(N)K#/~...u(N)K#/.¹¹*

- (19) a. draw - drew
b. blow - blew
c. grow - grew
d. know - knew
e. throw - threw

- (20) a. stick - stuck
b. sink - sunk
c. swing - swung
d. string - strung

One could suggest more detailed schemata (e.g. make reference to the initial consonant cluster structure most verbs in (19) seem to share: */CL.../,¹²*, etc.)

The difference between schemata and rules is not obvious. One factor that has been mentioned as distinguishing schemata from rules (and favouring the

¹¹Where *K* stands for a velar and *N* stands for a nasal.

¹²Where *L* stands for a liquid.

former) is that they interact with prototype theory (Köpcke 1998a). While rules are blind to what lexical items they apply to, schemata can take into consideration the prototype of a class. In (20), the prototypes would be *swing* or *string*, and new items will be more or less likely to belong to this same class according to how similar they are to these prototypical items. In a prototype approach to analogy, the analogical relation to the prototype(s) of a class is more important than the relation to non-prototypical items. In such a system, schemata do not need to be completely strict, but specify preferences. They can match items that are not a perfect match, but only partially fit them.

Schemata are usually more specific than rules, and list more phonological material, but this can be emulated equally well by rules. The supposed softness of schemata can also be modelled with either more specific, larger sets of rules, or with rule weights, as in the following section.

Croft & Cruse (2004: chapter 11.2–11.3) argue that schemata can be output-oriented, i.e. they can specify the specific value of certain output, independently of what the input would be (see also Bybee 1995). In (20), the output schema would be [... $\Lambda\eta$]_{past}. This schema then groups together all verbs that build their past form with / $\Lambda\eta$ /, independently of what their present form/stem is, and what processes would need to apply to them to form the past form.

It is important to note that output-oriented schemata are a way of generalizing over inflected forms. However, these kinds of schemata are not classifiers. From the schema [... $\Lambda\eta$]_{past} one cannot know whether a particular verb inflects according to this schema or not. There needs to be a different mechanism which links the present tense form with the past tense form, or the lexeme with this output schema. Therefore it remains unclear whether this kind of schemata are relevant for analogical classifiers.

The difference between schemata and rules is a subtle one, and it usually has more to do with cognitive representation and performance. Both rules and schemata would need to be formalized before one could establish that they are not equivalent. Currently, there is no way of assessing whether the difference is spurious. In any case, it is always possible to translate a rule-based system to a schema-based system and the other way around. In the end, the use of one or the other seems to be more determined by the theoretical background of the researcher. Formal linguists usually prefer the use of rules, while cognitive and usage-based linguists prefer schemata.

2.2.3 Multiple-rule systems

The generalization of simple rule-based systems is the use of multiple-rule systems. There is no unified theory of how multiple-rule systems (for the purpose of modelling allomorphy) should work. A system could include a specific order of application, follow Panini's principle¹³, or be entirely ordering agnostic. One can write rules that only look at endings of words, complete word forms, semantics, etc. Rules can be categorical, assign weights, or be probabilistic. Since there is no agreement regarding what the properties of these systems should be, I will briefly discuss two cases from the literature.

2.2.3.1 Estonian inflectional classes

An impressive example of classes modelled with multiple rules, is the Estonian inflectional system. There are around 40 inflectional classes for Estonian nouns depending on how one counts main classes and subclasses (Erelt et al. 1995; 1997; Mürk 1997; Blevins 2008), and there is no obvious systematic way of predicting the class of a noun. Blevins (2008: 242) gives the examples in Table 2.6 to illustrate the three main Estonian inflection classes (originally in Erelt et al. 2001).¹⁴ These three classes in turn can be subdivided into further subclasses.

