

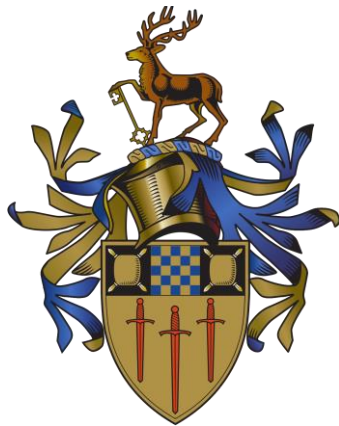
A Typological Approach to the Morphome

by

Borja Herce

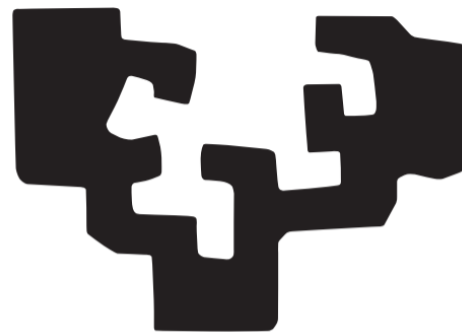
Submitted for the Degree of Doctor of Philosophy

University of Surrey



University of the Basque Country

eman ta zabal zazu



Faculty of Arts and Social Sciences

School of Literature and Languages

Supervisors:

Dr. Matthew Baerman

Dr. Iván Igartua

Dr. Oliver Bond

©Borja Herce 2020

Declaration of originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to their originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the Universities have the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so assessed, the University reserves the right to require an electronic version of the final document (as submitted) for assessment as above,

Borja Herce, University of Surrey, Guildford 7th of November 2019

Summary

This dissertation constitutes the first typologically-oriented monograph on morphemes, which is the term given to systematic morphological identities that do not map onto morphosyntactic or semantic natural classes like 'plural', 'past', 'third person singular' etc.

Chapter 1 discusses terminology and the relevant literature on this unusual phenomenon and Chapter 2 introduces some necessary clarifications with respect to the identification and definition of morphemes. From there, the discussion moves to more empirical matters. In Chapter 3, the notions of naturalness and economy are discussed in detail, and the connection of morphomicity to other morphological phenomena is explored. Diachrony takes then central stage, as Chapter 4 presents the different ways in which morphomic structures may emerge, change, and disappear from a language.

Chapter 5 constitutes the core of the dissertation and presents 110 morphomic structures that the author has identified across the world's languages. These structures are presented with great qualitative detail along with their diachronic history if known. On the basis of the synchronic variation across the morphemes, a dozen logically independent variables have been identified (in the spirit of Canonical or Multivariate Typology) as relevant to describing these structures in the most fine-grained detail. These variables have been operationalized into quantitative measures. After establishing the values they take in all 110 morphemes in the database, statistical analysis has been undertaken to spot correlations and dependencies between them which are subsequently discussed.

Another major finding to emerge from this synchronic database concerns the cross-linguistic recurrence of various morphomic structures. Some extramorphologically unmotivated structures like SG+3PL, 1SG+3, PL+1SG etc. have been found to be present in several genetically and areally unrelated languages. This is something which had been previously excluded from the phenomenon on definitional grounds.

The dissertation finishes by reiterating in Chapter 6 its findings and their implications for morphomic research, and for typology and morphology more generally.

Resumen

Esta tesis constituye la primera monografía de orientación eminentemente tipológica sobre morfomas. Este término denota estructuras morfológicas sistemáticas cuya extensión paradigmática no se corresponde con distinciones semánticas o morfosintácticas como 'plural', 'genitivo singular' etc.

El Capítulo 1 presenta y discute la literatura previa y cuestiones terminológicas, y el Capítulo 2 clarifica cuestiones relativas a la definición e identificación de los morfomas en casos concretos. La discusión se traslada a continuación a un plano más empírico. El Capítulo 3 discute las nociones de 'clase natural' y 'economía', y explora la relación entre morfomicidad y otras desviaciones morfológicas. La diacronía se convierte en protagonista en el Capítulo 4, donde se presentan y discuten las diferentes maneras en que pueden surgir, cambiar o desaparecer los morfomas en las lenguas.

El Capítulo 5 es el central de la tesis y presenta 110 morfomas identificados por el autor en lenguas de todo el mundo. Todas estas estructuras son presentadas detalladamente junto con su historia en muchos casos. En base a la variedad observada entre morfomas, se ha definido una docena de variables independientes en torno a las cuales se estructura dicha variación. Tras operacionalizar dichas variables y establecer su valor en los 110 morfomas mencionados, se explora estadísticamente su correlación.

Otro resultado derivado de esta base de datos sincrónica se refiere a la recurrencia cross-lingüística de morfomas concretos. Algunas estructuras, arbitrarias desde el punto de vista morfosintáctico o semántico (SG+3PL, 1SG+3, PL+1SG etc.), se encuentran presentes en lenguas independientes, es decir, no emparentadas ni relacionadas arealmente. Esto supone una novedad con respecto a la literatura anterior.

La tesis concluye reiterando en el Capítulo 6 los resultados principales de la investigación y explorando sus implicaciones en relación a nuestro conocimiento de los morfomas en particular y del campo de la tipología y la morfología en general.

Acknowledgments

Writing this dissertation has been a long trip. Someone told me recently that few people come into linguistics in a smooth way. It is definitely my case. For a long time, linguistics was (quoting J. R. R. Tolkien) my “secret vice”. Working at a bank and as a teenager I would spend a lot of my free time ‘conlanging’ and reading papers by Bybee, Givón or Haspelmath. I should therefore start by thanking them and others for luring me into the discipline. After realizing how miserable I was doing other things for a living, I finally dared to take the risk and do what I loved. I started taking English Philology courses while still at the bank (thank you so much, Mikel and Andoni, for sharing all those classnotes). When I could actually attend them, my first ‘truly linguistic’ courses turned out to be everything I had hoped for and more. I am particularly indebted to Carlos García-Castillero, my first teacher of historical linguistics. He is a true role model of what a professor and academic should be like: always eloquent and open to discussion, no matter how many questions I would ask during a single lesson (which in hindsight might have been too many). During my Bachelor, MA and PhD a lot of other people at the University of the Basque Country have helped me in more ways than I can name here: Joaquín Gorrochategui, Natalia Jardón, Nerea Madariaga, José Mari Vallejo... and of course my supervisor Iván Igartua. During my MA, he was a source of inspiration in his typology courses. He has continued to be so ever since. I am still amazed by Iván’s knowledge of the field and by his generosity with his own time, always ready to spare some for me even when he had none. Maybe as important, he has always had kind words towards me and my work, which can be a true game-changer during an emotionally challenging period like a PhD.

As for the English part of this transnational marriage that is my dissertation, I first met Matthew Baerman in a seminar he taught during my MA in the Basque Country. He may have been the first internationally renowned linguist I ever met and I was surprised by his affable, accessible nature. Matthew is probably the best supervisor I could have asked for: quick-witted, always available, generous and insightful with his comments and remarks, a source of calmness and balance in a hectic world. Still he has always managed to push me to do better, to be clearer, and to get out of my comfort zone. For all of that and what is still to come: thank you. I also owe a lot to the rest of the SMG gang both past and present. I want to thank in particular Olly Bond (my co-supervisor) for his time and detailed comments of my work. His

input has definitely improved this dissertation by forcing me to be more explicit and friendlier to a nonspecialist readership. I also thank Matt Carroll, Grev Corbett, Briana van Epps, Tim Feist, Steven Kaye, Enrique Palancar, Tatiana Reid, Erich Round, and Helen Sims-Williams for many interesting discussions over coffee about my work or theirs. I want to have particularly kind words towards my buddy Dávid Gyórfi, my roomie Jérémy Pasquereau, and my fellow Basque colleague Kristian Roncero for very many work and non-work-related conversations and for friendship. Of course, many other non-SMG linguists have also provided me with advice and feedback over the last years. Most relevant to this thesis has been that of fellow ‘morphomists’ Louise Esher and Paul O’Neill.

On more down-to-earth issues, I cannot forget to thank the Basque government for the scholarship that has provided my salary during the last four years, and Techne for paying my fees in Surrey. This thesis would not exist without them. Last of all I would like to thank my family and friends for putting up with my long absences, my absent-mindedness, and my general neglect of them over these years and in the future. Mum, dad, grandma, I’m sorry.

Table of contents

1 Introduction

1.1 Initial approximation and goals.....	page 11
1.2 Literature review.....	page 13
1.3 Terminology.....	page 17
1.4 Towards a working definition.....	page 21

2 Topics on morpheme identification

2.1 Morphemes and accidental homophony.....	page 25
2.1.1 Tests for the grammatical relevance of a formal identity.....	page 26
2.1.2 On the empirical status of homophony and polysemy.....	page 33
2.2 Morphemes and blocking	page 41
2.3 Morphological zeroes.....	page 44
2.4 Maximal domain of morphemes.....	page 46
2.5 Isolatability and the concept of the 'stem space'.....	page 54
2.6 Morphemes and cross-linguistic recurrence.....	page 58
2.7 Morphemes and locus of exponence.....	page 61
2.8 Segmentability.....	page 63
2.9 Difficult cases.....	page 66
2.9.1 The problem of the 1PL.....	page 66
2.9.2 Syntactically-licensed morphemes.....	page 69
2.9.3 Gender or morpheme? Other analytical uncertainties.....	page 72

3 Empirical aspects on the morpheme

3.1 What else can be morphomic?.....	page 83
3.1.1 Syncretism.....	page 83
3.1.2 Heteroclisis.....	page 85
3.1.3 Overabundance and defectiveness.....	page 86
3.1.4 Morphomicity in derivation.....	page 89
3.2 The scale from natural to unnatural.....	page 90
3.3 Independence of phonology.....	page 102
3.4 Economy and the morpheme.....	page 110

4 Morphemes in diachrony

4.1 The emergence of morphemes.....	page 116
4.1.1 Sound change.....	page 116
4.1.1.1 The paradigmatic distribution of zero.....	page 122
4.1.2 Semantic drift.....	page 126
4.1.3 Analogy.....	page 128
4.1.3.1 Morphosyntactically motivated analogy.....	page 129
4.1.3.2 Analogy motivated by form.....	page 131
4.1.3.3 A note on the motivation of analogy.....	page 134
4.1.4 Pattern interactions.....	page 135
4.1.5 Grammaticalization.....	page 138
4.1.6 Borrowing.....	page 141
4.1.7 Mixed origins.....	page 145
4.2 Morpheme loss.....	page 146
4.2.1 Loss of productivity and gradual erosion.....	page 146
4.2.2 Loss of morphosyntactic categories.....	page 149
4.2.3 Sound change.....	page 151
4.2.4 Analogy.....	page 152
4.2.4.1 Analogical change into a natural class.....	page 152
4.2.4.2 Analogical change into an unnatural class.....	page 156
4.2.5 Mixed causes.....	page 159
4.3 Discussion.....	page 160
4.4 Conclusion.....	page 165

5 Morphemes in synchrony

5.1 Criteria for inclusion into the synchronic morpheme database.....	page 167
5.1.1 Unmistakably unnatural paradigmatic distribution.....	page 167
5.1.2 Unmistakably systematic formal identity.....	page 171
5.1.3 Other requirements.....	page 175
5.1.4 Some excluded morphemes.....	page 177
5.2 Cross-linguistic database of morphemes.....	page 184
5.2.1 Achumawi.....	page 189

5.2.2 Aguaruna.....	page 190
5.2.3 Aragonese.....	page 192
5.2.4 Athpariya.....	page 194
5.2.5 Ayoreo.....	page 196
5.2.6 Bantawa.....	page 197
5.2.7 Barai.....	page 200
5.2.8 Basque.....	page 201
5.2.9 Benabena.....	page 202
5.2.10 Biak.....	page 202
5.2.11 Burmeso.....	page 204
5.2.12 Burushaski.....	page 205
5.2.13 Chinantec L.....	page 206
5.2.14 Chinantec, P.....	page 207
5.2.15 Daasanach.....	page 209
5.2.16 Daju, M.....	page 210
5.2.17 Darma.....	page 211
5.2.18 Ekari.....	page 213
5.2.19 English.....	page 213
5.2.20 French.....	page 215
5.2.21 Fur.....	page 216
5.2.22 Girawa.....	page 217
5.2.23 Greek.....	page 219
5.2.24 Icelandic.....	page 220
5.2.25 Iraqw.....	page 221
5.2.26 Irish.....	page 223
5.2.27 Italian.....	page 224
5.2.28 Jabuti.....	page 226
5.2.29 Jerung.....	page 227
5.2.30 Karamojong.....	page 228
5.2.31 Kele.....	page 229
5.2.32 Ket.....	page 230

5.2.33 Khaling.....	page 230
5.2.34 Khinalug.....	page 233
5.2.35 Koasati.....	page 234
5.2.36 Koiari.....	page 235
5.2.37 Kosena.....	page 236
5.2.38 Luxembourgish.....	page 237
5.2.39 Maijiki.....	page 238
5.2.40 Malinaltepec M.....	page 239
5.2.41 Mazatec.....	page 241
5.2.42 Mehri.....	page 241
5.2.43 Menggwa Dla.....	page 242
5.2.44 Mian.....	page 243
5.2.45 Murrinh-Patha.....	page 245
5.2.46 Ngkolmpu.....	page 246
5.2.47 Nen.....	page 247
5.2.48 Nimboran.....	page 248
5.2.49 Nivkh.....	page 250
5.2.50 North Saami.....	page 251
5.2.51 Northern Akhvakh.....	page 252
5.2.52 Old English.....	page 253
5.2.53 Páez.....	page 254
5.2.54 Pite Saami.....	page 255
5.2.55 Skolt Saami.....	page 258
5.2.56 Sobei.....	page 260
5.2.57 Spanish.....	page 260
5.2.58 Sunwar.....	page 262
5.2.59 Svan.....	page 263
5.2.60 Thulung.....	page 265
5.2.61 Tol.....	page 267
5.2.62 Turkana.....	page 270
5.2.63 Twi.....	page 271

5.2.64 Udmurt.....	page 273
5.2.65 Vitu.....	page 274
5.2.66 Vurës.....	page 275
5.2.67 Wambisa.....	page 276
5.2.68 Wutung.....	page 278
5.2.69 Yagara.....	page 281
5.2.70 Yorno-So.....	page 282
5.2.71 Zapotec Y.....	page 283
5.3 Typologizing morphemes.....	page 286
5.4 Variables of cross-morphomic variation.....	page 288
5.4.1 External morphosyntactic constraints.....	page 389
5.4.2 Word form recurrence.....	page 291
5.4.3 Paradigmatic recurrence.....	page 294
5.4.4 Cross-lexemic recurrence.....	page 296
5.4.5 Number of exponents.....	page 299
5.4.6 Shared form.....	page 301
5.4.7 Locus of marking.....	page 303
5.4.8 Informativity.....	page 304
5.4.9 Morphosyntactic coherence.....	page 307
5.4.10 Morpheme paradigm size.....	page 310
5.4.11 Cross-linguistic morpheme recurrence.....	page 314
5.4.12 Recurrence of geometrical patterns.....	page 320
5.5 Statistical analysis.....	page 330
5.6 Conclusion.....	page 345
6 Conclusion	
6.2 The importance of features.....	page 347
6.3 The importance of form.....	page 352
6.3 Summary and results.....	page 356
Appendix.....	page 365
References.....	page 369

1 Introduction

1.1 Initial approximation and goals

The present dissertation deals with morphemes, understood as morphosyntactically unnatural sets of paradigm cells that systematically share (some of) their exponence. The concept was introduced by Aronoff (1994) and popularized by Maiden's research on the diachronic behaviour of stem alternations in Romance varieties. These patterns have been extensively studied over the last few years and have even been given names of their own:

	<i>caber</i> 'fit' illustrating the L-morphome				<i>poder</i> 'can' illustrating the N-morphome			
	Pres. Indicative		Pres. Subjunctive		Pres. Indicative		Pres. Subjunctive	
	SG	PL	SG	PL	SG	PL	SG	PL
1	quepo	cabemos	quepa	quepamos	puedo	podemos	pueda	podamos
2	cabes	cabéis	quepas	quepáis	puedes	podéis	puedas	podáis
3	cabe	caben	quepa	quepan	puede	pueden	pueda	puedan

Table 1: Two morphomic stem alternations in Spanish

As Table 1 illustrates, the Spanish verb 'fit' has a dedicated stem in 1SG.IND+SBJV. The verb 'can', in turn, has a different stem in SG+3PL.¹ These stem alternations patterns are surprising because the sets of cells that share a stem do not constitute morphosyntactic natural classes (like e.g. 'subjunctive', 'speaker', 3PL etc.) and seem to be arbitrary instead. These morphological affinities are, however, systematic within the language, since they are repeated in hundreds of verbs and with different formal exponents. These facts are well-known from the research of linguists like Malkiel (1974), Maiden (1992, 2005, 2018b), O'Neill (2013) and others.

In stark contrast to the wealth of research on Romance morphemes, very few studies have explored the phenomenon at length by engaging with data from other languages families.² As

¹ The 2SG imperative (not shown in Table 1) also forms part of the Romance N-morphome.

² Notable exceptions include Round (2013) and Stump (2015:128-140). Cross-linguistically oriented research has been conducted, of course, on notions that are related to the morpheme, e.g. on

a consequence, our understanding of the phenomenon, both synchronic and diachronic, is likely to be incomplete and/or biased in important respects. This is the research gap that the present dissertation is set to fill.

A typological approach to the morpheme faces, of course, considerable difficulties. The most important of these is the sheer variation of the morphological component of grammar across languages. As pointed out by Baerman & Corbett (2007:115), “[o]f all the aspects of language, morphology is the most language-specific and hence least generalizable”. Consequently, there are important challenges to the extrapolation of meaningful principles.