From the examples in Table 2.6 we see that these classes show different markers for most cells. Despite its apparent complexity, the inflectional class of a noun is highly predictable from its phonological shape (with some exceptions). Viks (1995) shows a model that can successfully predict the inflectional class of most Estonian nouns (see also Viks 1994). Viks' model consists of a series of handwritten rules that make use of three features: number of syllables, final phonemes of the stem and medial phonemes. Of the final set of 117 rules, 28 alone offer some 73% coverage, while the remaining 89 offer around 27% coverage on their own. The total set of rules covers 93% of nouns¹⁵. The main point here is not a detailed description of all of Viks' rules, the interesting aspect of this system is that a small set of rules covers a relatively large portion of nouns, while a larger set of rules is there to account for the rest of the system.

As an example we can see the two rules for nouns in Table 2.7. In the description of the segments, Viks uses the symbols *c* to indicate any of the consonants:

¹³Panini's principle says that in cases where two rules compete with each other, the more specific rule will win the competition (Zwicky 1986).

¹⁴The grave accents indicate overlong syllables. The numbers in brackets indicate the inflectional subclass given in (Erelt et al. 2001)

¹⁵The coverage does not add up to 100% because there is some overlap.

Table 2.6: Main Estonian inflectional classes.

	Class I			
	SG	PL	SG	PL
NOM	maja	majad	`lipp	lipud
GEN	maja	majade	lip	`lippude
PART	maja	majasid	`lippu	`lippusid
ILLA2/PART2	`majja	maju	`lippu	lippe
‘house’ (3)			‘flag’ (20)	
	Class II		Class III	
	SG	PL	SG	PL
NOM	kirik	kiriku	inimene	inimesed
GEN	kiriku	kiriku	inimese	inimeste
PART	kirikut	kiriku	inimest	-
ILLA2/PART2	-	-	ini`messe	inimesi
‘church’ (12)			‘person’ (12)	

Table 2.7: Rule system according to Viks (1992).

	n. syllables	final sounds	medial sounds	class	coverage (n. nouns)
a.	1	c	0	22	2612
b.	3	cUS	0	11	2036

BDFGHJKLMNPRSDZPTV and capital letters stand for literal letters. The class is a number as defined in *A concise morphological dictionary of Estonian* (Viks 1992).¹⁶

To decide between the many different rules, Viks (1995)’s model uses a simple rule-ordering procedure, “as soon as the first matching rule is found it is implemented regardless of the following ones”. The rules follow an extrinsic order, designed to maximize the accuracy of the system. Viks (1995)’s model fulfills all characteristics of an analogical classifier: it makes use of phonological properties of lexemes to assign them an inflection class.

¹⁶Notice the class numbers are arbitrary and independent of the rules and rule-ordering.

2.2.3.2 English past tense formation (again)

A different example of a multiple-rule-based system is discussed by Albright & Hayes (2003). In this study, the authors compare three possible models for the formation of the past tense in English verbs: (i) a simple rule-based model, (ii) a weighted, multiple-rule-based model, and (3) an analogical model based on work by Nosofsky (1990).

The weighted rule-based model proposed by Albright & Hayes (2003) is based on the minimal generalization algorithm first proposed in Albright & Hayes (1999). The basic idea of this algorithm is as follows. For a given morphological process that applies to a set of items, the algorithm first tries to generalize across the set of items (in this case past tense formation) and then infer the minimal rules that captures all items. For example, if the algorithm only sees *shine-shined* and *consign-consigned*, it will make the generalization in Table 2.8.

Table 2.8: Minimal generalization learner.

	change	variable	shared feature	shared segment	change location
a.	$\emptyset \rightarrow d /$		\int	am	—] _{+past}
b.	$\emptyset \rightarrow d /$	kən	s	am	—] _{+past}
c.	$\emptyset \rightarrow d /$	X	$\left[\begin{array}{l} +\text{strident} \\ +\text{contin} \\ -\text{voice} \end{array} \right]$	am	—] _{+past}

The steps in Table 2.8 show how the minimal generalization algorithm works. In the first column, we see the phonological change that needs to be applied to the present tense form, in this case adding a /d/. As to the other columns, in (a) and (b) we see two individual instances of attested past tense forms with their corresponding present tense form. The step in (c) corresponds to the minimal generalization of (a) and (b). It assigns an X to the segments which are not common between both forms, generalizes over / \int / and /s/ in terms of their feature representation and keeps the shared segments /am/. This is all within the general context of the operation of forming the past tense.