Another very significant challenge is the nature of the morpheme itself. It is usually assumed that the notion is dependent on the cognitive status of the morphological associations. That is, morphemes, to be truly morphemes, must “constitute grammatical realities for speakers” (O'Neill 2014:32). This, however, is difficult to ascertain. The evidence that is usually presented in relation to this may be diachronic (e.g. the preservation or replication of formal alliances) or experimental (see e.g. Nevins et al. 2015). These types of evidence are regrettably unavailable for the vast majority of the world's languages. In addition, even when present, their contribution is often unclear and subject to different interpretations. For this reason, alternative approaches will have to be explored to approach the morpheme as a coherent object of analysis for a typological study.

The main contribution of this thesis, therefore, will be a typological study of morphemes (with a cross-linguistically varied sample of 110 of them). These data will be, of course, at the service of research questions such as: What types of morphemic structures are possible? What are the synchronic properties of morphemes? What patterns are common and which are infrequent and why? Synchronic data will be complemented with diachronic insights to inform us about: What are the most frequent sources and outcomes of morphemes? What role do frequency or morphosyntactic features play in their evolution?

Finally, this research will contribute to the broader discussion on the relation of the morpheme and the morpheme, and the phenomenon's overall place in grammatical and

'morphologically stipulated patterns of syncretism' (see Baerman et al. 2005).

morphological architecture. The diachronic and synchronic evidence gathered in this dissertation will help to answer the fundamental questions of the morpheme debate (Luís & Bermúdez-Otero 2016): What is the function of morphemes, if any? What makes them learnable? Is there a learning bias against morphemes? And ultimately: Are there any empirical properties distinguishing morphemes from morphemes?

The answers to these questions and the outcomes of this research will be relevant not only to theoretical morphology and typology in general but also to language description and documentation. At present, when field linguists come across formatives that do not have a clear-cut function, they resort to various terms such as “thematic suffix” (Dunn 1999:186), “epenthetic morpheme” (Morrison 2011:241), or “ligature” (Obata 2003:60) and list their uses and properties in the best way they can. A typology of morphemes, like the one presented in this dissertation, will provide them with the tools to describe these structures more thoroughly, more coherently and with a more homogeneous terminology.

1.2 Literature review

The term 'morpheme' and the adjective 'morphemic' are relatively new additions to linguists' analytical toolkit. They were famously coined by Mark Aronoff in his 1994 monograph *Morphology by Itself*. His basic claim was that morphology had organizing principles of its own so that “the mapping from morphosyntax to phonological realization is not direct but rather passes through an intermediate level.” (Aronoff 1994:25). He presented evidence of various phenomena (e.g. intraparadigmatic affinities, inflectional classes, etc.) that necessitated, in his opinion, the recognition of an autonomous morphological component in language.

Aronoff's monograph and term put autonomous morphological phenomena back at the forefront of language research. However, many before him had made observations that were difficult to reconcile with traditional morphemics. Well known examples are Maiden (1992), which set the stage for the vast subsequent literature on Romance morphemes, and Matthews's (1991:97), with his famous dictum that “one inflection tends to predict another”.

The syncretisms of Matthews, where one cell's inflection appears to take as a base the form of another cell, foreshadowed recent developments (e.g. Stump & Finkel 2013, Ackerman & Malouf 2013) in the set-theoretic and information-theoretic research of predictive relations within the paradigm.

Another researcher whose work casts doubt on traditional morphemic models was Hockett. His claim that sometimes “it is not the formal grammatical structure that yields the resonances; it is the resonances that induce the grammatical structure” (Hockett 1987:88) is very much in line with the core assumptions of current morphomic literature.

An alternative way of accounting for the problematic facts of language before Aronoff (1994) was to extend the notion of the 'morpheme' in a way that it would accommodate many (or all) of the phenomena that would be nowadays labelled morphomic. Wurzel (1989:30), for example, proposed a definition of the morpheme which “does not demand that a uniform meaning be assigned to the segment sequence”. In his opinion, an extraphonological property of any sort is sufficient to recognize a morpheme. Thus, he mentions that elements like *-mit* (in verbs like *permit* and *submit*), despite lacking a meaning of their own, should be regarded as morphemes by virtue of their identical behaviour in word formation: *permission*, *submission*, *permissive*, *submissive*. Similar evidence (the inheritance of irregular morphology from a root in the absence of compositionality: *stand>stood*, *understand>understood*, *withstand>withstood*) was presented by Aronoff (1994:28) as evidence for autonomous morphology.

A still earlier, and little-known reference that preceded the re-emergence of autonomous morphology and the morpheme is Janda (1982). There it was argued, for example, that “morphological homophony in languages is too extensive and too widespread to be due to chance” (Janda 1982:185) and also that “a language's system of inflectional and derivational morphology is more highly valued if the same formative appears in more than one word-formation rule” (Janda 1982:190). To account for the facts, Janda advocated for autonomous morphology and entertained the possibility of allowing morphemes to have either a very general meaning or no meaning whatsoever.

The field of Romance philology was, for obvious reasons, especially reluctant to ever fully buy into the notion of the morpheme as always involving a strict pairing of form and meaning. Malkiel (1974:307), for example, already reflected on elements like the *-iss-* in French *fin-issions*, which, he argued, “serve no identifiable purpose”. In the absence of a better term, he seemed to begrudgingly accept calling these elements 'empty morphs'.

Even during its zenith, the problems of the morphemic model were never completely forgotten. Uhlenbeck (1952:326), for example, remained true to the spirit of the classical word-and-paradigm model when he argued that “the morpheme, in contradistinction to the word, is not a linguistic unit [and] only has meaning via a word”. Even before that (Hockett 1947, Harris 1942), there was already a tendency to regard the morpheme more as a grammatical distributional element of form, than as the meaning-bearing unit that the term has come to denote.

This was, undoubtedly, also the spirit of Bazell (1938:365) when he proposed the term 'phonomorpheme' to refer to those situations (e.g. dative and ablative plural syncretism in Latin, or genitive singular and nominative plural syncretism in some IE declensions) where various functions tend to be covered by a single formative. Thus, Bazell's concept of the 'phonomorpheme' predates Aronoff's 'morpheme' by more than half a century but seems to be driven by largely the same concerns.

The idea that grammatical units of some kind can sometimes exist independently of meaning has, therefore, been among us for a very long time. This conviction seems to have been present, whether consciously or not, even amongst the most zealous morphemists like Bloomfield. One can, for example, detect a certain degree of logical dissonance in his famous 1926 paper, where, even after explicitly defining a morpheme³ as a meaningful unit

³ Although it is not my purpose here to comment on the history and meaning of the term 'morpheme' (see Anderson 2015 for such an endeavor) it is appropriate to point out at this point that the extent to which meaning was part of the definition of 'morpheme' has not been stable throughout history. Baudouin de Courtenay (1895 [1972]), thus, coined the term to refer to any atomic subword unit with psychological autonomy. Only later (e.g. in the work of Bloomfield) did the conviction spread that this unit (the morph or formative) needed a meaning (a sememe) of some sort. However, what exactly a possible meaning was (for example whether disjunctive or list-like entries are allowed) was usually not explicitly discussed (e.g. Bloomfield 1943).

(1926:155), Bloomfield uses the same term to refer to the (meaningless) sequence *-end-* present in Latin verbs like *prendere, pendere, rendere* and *attendere* (Bloomfield 1926:163).

Both before and after Aronoff (1994), therefore, abundant evidence has accumulated that some units of grammar are either not about meaning (see Bickel's [1995] notion of the "eideme") or even exist at odds with it. If this is the case, dissociating form and function (see Beard's [1995] so-called 'Separation Hypothesis') may well be the only way of accounting for many of the less "well-behaved" distributions in morphological exponence. Be that as it may, after Aronoff's 1994 monograph called attention to the problem, the literature has fortunately been able to move beyond the theoretical recognition of the problem and into the empirical exploration of the phenomenon. Maiden (2001, 2005, 2011), for example, has done extensive research on the diachronic behaviour of stem alternations in Romance varieties. His research has shown conclusively that paradigmatic affinities that are purely morphological exist, can be extremely resilient, and can even constitute productive units in cases of morphological change.

These empirical investigations have also, in turn, fed theoretical discussion. Because these formal alliances are clearly not just diachronic junk, formal models and mechanisms have been proposed that make it possible to have non-trivial mappings from morphosyntactic features to phonological form. Consider for instance the form and content paradigms proposed by Stump (2001) for Paradigm Function Morphology.

Research around the morphome has been done for over two decades now but there is still no absolute consensus regarding the answer to even the most fundamental questions such as, for example, whether morphomes have a learnability disadvantage over morphemes, and most importantly, whether they exist or not in the first place (see Bermúdez-Otero & Luis 2016, Steriade 2016). One should be sure, that is, that they constitute real categories for language-users and not spurious or accidental formal resemblances. Although most of the morphomic literature, understandably, leans towards a positive answer to the existence

question,⁴ some linguists continue to be very critical of the concept. Embick (2013:5-6), for example, argues against the separation hypothesis and against morphomic stems, and he complains that the whole enterprise does “not hold more theoretical interest than an enumeration of the facts” (Embick 2013:22). Others like Koontz-Garboden (2016) lament the lack of positive diagnostics or empirical predictions in relation to the morphome.