After this process is iterated, the algorithm arrives at a series of rules, of different degrees of generality, that cover the attested items. Using the accuracy of the rules and their coverage (how many items they apply to), the model then calculates weights for these rules. The weights allow the model to infer degrees of confidence for each rule and to the forms derived from them. This model can thus emulate, to a certain extent, the schemata proposed by Bybee & Slobin (1982),

in that the clusters of similarity like *fling-flung*, *sting-stung*, *cling-clung* can be captured by small rules that specifically apply to them. For these three items, the minimal generalization learner produces the rule: $/t/ \rightarrow /ʌ/$ / $[[\text{-voice}] \text{ } 1_ _ \eta]]\{[+past]]\}$. For the larger, more general set that adds *win*, *swing*, *dig*, *spring*, *spin*, *sting*, *wring*, *string*, the model has the more general rule: $t \rightarrow \Lambda$ / $[XC_ _ _ [+voice, -continuant]]\{[+past]]\}$. And so on for the other cases. With these sets of rules, Albright and Hayes’s model predicts that there should be “islands of reliability” in the irregular past tense, where verbs that look alike, by conforming to the context of the rules, will behave according to said rules.

To evaluate their model against the purely analogical model, Albright & Hayes (2003) performed two wug experiments where they asked speakers to produce the past tense of nonce verbs. These words were selected to either belong, or did not belong to the islands of reliability predicted by their model. The authors compared the responses given by the speakers with the probabilities predicted by the three different models. In the end, the multiple-rule-based model outperformed other computational models, including a multiple-rule-based model that did not include weights.

Since Albright & Hayes (2003)’s original model works from inflected forms to inflected forms, it is not, in the strict sense, an analogical classifier. However, the minimal generalization learner as a method for inferring rules could easily be deployed in an analogical classifier. An important aspect of Albright & Hayes (2003)’s system is that the rules it produces are weighted rules, unlike the rules in Viks (1994)’s system. This also means that there is no rule-ordering but weight comparison. If two different rules make different predictions for the same input lexeme, the prediction with the highest weight wins. Rule weights correspond, to a certain extent, to the idea of prototypes in the schema-based model. Rules with stronger weights capture the more prototypical shapes in the system.

2.2.4 Neural networks and analogical modelling

Two of the main computational implementations of analogy, and the ones I will focus on in this section, are neural networks and Analogical Modelling (AM).¹⁷ The use of neural networks in linguistics has a relatively long history (Bechtel & Abrahamsen 2002; Churchland 1989; McClelland & Rumelhart 1986; Rumelhart & McClelland 1986a,b). The early models were labelled connectionist models and were aimed at explaining much more than just the choice between alternatives.

¹⁷Other exemplar-based models have received considerably less attention, see Matthews (2005) for an overview.

In the second part of this book I will give a more detailed explanation of how neural networks work, but the basic idea of neural networks is that they represent (linguistic) systems in the form of weights between input, hidden and output nodes. In the context of connectionist models, input nodes see the surface linguistic forms, hidden nodes are used by the networks to represent the system in a non-symbolic way and output nodes produce the surface outputs.¹⁸

Roughly speaking, there are two kinds of neural network implementations. Early connectionist models tried to directly link meaning to form, without any kind of category assignment. That is, in a neural network predicting past tense formation in English, the network would directly learn the past tense forms of verbs and directly produce inflected verbs. The alternative approach is to train the model to learn categories. Instead of directly learning that the past tense of *fly* is *flew*, the model would learn that *fly* belongs to the class of verbs that form the past tense with a vowel change to /ew/ (i.e. an analogical classifier).