A possible solution to the former may be found in the recently emerging set-theoretic (Stump & Finkel 2013) and information-theoretic (Ackermann & Malouf 2013) approaches to morphology. Blevins (2016:105), for example, proposes to regard morphemes as units of predictive value. Besides predictability-based approaches, some other newly-emerging research paradigms and concepts (such as so-called 'niches' [Lindsay & Aronoff 2013]) as well as some older ones like 'stem spaces' [Boyé 2000, Boyé & Cabredo-Hofherr 2006, Montermini & Bonami 2013], or the principle of 'synonymy avoidance' or 'no-blur' (Carstairs-McCarthy 1994, 2010) also relate to the morphome in ways which are not always entirely appreciated or explicitly discussed.

1.3 Terminology

Despite the increasing appearance of the term in linguistic literature, the concept of the morphome is notoriously confusing. The term and its adjectival form 'morphomic' have been used to refer to various linguistic objects such as meaningless stems, unnatural sets of paradigm cells, inflection classes etc. (for a more exhaustive survey of the different uses see O'Neill 2011:44 and O'Neill 2013:221). These objects' only common property, as far as I can see, is that they could be regarded as autonomous morphological phenomena. The terms 'morphome' and 'morphomic' are also used frequently to refer to a particular formalization, theoretical construct or hypothesis related to these linguistic phenomena (see e.g. Round 2011, Spencer 2016:5, Bermúdez-Otero & Luís 2016, Koontz-Garboden 2016:90 etc.). This polysemy constitutes sometimes a notable hindrance to successful reasoning and dialogue.

⁴ I can already advance that my answer to that existence question will be positive too. I consider that the existence of morphemes has been shown convincingly enough by others before me, most notably by Aronoff and Maiden, and I will thus not be concerned specifically with it in this dissertation.

Fortunately, some contributions have recently spotted the problem and have proposed some solutions.

Smith (2013) distinguished between what he called ‘class morphemes’ (i.e. inflection classes) and ‘paradigm-subset morphemes’. Yet another contribution to terminological clarification is Round (2013). In his attempt at distinguishing the various senses of the terms ‘morpheme’ and ‘morphemic’ in the literature, he coined the terms ‘rhizomorpheme’ (for inflection classes), ‘metamorpheme’ (for sets of paradigm cells characterized by common exponents) and ‘meromorpheme’ (for the actual forms that reveal a metamorphemes). Illustrating the referents of these terms with an example familiar from the morpheme literature will help clarify what each term refers to:

	<i>venir</i> ‘come’		<i>nacer</i> ‘be born’		<i>caber</i> ‘fit’	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	ven-g-o	ven-g-a	naθ-k-o	naθ-k-a	k-ep-o	k-ep-a
2SG	vienes	ven-g-as	naθes	naθ-k-as	kabes	k-ep-as
3SG	viene	ven-g-a	naθe	naθ-k-a	kabe	k-ep-a
1PL	venimos	ven-g-amos	naθemos	naθ-k-amos	kabemos	k-ep-amos
2PL	venis	ven-g-ajs	naθeis	naθ-k-ajs	kabejs	k-ep-ajs
3PL	vienen	ven-g-an	naθen	na-θ-k-an	kaben	k-ep-an

Table 2: L-morpheme in Spanish (shaded cells)

The lexemes *venir* and *nacer*, for example, belong to two different rhizomorphemes by virtue of their inflecting in different ways (contrast e.g. ven-i-mos vs naθ-e-mos). A rhizomorpheme, thus, would be a set of lexemes that inflect in the same way. Much like gender, they are partitions of the lexicon, but partitions without extramorphological effects. Because, in my opinion, inflection classes are a phenomenon quite different from the other ones referred to by the term ‘morpheme’, the two can be explored with relative independence from one another. This dissertation, therefore, is not going to be concerned with inflection classes or rhizomorphemes.

More subtle is the distinction between the other two notions. A metamorphome, represented above by the reknown L-morphome, is a set of paradigm cells which behave, within a given lexeme, in the same way regarding some morphological aspect. In the case of this particular metamorphome, it encompasses the 1SG present indicative and all the present subjunctive cells. However, as the verbs in Table 2 illustrate, the forms that reveal the metamorphome can be diverse. In the case of the verbs *venir* and *nacer*, the L-morphome cells share a /g/ or /k/ velar extension (i.e. /ven/ > /veng/, /naθ/ > /naθk/) to the stem which is found in other cells. In the case of the verb *caber*, these cells (and only these cells) have a mildly suppletive stem alternant (i.e. /kab/ > /kep/). These alternations are reminiscent of other morphological oppositions in other languages, for example, Eng. *kill* > *killed*, *think* > *thought* etc. There is, however, a fundamental difference between these two: one opposition (English) has a straightforward semantic/functional correlate (past), whereas the other (Spanish) does not and appears to be morphosyntactically arbitrary.

Distinguishing between formal elements (e.g. /g/) or operations (e.g. 'add /g/') and the set of morphosyntactic contexts where these apply (e.g. 1SG.PRES.IND+PRES.SBJV) is sometimes necessary for clear argumentation. These two senses are, however, two sides of the same coin. The unnatural set of contexts or paradigm cells that often share a morphological affinity may be termed metamorphome while the term 'meromorphome' is preserved to denote the actual form(ative)s which reveal the existence of the metamorphome in the first place. In the examples above, the stem augments -g- (in *venir*) and -k- (in *nacer*), and the stem change -ab- > -ep- (in *caber*) would, thus, all be 'meromorphomes', that is, the pieces of form whose unnatural yet identical morphosyntactic distribution we would like to account for in some principled way.

The question to be asked is whether we need to distinguish terminologically between a form and its distribution. The tendency of the last decades in the morphemic literature has been to conflate the two, so that the erstwhile notions of 'morph' (a unit of form) and 'sememe' (a unit of meaning) have been both increasingly referred to as 'morpheme'.

Most authors in the morphomic literature (e.g. Smith 2013 or Stump 2016:175) have indeed made no terminological distinction between the meta- and the meromorphome. The two

concepts are, obviously, intimately linked, since one cannot exist without the other.⁵ I believe, in addition, that the possibilities for actual confusion of the two senses are very limited when used in context. A terminological distinction between meromorphome and meta-morphome could therefore do more harm than good. On the one hand it would empty the original and better-known term 'morphome' of any content. Alternatively, it would demote the term to denoting just a hyperonym of all autonomous morphological phenomena, which is something that, as far as I can see, we do not need a term for. On the other hand, distinguishing meta- and meromorphomes would introduce new terminology that would (unnecessarily) degrade the readability of morphomic literature for outsiders. Consequently, I will not adopt throughout this thesis Round (2013)'s terminology and I will stick to the traditional terms 'inflection class' to denote a set of lexemes that inflect in the same way, and 'morphome' to refer to unnatural systematic syncretisms, both on their form and their meaning side.

A sense of the term which I believe can occasionally come in the way of clear discussion is the use of the term 'morphome' to denote a particular formalization of a given morphological pattern or a theoretical hypothesis about morphological architecture. I would like to draw attention here to the fact that, although description and analysis are more closely intertwined in linguistics than in most other sciences, the two can and sometimes should be distinguished. To put a parallel example, the term 'syncretism' usually refers to the “thing in the language” regardless of its formalization. The possible ways of formalizing or analyzing syncretism (e.g. as “underspecification”, with a “rule of referral” etc.) are referred to by dedicated terminology, which prevents sloppiness in argumentation. I believe there is a strong need to introduce a similar distinction in the morphome.

Similarly to syncretism, one can sometimes simply “observe” recurrent elements of form in a language whose distributions are not conjunctively definable (by some measure). We may call this as we please (e.g. unnatural syncretism, morphome, homophony etc.) but this descriptive

⁵ Sometimes, for example in the Kayardild case/tense markers that Round (2013) discussed, systematic morphological affinities can be found between formatives in different word classes. In these cases, of course, meromorphomes single out cells in different paradigms (e.g. FUT+DAT) rather than within a single lexeme's paradigm (e.g. 1SG+2PL). A terminological distinction between inter- and intraparadigmatic morphological affinities might, indeed, be useful (see Section 2.4) but has not been proposed as far as I know.

entity should ideally be distinguished from its more sophisticated theoretical analysis, which might involve, for example, positing a purely morphological component of grammar, or an underlying distribution different from the one observed in surface, or arguing that there are in reality two or more elements that just happen to be homophonous. A terminological clarification would be, therefore, most welcome in this respect since, currently, ‘morphome’ and ‘morphomic’ denote both a morphological phenomenon and a particular theoretical stance or analysis.