The framework of AM was initially developed by Skousen (Skousen 1989; Skousen et al. 2002; Skousen 1992) and has been applied to a variety of different phenomena like gender assignment (Eddington 2002; 2004), compounding (Arndt-Lappe 2011), suffix competition (Arndt-Lappe 2014) and past tense formation (Derwing & Skousen 1994), among others. Derwing & Skousen (1994: 193) summarize the logic behind AM as follows:

to predict behavior for a particular context, we first search for actual examples of that context in an available data base [...] and then move outward in the contextual space, looking for nearby examples. In working outward away from the given context, we systematically eliminate variables, thus creating more general contexts called supracontexts. The examples in a supracontext will be accepted as possible analogs only if the examples in that supracontext are homogeneous in behaviour. If more than one outcome is indicated by this search, a random selection is made from among the alternatives provided.

The idea is that the classification of an item is made based on how other similar items are classified. The mathematical implementation is not too important here, what is important is that AM has essentially the same properties as a neural network.¹⁹ To be clear, computationally AM and neural networks are very

¹⁸In principle, neural networks simply relate inputs to outputs, with an arbitrary number of intermediate hidden layers. Inputs and outputs can be anything, not just surface linguistic forms.

¹⁹This should not be taken to mean that both produce exactly the same result, but that the results they produce are very similar.

different from each other. The point is that they are conceptually very similar. This point has already been argued by Matthews (2005: 289), who explains that there is no crucial difference between AM and connectionist models, as long as the connectionist model is trained as a classifier:

a [neural] network designed to produce the same category mapping would have exactly the same property [as AM]. Indeed, when a network is constructed to produce just classificatory outputs, its behaviour is almost identical to that produced by AM.

It also follows that other approaches to analogical classifiers do practically the same job. Schemata are a way of measuring and finding groups of items that are surface similar, the same as the weighted rule approach. Even simple context rules like those found in phonology delimit groups of similar items.

2.2.5 Analogy or rules

The discussion of analogy/similarity systems vs rule-based systems is not new. Nosofsky et al. (1989) observed that rules can be used to compute similarity, which in turn would produce analogical systems. The distinction between both kinds of processes is not a simple one. The most explicit treatment of the differences between analogy and rules is given by Hahn & Chater (1998). The authors first acknowledge that with the common conception of rules vs analogy (the authors use the term ‘similarity’) “the best empirical research can do is to test particular models of each kind, not ‘rules’ or ‘similarity’ generally” (199), but then attempt to provide a clear way of distinguishing between rules and analogy.

They identify two distinctions: (i) absolute vs partial matches, and (ii) relative degree of abstractness of the stored pass elements. Regarding (i) the authors say that:

the antecedent of the rule must be strictly matched, whereas in the similarity comparison matching may be partial. In strict matching, the condition of the rule is either satisfied or not - no intermediate value is allowed. Partial matching, in contrast, is a matter of degree - correspondence between representations of novel and stored items can be greater or less (Hahn & Chater 1998: 202)

and regarding (ii) that:

2 *Remarks on analogy*

Second, the rule matches a representation of an instance [...] with a more abstract representation of the antecedent of the rule [...], whereas the similarity paradigm matches equally specific representations of new and past items. The antecedent ‘abstracts away’ from the details of the particular instance, focusing on a few key properties (Hahn & Chater 1998: 202)

These arguments for distinguishing rules from analogy are unconvincing, however. The argument in (i) only really matters if we can determine, with some independent method, the size of the units that the rules or similarity relations should have. Otherwise, any partial matching process can be emulated with ranked constraints, decision trees, or weighted or ordered rules, as long as these rules are smaller than the larger partial match. So, for example, partial string matching of two strings can be decomposed into categorical matching of their corresponding substrings: given the strings “aabc” and “aabb”, a categorical rule will find a partial match, as long as the rule compares 3 letter substrings and returns `true` whenever at least one of the possible substrings is correctly matched. So, unless there is some external reason for stating that the size of the comparison should be four letter substrings, the distinction between categorical rule-based and similarity-based comparison is a blurred one.

An additional difficulty with (i) is that it makes rule-based systems a special case of similarity-based systems. This is because perfect matching will happen in similarity-based systems, which means that any similarity-based system can easily emulate a rule-based system.

Finally, partial matching has the problem that it is not easily computationally implementable. Systems which implement partial matching usually do some sort of statistical evaluation as in the model by Albright & Hayes (2003), or decompose matches into smaller pieces. For example, the schema $[kl...iNK]$ can be simulated by doing smaller exact matches of its individual elements. A computer can be programmed to do matching based on estimated probabilities or confidence values, but in the end there is either a strong threshold, or some randomization process, neither of which really constitute partial matching.

The difficulty with (ii) is that, for the purpose of distinguishing between rules and similarity, it is a statement that is important from a psycholinguistic perspective, but not from a modelling perspective, as the authors admit (p. 203–204):

Rule-based reasoning implies rule-following: that a representation of a rule causally affects the behavior of the system and is not merely an apt summary description. Thus, only claims about rule-following are claims about cognitive architecture.

Their point is that the distinction about abstractness is important if we are concerned about cognitive architecture, because from a purely descriptive perspective the distinction between rules and similarity breaks down. Thus, (ii) is more a statement about how speakers store and represent previously encountered items and the nature of those representations. Although the question of rich memory is an interesting and important one (see for example Bybee 2010; Kapatsinski 2014; Port 2010a; Port 2010b, among many others), it is completely tangential to the issue at hand.

Albright & Hayes (2003)'s attempt at distinguishing rules from analogy is even vaguer. The authors claim that the key difference between analogy and a rule is that rules represent **structured similarity**, while analogy represents **variegated similarity**. Structured similarity occurs when the similarity function is restricted by some structural property of the items it operates on, while variegated similarity occurs when it is not. If, for example, the similarity function can only look at the final syllable of a word, it is making use structured similarity. The toy example in (21) illustrates the difference between variegated and structured similarity. The rule in (a) makes use of structured similarity while the rule in (b) makes use of variegated similarity. While both rules match the same segments, the rule in (a) makes use of phonological structure because it restricts the position of the similarity to the final syllable of the word. The rule in (b), on the other hand, matches any lexemes that contain the sequence /at/ in any position.

- (21) a. *class-X / .at#*
 b. *class-X / at*

This distinction is not very convincing, because it simply makes reference to a way of capturing similarity, which is mostly tangential to all other properties of analogical models. As Albright & Hayes (2003: 5) then point out, most connectionist models can infer structured similarity, which is why they do not consider these models as pure analogy. Albright & Hayes (2003) show that structured similarity seems to be a fundamental property of the linguistic systems they investigate, which they take to be support for rule-based models over analogical models. However, although it is true that some models ignore structure altogether, lumping connectionist models together with rule-based models based on whether phonological structure is at play or not draws an unnecessary ad-hoc line between analogy and rules. From this perspective, none of the models I use for the case studies are purely analogical, since they heavily make use of structural constraints on the similarity function, but they certainly are nothing like typical rule-based models.

Finally, authors like Pothos (2005), working on analogy from a more general perspective and not specifically on linguistic systems, have also arrived at the conclusion that similarity (analogical) models and rule models are simply two extremes of the same gradient. For that reason, I will not attempt to draw clear distinctions between analogical and rule-based systems. I will employ neural networks for the case studies, but these models would work equally well with hand written rules or AM.

2.2.6 Mental representations vs grammatical relations

Analogical models of grammar, and more generally, analogical accounts of grammatical phenomena are very often mixed in with discussions of mental storage, processing and psycholinguistic models (see for example Bybee (2010) and references therein). Eddington (2009: 419–420), for example, claims that “[i]n contrast to rule systems, analogy assumes massive storage of previously experienced linguistic material” and that “linguistic cognition entails enormous amounts of storage and little processing”. This is not restricted to usage-based linguistics, for example Gouskova et al. (2015)’s model explicitly mentions of storage and processing by speakers (see Chapter §6 and the next section). The questions of language processing and mental representation of language are important, but we can study analogical relations in the lexicon independently of them.