1.4 Object of study

Since this dissertation is mostly empirically-oriented, the term ‘morphome’ will be used here most often in its near-observational formal-identity sense and not to refer to its higher-level theoretical analysis. The reason to focus on this sense of the term is straightforward. If we want to make any claims or empirical discoveries about the morphome, it has to be possible to define it and identify it in a language in a way that does not hinge upon a particular formal analysis. For this reason, in the context of typological investigations like this one, concise working definitions of the object of study could well be sufficient initially. Trommer (2016), for example, defines a morphome simply as:

Morphome: *“a systematic morphological syncretism which does not define a (syntactically or semantically) natural class.”* (Trommer 2016:60)

This is the kind of definition which I consider most appropriate for a typological investigation.⁶ A definition such as this one could make it possible for us to agree on the morphomic or nonmorphomic status of particular exponents when we see them, provided we had clear criteria for recognizing i) syncretisms and ii) natural classes, and that we operationalize iii)

⁶ In the context of more theoretically-oriented disquisitions, a different definition might well be called for. Spencer (2016:7), for example, proposes that “An expression E is morphomic_{strict} iff E does not consist of a pairing of a form and a (natural) class of grammatical properties (feature-value pairs); E does not alter the set of grammatical properties (feature-value pairs) in the representation of a word form; E does not serve as the realization of any grammatical property set (set of feature-value pairs)” It is clear why this definition would be unsuitable for a typological investigation. Outside a particular theoretical framework there is no way to tell if an expression “alters the class of grammatical properties” or “realizes a property set”.

“systematicity” in some way. Because a consensus on these is woefully lacking, I will set next to put these matters to rest.

Syncretism is a term with a long history and has been widely adopted by morphologists. This does not mean, however, that everyone uses it in the same way. One can find even antagonistic definitions of what a syncretism actually is. For Haspelmath & Sims (2010:174), a morphological identity counts as syncretism (as opposed to accidental homophony) only if the formally-indistinguishable values constitute a natural class. By contrast, Boyé & Schalchli (2016:208) argue that we should only recognize a syncretism when forms are the same “for contexts not belonging to a natural class”. This shows clearly the importance of homogeneous terminology and of agreeing upon our definitions and grammatical primitives. Most morphologists (e.g. Baerman et al. 2005) do not make any reference to the (un)naturalness of the pattern when defining what a syncretism is. I will follow that usage here and will use the term ‘syncretism’ to refer to the (total or partial) morphological identity between different morphosyntactic values (e.g. 1 and 3) or conjunction of values (e.g. 1PL, 2SG, and 3SG) within a paradigm.

What counts as a **natural class** is an even more controversial matter, as this is dependent on feature structure and morphological architecture, theoretical aspects on which there is no consensus. Unlike most extant formalisms suggest and/or allow, this is most likely to be a gradient matter (see Section 3.2). That is, it is unlikely that one can say in a dichotomical fashion whether two values (e.g. accusative and dative), simply form or don't form a higher-level class. It is more likely that, quite on the contrary, the semantic and morphosyntactic affinity between any two particular values may be stronger or weaker:

	A) most natural			B)			C)			D)			E)			F) least natural		
	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL
1	■			■			■			■			■			■		
2	■						■						■	■				■
3	■									■							■	

Table 3: Some paradigmatic distributions ordered for their naturalness

In the above Table 3, pattern A) and pattern B) are unmistakably natural because they can be captured with reference to a feature value 'SG' or to a conjunction of values '1SG'. Pattern F) is the farthest from a natural class and thus the most unmistakably morphomic. The intermediate configurations could be considered natural or unnatural (or a possible or impossible meaning for a lexical entry) depending on the particular researcher and framework (see Section). Because of this, my approach here will not be to take a particular immutable feature structure as the standard to taxonomize individual cases as morphemes or morphomes.⁷ Instead, I will usually present patterns on a scale of variation according to the plausibility or implausibility of a natural-class analysis. Apart from avoiding arbitrary taxonomies, a beneficial side-effect of this approach will be that the findings will be largely theory-neutral and as such 'usable' by researchers of various theoretical convictions, which is specially convenient given the controversial nature of the morphome as a whole.

Having clarified the notions of syncretism and natural class, it must still be clarified how can the **systematicity** of a pattern be assessed. There are many ways to understand 'systematic' or its opposite 'accidental'. The term could apply to a pattern's diachronic origin or to its synchronic status in the language (see Section 2.1 for a more extensive discussion). Straightforward grammatical tests for the cognitive relevance of a syncretism are woefully lacking (see Section 2.1.1) and we cannot look into the brain of language users either to see if they regard a particular syncretic pattern as involving two homophonous elements or just one marker. And yet, if we study morphomes we would like to make sure somehow that we are analyzing morphological single units of some sort, that is, generalizations that the language users spot and abide by and not instances of mere homophony.

As mentioned by Aronoff (2016), a polyvalent morph by itself does not provide any evidence for systematicity. For example, the fact that 3SG and 2PL agreement in German are expressed with the same suffix -t could well be a quirk of the language that is not exploited by native speakers in any way. They could perfectly well have learned the pattern as two different elements: a -t₁ triggered by 3SG subjects and a -t₂ triggered by the 2PL. Thus, to be on a safer

⁷ This does not hold for inclusion of a morphological pattern into the synchronic morphome database in Section . In order to minimize subjectivity there, clearcut criteria will be laid out in that section to make consistent dichotomous judgements on morphomhood (i.e. morphome or not-morphome).

ground, it is usually not until an unnatural distribution is replicated with different forms that morphologists recognize a morpheme (consider also the economy considerations in Section). This is the case, for example, of the following morphological identity in Udmurt:

	1 st conjugation Indicative suffixes				2 nd conjugation Indicative suffixes			
	PRES		FUT		PRES		FUT	
	SG	PL	SG	PL	SG	PL	SG	PL
1	-iško	-iškom	-o	-o-m	-ško	-škom	-lo	-lo-m
2	-iškod	-iškodi	-o-d	-o-di	-škod	-škodi	-lo-d	-lo-di
3	-e	-o	-o-z	-o-zi	-∅	-lo	-lo-z	-lo-zi

Table 4: Inflectional suffixes in Udmurt (Uralic) (Csúcs 1988:142)

The sharing of form by the 3PL present and by all the future forms is repeated in the two conjugations of the language and with different formatives: -o, and -lo. This fact provides a stronger evidence for the induction of the generalization that those values indeed share the same exponent. Such a generalization would also allow an Udmurt language user to make reliable inferences concerning the presence of these forms in the paradigm (e.g. a 3PL.PRES in -o implies a 1SG.FUT in -o and *vice versa*). It is thus safer to require that an unnatural morphological pattern be repeated before classifying it as a morpheme. This is a criterion I will adopt throughout this dissertation, particularly in Chapter 5.

Having provided the literature background, clarified terminology, and provided a working definition of the object of enquiry of the present research, this dissertation will seek a deeper understanding of the phenomenon of morphomicity. In order to do this I will deal, in the coming Chapter 2, with several problematic issues around the definition of morphemes and their identification in a language. Before I do so, however, I would like to point out that, although the very existence of morphemes may still not yet be an undisputed *communis opinio*, I believe that the objections to morphemes' existence are more theoretical and philosophical (i.e. regarding what to say about them or how to best analyze them) that empirical in nature at this point. However one may wish to conceive or formalize them, it is my conviction that a greater empirical understanding of unnatural morphological patterns will be valuable for both defenders and detractors of autonomous morphology.

2 Issues on morpheme identification

2.1 Morphemes and accidental homophony

As noted by many theoretically-inclined linguists, “[a] recurrent problem in linguistic analysis is the existence of multiple senses or uses of a linguistic unit” (Haspelmath 2003: 1). The difficult point is, usually, to distinguish cases of polysemy, which are generally regarded as grammatically significant formal identities, from cases of so-called accidental homonymy, which are frequently dismissed as irrelevant to grammar and uninteresting.

Some of the criteria which are usually employed for grounding this distinction are semantic relatedness and cross-linguistic comparison. If the meanings expressed by a given formal element are completely unrelated and/or if they are not usually found outside a particular language or language family, the formal identity is taken to be accidental and hence irrelevant for grammatical theory. This is, for example, usually held to be the case of English plural and genitive -s (e.g. Haspelmath 2003: 5).

It is, however, necessary to think twice about the meaning of “accidental”. Should it apply to the presence of a particular formal identity in a particular language? Should it be understood as referring to the (cognitive) status of that formal identity? It seems unlikely for all the possible tell-tale signs for “accidentality” to converge. If the first criterion is used, the English plural-genitive formal identity should probably be described as accidental, since cross-linguistic comparison does not seem to provide evidence for an above-chance-level formal identity of those meanings. However, when one considers the facts of the English s-exponent, there may be some reasons for concern. It is true that without s-containing exponents for plural and genitive in the ancestral stages of English, that identity would probably not have arisen. However, there is diachronic evidence that this formal identity may have been sought and reinforced. Both s-plurals and s-genitives were preferred over alternative strategies and were progressively extended to become fully productive (compare to the situation in related languages like German or Icelandic). The same thing happened, yet again, when 3SG -s achieved supremacy over -th in verbal inflection.

All of these morphosyntactic contexts share an identical distribution of allomorphs (/s/, /z/, /iz/). Furthermore, when these formatives occur together, one suffices to express both meanings (e.g. tigers'). This constraint is not merely a phonologically-triggered haplology but can be shown to have grammatical import (consider the ungrammaticality of **the kings of England's crown*). In addition, other morphs with the same form and syntagmatic locus occur elsewhere in the grammar as clitic versions of *has* and *is*. These have also taken the upper hand over alternative conventions ('tis>it's). It is, in my opinion, quite striking that so many different functions have come to be expressed by the exact same form, especially given the scarcity of morphology in the English language. All in all, therefore, the synchronic and diachronic facts of English would seem to favour, if anything, a non-accidental understanding of this particular formal identity.

Be that as it may, from the perspective of the morpheme it is of course an unwarranted aprioristic assumption to always regard as grammatically uninteresting those formal identities which lack cross-linguistic generality or any morphosyntactic or semantic features in common. If some of those patterns are shown to be systematic within a given language, then they must surely deserve attention and inform our morphological models.

A morphosyntactically coherent exponent (for example, one which occurs across all 1PL verb forms, like /mos/ in Spanish) can usually be classified as systematic by virtue of this morphosyntactic coherence alone so that no other proof must be offered to support the relevance of the formal identity of e.g. *crece-mos*, *crezca-mos*, *ande-mos*, *tuvi-mos*, *amare-mos* etc. However, this straightforward diagnostic is unsuitable for morphomic exponents because, by definition, they lack a morphosyntactically coherent description. Evidence for the non-accidental character of a formal identity, therefore, must in these cases be sought somewhere else.

2.1.1 Tests for the grammatical relevance of a formal identity

It would be ideal to have a hard-and-fast (e.g. syntactic) test to ascertain whether two formally identical elements are also 'the same' at some deeper grammatical level. Some such tests have sometimes been proposed:

a) Feature conflict resolution

As discussed by Zwicky (1991), in some cases, but crucially not always, a syncretic form has the ability to resolve a conflicting morphosyntactic requirement. Because of this, Zwicky suggested using this test to distinguish accidental homonymies from systematic identities:

1) *Entweder wir oder sie spielen gegen Bulgarien.*

‘Either we or they will play in the Bulgaria match.’

2) **Entweder Bierhoff oder ihr spielt gegen Bulgarien.*

‘Either Bierhoff or you(pl) will play in the Bulgaria match.’

The above contrast, presented in Haspelmath & Sims (2010:175), therefore, would suggest a systematic status for the formal identity of 1PL and 3PL verb forms in German but an ‘accidental homophony’ status for the identity of 3SG and 2PL. This seems intuitively appealing because the former two forms are always whole-word syncretic whereas the latter are not (contrast e.g. 3SG *fährt* ‘drive.3SG’ and *fahrt* ‘drive.2PL’). However, it is not difficult to find limitations that severely compromise the usefulness and validity of this test. For example, we are likely to sometimes not find any construction which could be used to induce the required feature conflict. In addition, the test is unsuitable for formal identities smaller than the whole word (e.g. when only the stem or only an affix are formally identical). For a fuller discussion of the test and its limitations, see Johnston (1996:13-14).

b) Co-occurrence restrictions

In an ideal world we would not expect the same formative to appear two times in the same word or domain. This is, patently, not always the case (consider the phenomenon of multiple exponence, see Harris 2017). In principle, however, we could expect it to remain a strong universal tendency for a given morphosyntactic feature specification to be expressed only once in a word. Formatives which are ‘different’, by contrast, are expected to be able to co-occur, provided that they are semantically compatible. One could, thus, attempt to use co-

occurrence restrictions as tell-tale signs of the accidental vs systematic formal identity of different formatives.

There is a suffix in Turkish, for example, (*-miş/-miş/-muş/-müş* depending on vowel harmony) that has both perfect and hearsay uses (Slobin & Aksu 1982). The two uses are very likely historically related. However, they are synchronically compatible semantically and the two can co-occur within a single word, suggesting that they should be considered two different elements at a deeper level, rather than one single formative with broad (or complex) modal-aspectual semantics:

3) Kemal gel-miş-miş

Kemal come-PERF-EVID

'(It is said that) Kemal had come' (Slobin & Aksu 1982:194)

Another Turkish suffix *-lar/-ler* can mark the plural of a noun and the plural of a third person possessor (Stump 2015:176). That is, *adam-lar* (man-PL) means 'men', and *adam-lar-ı* (man-PL-3) means 'their man' (consider also *adam-ı* (man-3) 'his/her man'). Although semantically the two uses could be compatible, in order to express 'their men', instead of the expected **adam-lar-lar-ı* the form *adam-lar-ı* is used instead, which is thus three-way ambiguous:

	Possessor 3SG	Possessor 3PL
Possessee SG	adam-ı	adam-lar-ı
Possessee PL	adam-lar-ı	adam-lar-ı

Table 5: Turkish noun number and possessor (Stump 2015: 176)

Simultaneous use of the two *-lar* is impossible, thus suggesting some inherent grammatical incompatibility, maybe because they are analyzed by the language user as one and the same element. This might be attributed to the semantic affinity [plural] of the two uses.

This co-occurrence test could, therefore, provide evidence for the analysis of these polyfunctional elements as one complex vs multiple homophonous elements, thus helping us distinguish accidental from systematic formal identities synchronically. However, there are

also severe limitations to the validity and applicability of this test. First, as in the previous one, in many cases there might simply not be a word or construction in the language where the two elements could potentially appear side-by-side. Second, the phenomenon of Obligatory Contour Principle, as usually portrayed (e.g. Yip 1988), constitutes an occasional obstacle to the appearance of phonologically identical contiguous sequences. This may be independent of grammatical considerations, and the effects of phonological and grammatical identity would be, in most cases, difficult to distinguish. Last and most importantly, like the English genitive-plural /z/ incompatibility suggests, it is unclear whether or to what extent incompatibility between two formatives correlates with a greater syntactic, semantic or general cognitive proximity between them.

After surveying two tests that might have served as potential sources of evidence for/against systematicity, the conclusion is that, unfortunately, there is none that we can reliably apply to obtain independent evidence for the cognitive status of a morphological affinity. Other clues must therefore be sought. Evidence for systematicity within a given language may be plausibly sought from sources such as i) diachronic developments (e.g. analogical changes), ii) allomorphic variation, or morphophonological processes affecting all the forms in the same way, iii) evidence for a morphosyntactic rationale in the distribution and iv) general robustness of the pattern.

i) Diachronic evidence for non-accidentality

Cases of formal identity which have come about solely as a result of regular blind phonological change provide no evidence concerning whether speakers regard those identities as grammatically significant or not. However, those cases of formal identity which are reinforced, extended or created by means of speakers' analogical changes must surely be regarded as systematic. That is at least the prevalent opinion in the Romance morphomic literature. Well-known examples of formal identities which are reinforced and extended are the N- L- or PYTA morphemes of Romance languages (see e.g. Maiden 2011a).

One could think that morphological changes giving rise to morphomic exponents from scratch might be rare. However, cases of what might be called 'gradient attraction' (Burzio 2001), whereby forms that are similar become still more similar or identical, are not rare.

One example, evidence for which comes from the comparison with related languages, is the case of Acazolco Otomí, discussed by Baerman et al. (2017), which analogically modified the form of a suffix to match that of (what was) an unrelated suffix with a different paradigmatic distribution. Another example of 'gradient attraction' is the formal identity in Modern English of the nominalizer and progressive -ing suffixes (cf. 'The *beating* he got was dreadful' vs 'They are *beating* him badly'). The formal identity of these two uses cannot be attributed solely to phonological change (compare to the forms in Old English: *-ing/-ung* vs *-end*). Diachronic morphological-analogical developments like these (see also Section 4.1.3.2), could thus be taken as diagnostic of the cognitive relevance of some formal syncretisms, and of a deeper affinity between these forms.

ii) Allomorphic or morphophonological evidence for non-accidentality

For many languages there is unfortunately not enough diachronic or comparative data to work with. However, synchronic grammar may sometimes also provide evidence for non-accidentality. Consider the following Spanish verbs: *conducir* 'drive', *reducir* 'reduce', *inducir* 'induce', *seducir* 'seduce' etc. There isn't synchronically any verb with the form 'ducir' and the verbs do not have any obvious semantic affinity. If these were all the facts we may have had to conclude that the formal similarity in the form of these verbs is accidental and grammatically moot. However, all of them are subject to the same phonologically unmotivated alternations in inflection and word formation: *conduzco* 'I drive', *conduje* 'I drove', *conducción* 'driving'. It is hard to believe that every verb ending in *-ducir* (and only those in *-ducir*) is independently and by chance subject to these same operations.

The alternative explanation is that speakers do posit, on the basis of form alone, a grammatical unity at some level despite the lack of shared semantic content. This is not to

say that this unity cannot sometimes be eroded as it can happen to any other grammatical item. The verb *seducir*, for example, is more prone to losing some of these alternations (e.g. *seducí* 'I seduced' instead of *seduje*) because, unlike *con-*, *re-* or *in-*, *se-* is not a recurrent prefix in Spanish. This fact may make it more difficult to identify an element *-ducir* in *seducir* than to identify an element *-ducir* in *inducir*. Kayardild's morphemes (see Round 2013), similarly, also give evidence of their non-accidental nature by means of the morphophonological processes and allomorphic variation they are subject to in the various morphosyntactic contexts in which they appear in the grammar.