Distinguishing between mental representations and grammatical descriptions is already commonplace in most formal approaches to grammar. Stump (2016: 63–64), for example, makes a distinction between the mental lexicon (the set of forms speakers actually store) and the stipulated lexicon (“the body of lexical information that is presupposed by the definition of a language’s grammar” (64)). Rich mental storage does not go against the idea of a stipulated lexicon, but mental storage of derived or inflected forms is a tangential question to the items that need to be in the stipulated lexicon. Whether speakers only stored inflected and derived high frequency forms (Pinker & Ullman 2002; Ullman 2001; 2004)²⁰ or (possibly) every single form they ever encounter (Baayen 2007; De Vaan et al.

²⁰This position is relatively common among formal linguists who accept that frequency plays a role in processing (see for example Stump 2016 or Müller & Wechsler 2014), but it presents a problem with no solution as of yet: in these models, the only way of knowing whether a form has high frequency or low frequency, is to know its frequency. And the only way to know the frequency of a form is if said form has already been stored (Bybee 2010, but compare Baayen & Hendrix 2011). The issue could be circumvented with more complex mental storage architectures which can model frequency learning without direct frequency representations (Baayen 2011; Baayen et al. 2011; Baayen 2010; Baayen & Hendrix 2011).

2007), has no real impact on the number and nature of the items in the stipulated lexicon.

Nevertheless, the linguistic discourse on analogy has not been free from the confusion between mental representations and structural properties. The definitions usually given for analogical models make explicit reference to the mental lexicon, storage and actual speaker performance:

The analogical approach, on the other hand, deals with complex and simplex lexemes and the way they are connected to each other in the mental lexicon. It is argued that the formation of new complex lexemes is based on the paradigms of similar existing complex lexemes and their formal properties rather than on abstract rules. (Schlücker & Plag 2011: 1540)

or:

An important source of creativity and productivity in language that allows the expression of novel concepts and the description of novel situations is the ability to expand the schematic slots in constructions to fill them with novel lexical items, phrases or other constructions. Considerable evidence indicates that this process refers to specific sets of items that have been previously experienced and stored in memory. A number of researchers have used the term ‘analogy’ to refer to the use of a novel item in an existing pattern, based on specific stored exemplars (Bybee 2010: 57)

Analogical relations do not require us to postulate mental storage or psychological processes and can be formulated independently of how speakers process language. The main point linguists working on analogy want to make is that analogy expresses a relation between word forms (Becker 1990: 11).

While it is likely that speakers make use of some form of rich memory, and that analogy is closely linked to it, the model developed in this book does not require this assumption, but is compatible with it. The model I will develop in the following chapter is agnostic about these issues. The advantage of this approach is that we can avoid unnecessary debates and, most importantly, remove possible confounds.

2.2.7 Summing up

From a systemic perspective, there is not a real categorical distinction between schemata, computational systems and rules for modelling analogical classifiers.

In the end, all these systems are used to find abstractions about the shape or meaning of lexemes and find similar clusters of lexemes which belong to the same class. The real difference appears when the question of learning and mental representation is put forward. Similarity-based systems, be they computational or schema-based, assume a rich memory model of language, where speakers store most of the items they encounter and actively use stored forms to process new ones. Rule-based models, on the other hand, make the assumption that rules are learned from very few items and stored independently and abstractly. The latter type of models do not usually assume rich memory.

If the main research goal is to address the question of how speakers process and represent linguistic structures, then the distinction between rule-based and similarity/analogy-based systems is important. However, as far as modelling is concern, we are simply talking about a matter of degree. Schema and rule-based models explicitly write what the similarity relations between items must be, while AM and other computational models use mathematical objects to infer and store these similarity relations. Because in this book I do not explore the question of mental representation or acquisition, I will not argue for or against any particular implementation. The main claims of this book hold true for any of the approaches described above.²¹

2.3 Missing pieces

Despite the great progress that has been made in terms of computational implementations of analogical classifiers (from now on also simply *analogy*), as well as in the coverage of different phenomena, there are still a few conceptual issues that have been ignored and which require an answer. Broadly speaking, most work on analogy has been carried out within the cognitive linguistics and usage-based linguistics communities (the most recent exception being Gouskova et al. (2015), who seems to mostly ignore work coming from these two communities). For better or worse, research on analogical classifiers has mainly focused on developing new and better computational models, as well as trying to find out what the limits of analogical classifiers are, by applying them to all kinds of phenomena. This, however, has come with a relative lack of attention to proper formalization of what analogical classifiers actually are and how they relate to grammar.

²¹One possibly incompatible approach is Optimality Theory (OT). The problems that an OT model would face will become clear in the next chapter when I present the formal model of analogy.

Some of the glaring problems were mentioned by [Wills & Pothos \(2012\)](#). The authors argue that analogical classifiers (what they call “categorization models”) suffer from not being explicit about their scope, and because they are fitted to each individual phenomenon, models are not consistent with the variables they work on. This is an important point. In the examples mentioned in the previous section, the analogical classifiers were built to deal with only one alternation. Each classifier looks at specific predictors relevant for each phenomenon, and it is strictly confined to that phenomenon. So far, there is no theory of how this restriction takes place or how it relates to the rest of the system. The main question missing and answer is given in:

- (22) How do analogical processes (understood as analogical classifiers) interact with grammar and with each other?

This is not trivial. There is, so far, no analogical model that can capture most of, let alone all, language domains. There are not even analogical systems that can capture most of a single domain. In other words, analogical classifiers are designed to capture specific phenomena within a well defined and limited domain, but they cannot capture the whole (or a sizable portion) of the morphology, phonology or syntax of a language. This basically means that even if we accept that a large number of phenomena in language require and are best accounted by, analogical systems, grammar (in the form of constructions, features, rules, etc.), still needs to take care of the rest. This also means that grammar needs to interact with analogy in a clearly defined way.

Unless the claim is put forward that *all* grammatical phenomena in language can be accounted for with analogy, a formally well defined interface between analogy and grammar is required. This interface must make explicit what kinds of interactions we see between analogy and grammar, what the domains are and, importantly, where the limits of analogy lie.

The interactions between analogical classifiers are also poorly understood. Supposing that a language can have more than one phenomenon which is explained by an analogical classifier, it is not clear whether these two classifiers interact with each other and how. If a language organizes irregular verbs and nouns according to analogical classifiers, one would like to also know whether these classifiers are independent from each other and to what degree.

Another pressing issue relates to the targets of analogy, or the features analogy can and cannot see. [Albright \(2009: 185\)](#) correctly points out that “an adequate model of analogy must [...] be restrictive enough to explain why speakers generalize certain statistical properties of the data and not others”. This question has

mostly been ignored. Bybee (2010: 54) emphasizes that “[m]ost analogical formations in language are based on semantic or phonological similarity with existing forms”, but acknowledges that

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 strung verbs more important than the initial ones? (Bybee 2010: 62)

There have been a couple studies which have, only indirectly, dealt with some of these questions. In the generative literature most of the phenomena of phonologically conditioned allomorphy are dealt with either context rules, or OT, or sometimes simply just described but not really modelled (e.g. Rubach 2007: 119). In the usage-based literature the issue of analogy-grammar interaction is mostly ignored, or taken from granted. As far as I am aware, there have been no attempts at explicitly answering the question in (22), only a few informal approaches.

Probably the most explicit formulation of how analogy interacts with grammar is given by Bybee & Beckner (2015). Bybee & Beckner suggest a model where analogy classifies lexical items according to whether they are compatible with different constructions or not. For Bybee & Beckner (2015), a construction like: $[X_{\text{VERB}}-d] \leftrightarrow [\text{past}(\text{SEM}(X))]$,²² is responsible for producing the past tense form of regular English verbs. What the analogical classifier does is simply decide which verbs can be combined with this construction.²³ However, Bybee & Beckner (2015) are not really explicit on how this happens or where. There are multiple alternatives: The analogical classifier could apply immediately whenever any new verb is learned and assign a feature to it specifying which inflectional construction it is compatible with, or it could apply every time a speaker wants to inflect said verb. It is also not clear how different constructions compete with each other. One could have a classifier which directly decides which constructions a lexical item is compatible with, or there could be individual classifiers for each constructions deciding whether some given lexical item is compatible or not.