Morphological affinities can be detected (and can be repeatable with different exponents) between lexemes (Spanish *-ducir*), between inflectional affixes in different parts of speech (Kayardild), and of course also between different paradigm cells within a single lexeme's paradigm, as in the most renowned Romance morphemes:

	<i>venir</i> 'come'		<i>nacer</i> 'be born'		<i>caber</i> 'fit'	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	ven-g-o	ven-g-a	naθ-k-o	naθ-k-a	k-ep-o	k-ep-a
2SG	vienes	ven-g-as	naθes	naθ-k-as	kabes	k-ep-as
3SG	viene	ven-g-a	naθe	naθ-k-a	kabe	k-ep-a

Table 6: L-morphome "allophones" in Spanish, partial paradigm

The replication of a particular morphological affinity with different forms, thus, is taken to decrease the likelihood of the affinity being just an accident.

iii) Morphosyntactic evidence for non-accidentality

Due to their morphosyntactically well-behaved nature, the systematicity of run-of-the-mill morphemes is not usually questioned. For example, /mos/ appears at the end of every 1PL verb form in Spanish, which seems systematic enough to not require further proof. Morphemes are, by definition, not reducible to morpho-syntactic determination. However, this is not the same as positing complete orthogonality to morphosyntax. Relatively

prototypical examples of morphemes (e.g. the Romance L- and N-morphemes), in fact, often obey some morphosyntactic rationale, as concrete formatives can be limited to specific tenses (e.g. the present) and in this way, they could be argued to 'mean' at least that.

Some languages have other kinds of exponents whose distribution cannot be determined by morphosyntactic features alone but which still are in some way constrained by them. Cases of so-called polyfunctionality (Stump 2015:229), but also cases of deponency, illustrate the capacity of the same morphological forms to be used for more than one purpose. In the case of Noon (see Stump 2015:235), for example, a similar set of affixes is used, in different grammatical categories with different but related meanings. The suffix *-ríi*, for example, can code a 1PL.EXCL object or a 1PL.EXCL possessor. The morpho-semantic core is clear but is not enough to delimit the distribution of those forms. A somewhat different case is that of Nuer nominal inflectional morphology (see Table 18 and Baerman 2012), where some suffixes seem to have a problematic morphosyntactically distribution, as this changes from one lexeme to another. However, looking across all paradigms, the range of particular suffixes appears to be limited to natural morphosyntactic classes (*-ni* to the plural, and *-kä* and *-ä* to the oblique singular).

As explained in the above cases, thus, although perfect morphosyntactic determination is definitionally impossible in morphemes, a morphosyntactic rationale may still be offered as proof for systematicity in some cases.

iv) Other tell-tale signs of systematicity

The general robustness of a pattern of morphological identity, measured as either the amount of shared (segmental or suprasegmental) phonological material, the proportion or number of lexical items displaying the pattern, the number or proportion of exceptions to the pattern etc. could all be potentially used to evaluate the systematicity of a morphological pattern. However, these and others will be considered independent variables of cross-morphomic variation (see Section 5.4) and will be allowed to vary freely.

2.1.2 On the empirical status of homophony and polysemy

As mentioned before, much of the literature regards the phenomenon of the morpheme as necessarily involving cognitive reality and not simply formal identity. Consider, for illustrative purposes, the following data from Basque:

Future

4) Leioha ireki-ko dut (I will open the window)

window open-FUT have.1SG

5) Madrilera joan-go naiz (I will go to Madrid)

to.Madrid go-FUT be.1SG

6) Horrela egin-en dute (They will do it that way)

thus do-FUT have.3PL

Genitive

Hiri-ko atea (The door of the city)

city-GEN door

Irun-go neska (The girl from Irun)

Irun-GEN girl

Mikel-en aita (Mikel's dad)

Mikel-GEN dad

Future and genitive suffixes in Basque are identical and share many of their allomorphic and morphophonological traits. On this evidence, one may wonder whether we should describe these situations as one element with an unnatural distribution (i.e. one *-ko* which may appear in genitive and in future contexts) or two homophonous elements (i.e. a genitive suffix *-ko₁* and a future suffix *-ko₂*). Many linguists seem to think it is important whether these situations are perceived by language users as different elements or as just different uses of a single element. Although it is likely to be much more complicated than a simple dichotomy, these two scenarios have come to be labelled 'polysemy' and 'homophony' respectively (see e.g. Panman 1982, Klein & Murphy 2001).

As I mentioned, a lot of effort has been devoted to answering this polysemy vs homophony question in specific cases (see e.g. Harbour 2008). However, one could wonder whether all these discussions are worth having. In the end, even if we accepted, for example, that there is just "one *-ko*", language users would still have to know in which specific contexts to use the form. Is that any different, ontologically, from saying that there are two *-ko*? Or conversely, is saying that there is a *-ko₁*, and a *-ko₂* any different from saying that there is one *-ko* element

with a complex distribution? Are these decidable statements like the ones science is supposed to deal with? Or is it merely an analytical preference of the linguist with no extra-theoretical bearing?

Language is an idiosyncratic object of study in that it exists exclusively in the mind of language users. Because of this, it is very hard, if not impossible, to separate facts from analysis. Human beings inevitably have to analyze their language input (i.e. posit some categories, make certain analytical choices etc.) to make some sense of it and be able to use language productively. It is this very analysis that constitutes their grammar of the language. Because of this, phenomenon and analysis are not genuinely different things in linguistics. The analysis of the (native) language user constitutes the phenomenon itself and should be the object of study.

This does not mean, of course, that the analyses of linguists will always match those of language users. On the contrary, it is often the case that linguists' analyses are not interpretable outside some particular theoretical framework or even that they are completely divorced from language users' intuitions and from (some of the) available data. When this happens, that is, unquestionably, unfortunate. Consider, by way of example, the following agreement patterns with some Spanish nouns:

7a) **la** costa(F) peligros-a

The coast dangerous-F

7b) **las** costas peligros-a-s

the coasts dangerous-F-PL

8a) **el** arma peligros-a

the weapon dangerous-F

8b) **las** armas peligros-a-s

the weapons dangerous-F-PL

9a) **el** tema(M) peligros-o

the issue dangerous-M

9b) **los** temas peligros-o-s

the issues dangerous-M-PL

The traditional account of this phenomenon is that the form *e/* in 8a) is not “the same thing” as that in 9a) and that they just happen to be accidentally homophonous. The 'official' analysis (see e.g. RAE-ASALE 2009: 23, 265-267) argues that *e/* is, in contexts like 8a), merely an

allomorph of *la*, the usual feminine singular article, seen in 7a). It is supposed to be a phonologically triggered allomorphy that occurs in 8a) because the following noun begins with a tonic /a/. The nouns which trigger this form are, indeed all of that phonological form (e.g. *alma* 'soul', *águila* 'eagle', *agua* 'water', *hambre* 'hunger', *ala* 'wing', *aula* 'classroom' etc.) and it is clear that the phenomenon did originate from some differential treatment of the feminine article in these phonological contexts.

However, there is sufficient synchronic evidence that this is no longer the analysis of (most) language users, which regard the 'el' of 8a) as a genuinely masculine form synchronically and not as a phonologically determined allomorph of the feminine. This is supported by various facts. For example, it is just nouns, and no other grammatical category that trigger this allomorphy (e.g. *la alta torre* [*el alta torre], *la hábil secretaria* [*el hábil secretaria]). Even in nouns, the allomorphy is not triggered by every single noun starting with tonic /a/ (e.g. *la Ana* [*el Ana], *la A* [*el A]). In addition, we observe that the use of a masculine agreement form in these nouns is not limited to the definite article but has been gradually extended by speakers to many other morphologically singular elements including the indefinite article (*un/una*), the demonstratives (*este/esta, ese/esa, aquel/aquella*) and even, occasionally, to adjectives and quantifiers, and to articles and demonstratives not immediately before the noun (e.g. “*un hambre tremendo(M)*” or “*un(M) bonito(M) águila*”, which is five times more frequent on the Internet than the 'correct' “*una bonita águila*”). All of these developments, which happen even despite linguistic prescription, would make absolutely no sense if language users indeed regarded the article of “*el águila*” as feminine.

The formal convergence of the feminine article before tonic /a/ with the masculine, and its divergence from the more usual feminine article, may have been a more-or-less fortuitous outcome of sound change (**ela kasa* > *la=kasa*, **ela alma* > *el=alma*). However, once it happened, language users had the understandable impulse to associate the form with other *el* rather than with other *la* and the nouns taking *el* with other (masculine) nouns taking *el*. In the noun *arte* 'art', interestingly, the whole set of agreement targets has been changed in the singular so that in the modern language, *arte* is usually described as masculine in the singular and feminine in the plural.