All this being said, this book is mostly an attempt at formally implementing and testing the Bybee & Beckner proposal, where analogy and grammar are independent but closely interlinked with each other.

²²Bybee & Beckner (2015) use a slightly different notation, but the idea is the same.

²³A very similar model of analogy-grammar interaction is discussed by Gouskova et al. (2015). Working in the framework of Distributed Morphology (Halle & Marantz 1993), Gouskova et al. (2015) propose a model for dealing with Russian diminutives based on more or less the same principles. See Chapter §6 for a discussion of this model.

2.4 Final considerations

In this chapter I provided a short summary of some of the main different uses of analogy in linguistics. I presented single case analogies, proportional analogies and analogical classifiers. The main difference between single case analogy and analogical classifiers is that the former directly links forms to forms, while the latter uses an intermediate abstraction step that links forms to classes.

Analogical classifiers are of interest to both usage-based and formal linguists. Analogical classifiers are capable of capturing what has been seen as different processes (phonologically conditioned allomorphy, inflection classes, gender assignment, etc.) and treat them as a single phenomenon. There are several techniques used for implementing analogical classifiers (rules, schemata and computational implementations), and although superficially very different from each other, they are, at their core, very similar and often interchangeable.

Although there has been a considerable amount of research on analogical classifiers, there are still several questions pertaining to the interaction between analogical models and grammar. Answering these questions is crucial if we want to have a better understanding of exactly how much analogy can do and how much it cannot do. We want to avoid waiving away phenomena by simply invoking analogy as a magical solution, but we also want to avoid overly complicating grammatical analysis by trying to explain those aspects that analogical models can more easily capture.

References

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

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

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

References

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

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

References

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

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

References

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

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

References

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

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

References

- Kempe, Vera, Patricia J. Brooks & 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*. Stanford: CSLI publications.
- Kohavi, Ron. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on artificial intelligence*, 1137–1143. Morgan Kaufmann.
- Köpcke, Klaus-Michael. 1988. Schemas in German plural formation. *Lingua* 74(4). 303–335.
- Köpcke, Klaus-Michael. 1998a. Prototypisch starke und schwache Verben der deutschen Gegenwartssprache. *Germanistische Linguistik* 141(142). 45–60.
- Köpcke, Klaus-Michael. 1998b. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language* 25. 293–319.
- Köpcke, Klaus-Michael, Klaus-Uwe Panther & David A. Zubin. 2010. Motivating grammatical and conceptual gender agreement in German. In Hans-Jörg Schmid & Susanne Handl (eds.), *Cognitive foundations of linguistic usage patterns*, 171–194. Berlin, Boston: De Gruyter Mouton.
- Köpcke, Klaus-Michael & David A. Zubin. 1984. Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*. 26–50.
- KordiĆ, Snježana. 1997. *Serbo-Croatian*. München: Lincom Europa.
- Kramer, Ruth. 2015. Impoverishment, gender and number: Predicting the patterns of syncretism. In *Roots IV*. Georgetown University.
- Krott, Andrea, R. Harald Baayen & Robert Schreuder. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39. 51–94.
- Kuryłowicz, Jerzy. 1945. La nature des procès dits «analogiques». *Acta Linguistica* 5(1). 15–37.

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

References

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

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

References

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

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

References

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

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

References

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

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

Language index

Bosnian, Croatian and Serbian, 105–
111

Derivation, 111–117
Dutch, 5

English, 6, 24
Estonian, 22

German, 6, 7, 11
Gothic, 13–15

Hausa, 130–135
Highland Otomi, 127–129

Kasem, 164–207

Latin, 7, 77–81

Romanian, 81–103
Russian, 111–117

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

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

Subject index

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

Analogical classification in formal grammar

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

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

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

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

ISBN 978-3-96110-186-3