This case illustrates at least two things. The first is that linguists' explicit analysis of a phenomenon does not always coincide with the way in which language users implicitly analyze it. The second is that speakers usually prefer to analyze sameness of form as sameness of function, a fact which is sometimes questioned (see e.g. Harbour 2008). Form, along with meaning, is evidence of the utmost importance for language users' construction of their grammars and should therefore be given the utmost consideration by linguists.

Concerning linguistic analysis, therefore, it is not the case that 'anything goes'. If our goal is to understand language, we should aim at understanding language users' grammatical system. Even if this is really difficult in practice, we should not be satisfied with an analysis or formalization that simply mimics speaker performance. Because of this, I argue that it is indeed a relevant distinction, in linguistics, whether the *e/* in 'el arma' is the usual masculine singular article, a feminine singular allomorph of *la*, or something else entirely. It is, therefore, important whether some pattern of identity is cognitively relevant, i.e. part of the grammatical system of native speakers, or merely reflects the inert outcome of some historical accident.

On the other hand, even if we currently lack direct access to the mind of language users, there seems to be experimental evidence that the homophony vs polysemy distinction is, indeed, a cognitively real one. Pykkänen et al. (2006) for example, found that there are noticeable differences in the speed at which polysemous and homophonous pairs are processed. This suggests that the difference that linguists intuitively sense between these pairs is not a mere illusion.

There are many diachronic changes which can also be offered as evidence that whether or not language users make a generalization over two forms is, indeed, of the utmost importance. Most revealing, in my opinion, are those cases where an originally single lexeme splits into two. This may happen in two main scenarios: i) when the meanings of a single lexeme become too different or ii) when the forms of a single lexeme become too different.

i) Semantically-motivated split⁸

The Spanish verb *saber* can mean both 'know' and 'taste'. Under both senses, it is a descendant of Latin *sapere*. Because of this, prescriptive grammarians insist that it should be conjugated in the same way (*sé, sabes, sabe* etc.) regardless of its meaning. This, however, does not match the intuitions of many language users. Under the meaning 'taste' the verb is used almost exclusively in the third person, but when native speakers produce the rest of the forms, these are often: *sepo*, (e.g. *yo sepo salado* 'I taste salty'), *sabes, sabe* etc. The 1SG present form, thus, may differ from the one found under its sense 'know':

	<i>saber</i> ₁ 'know'				<i>saber</i> ₂ 'taste'			
	Present Indicative		Present Subjunctive		Present Indicative		Present Subjunctive	
	SG	PL	SG	PL	SG	PL	SG	PL
1	sé	sabemos	sepa	sepamos	sepo	sabemos	sepa	sepamos
2	sabes	sabéis	sepas	sepáis	sabes	sabéis	sepas	sepáis
3	sabe	saben	sepa	sepan	sabe	saben	sepa	sepan

Table 7: Spanish present tense forms of *saber* 'know' (left) and *saber* 'taste' (right)

It seems, thus, that a morphological change has occurred from the original paradigm *saber*₁ to that of *saber*₂. The most obvious explanation for the change is that, when the two main senses of *saber* drifted sufficiently away from each other, language users ceased to make the generalization that they constituted a single lexeme. When this happened, the necessity to have them both inflect by the same paradigm disappeared. Since the first and second person forms of *saber* (e.g. irregular *sé*) were only ever encountered in the input under their meaning 'know', they did not count as evidence for language users' deduction of the full paradigm of *saber*₂. This meant that the first and second person forms of *saber*₂ 'taste', when needed, had to be constructed "online" on evidence exclusive to its sense 'taste' (i.e. third persons and impersonal forms), as well as, more generally, on the evidence of recurrent patterns of allomorphy in Spanish conjugation.

⁸ This section largely paraphrases arguments in Herce 2018a.

It might seem strange at first that an analogical reshaping of the first person singular would not have resulted in the apparently more regular *sabo*. This, indeed, would have resulted in stem alternants (*sab-* vs *sep-*) correlating with natural classes (indicative vs subjunctive). The chosen form, however, makes more sense when one considers the patterns of other verbs:

	saber ₁ 'know'	saber ₂ 'taste'	tener 'have'	conocer 'know'	caber 'fit'	caer 'fall'
1SG PRES IND	sé	sepo	tengo	conozco	quepo	caigo
2SG PRES IND	sabes	sabes	tienes	conoces	cabes	caes
3SG PRES IND	sabe	sabe	tiene	conoce	cabe	cae
1SG PRES SUB	sepa	sepa	tenga	conozca	quepa	caiga
2SG PRES SUB	sepas	sepas	tengas	conozcas	quepas	caigas
3SG PRES SUB	sepa	sepa	tenga	conozca	quepa	caiga

Table 8: Partial paradigms of some Spanish verbs

Unlike *saber₁*, verbs whose stem differs between the third person indicative and subjunctive (e.g. *tiene* vs *tenga*, *cabe* vs *quepa* etc.) consistently have the same form in the 1SG indicative as in the subjunctive. The knowledge of this pattern is what may lead Spanish language users to innovate a form *sepo* over *sabo*.

The analogical reshaping operated from the paradigm of *saber₁* to *saber₂* suggests that these purely morphological patterns (the so-called L-morpheme in this case, see Table 1) do exist as a part of language users' synchronic knowledge of grammar. The stem used for '1SG present indicative + all subjunctive forms' cannot be attributed any coherent function and only exists by virtue of the formal relations holding between those cells across paradigms. The fact that this purely morphological solution was preferred to a semantically coherent one suggests that the pattern of root alternations illustrated by verbs like *tener* or *caer* might attract new members under the right circumstances and can hardly be pronounced “dead” synchronically (*contra* Nevins et al. 2015).

ii) Formally-motivated split

Similarly to what happened with the verb *sapere*, a single Old Latin noun *deivos* gave rise to two different lexemes (*dīvus* and *deus*) in Classical Latin (see e.g. Meier-Brügger 2013:89). The noun would have had a uniform stem /deiw/ in Old Latin and would have been declined unproblematically (e.g. genitive *deivī*). However, the loss of /w/ before back vowels /o/ and /u/ and long vowel shortening before another vowel (*deiwos* > *de:wos > *de:os > *deus*) meant that the form of the stem diverged in different paradigm cells:

	SG	PL
NOM	deus	deī < deiwoi
VOC	dīve	deī
ACC	deum	deōs
GEN	dīvī	deōrum
DAT	deō	deīs < deiwois
ABL	deō	deīs

Table 9: Expected paradigm of *deus* (Thurneysen 1887:155)

Undoubtedly because of the resulting formal difference, forms in *dīv-* and forms in *de-* ceased at some point to be interpreted as belonging to a single lexical item and as a consequence, the two forms parted ways definitely as language users innovated the “missing” forms to generate complete inflectional paradigms:

	SG	PL	SG	PL
NOM	deus	deī	dīvus	dīvī
VOC	dee	deī	dīve	dīvī
ACC	deum	deōs	dīvum	dīvōs
GEN	deī	deōrum	dīvī	dīvōrum
DAT	deō	deīs	dīvō	dīvīs
ABL	deō	deīs	dīvō	dīvīs

Table 10: Latin paradigms of *deus* and *dīvus*

The two cases that have been presented in this section suggest that whether or not language-users think of two forms as constituting the same grammatical unit is, indeed, crucial. This even allows to make some “soft” predictions: When a unified cognitive status does not hold, changes that put an end to the surface identity are either not resisted or, in some cases, may even be immediately derived from the loss of the former cognitive generalization.

These lexemic splits also suggest that, as will be argued throughout this dissertation (e.g. in Section 2.4), both semantic-functional distance (in *sapere*) and formal distance (in *deiwos*) can hamper or prevent the induction of a generalization. Thus, the likelihood of a cognitive generalization encompassing two elements increases as a function of their formal and functional similarity.

Whether or to what extent a generalization is drawn or an identity (formal or semantic) is perceived by language users is, unfortunately, not directly accessible to linguists. Already before any change revealed it in the surface, the lexemic unity may have been already broken in the cases presented above. Thus, we cannot always conclude that, in the absence of surface changes, the grammatical unit still holds intact. As linguists or language users, we may have intuitions about whether or not it does. However, as Elbourne (2011:34) points out, “there is no evident reason why intuitions that purport to be about complex internal mental structure (or epistemically inaccessible abstract objects) should be trusted”. It is important, however, to recognize that this fact makes the problem more difficult to solve and not less of a problem. In my opinion, the fact that often “you just can't tell” does not render the whole polysemy vs homophony distinction a figment of the imagination of linguists, but simply a harder nut to crack.

2.2 Morphemes and blocking

The notion of blocking has important ramifications for the definition and identification of autonomous morphological patterns. Blocking is a conflict-resolution principle which is usually assumed (see e.g. Bonami & Stump 2016) to operate between mutually compatible morphemes or realizational rules. It states that, in cases where two rules are in a subset-superset relation, the most specific one will take priority over the more general one. Consider the situation in these paradigm:

	SG	PL
1	war-an	war-aton
2	war-aan	war-ayan
3	war-en	war-an

Table 11: Past tense forms of 'get' in Daga (Dagan, PNG) (Murane 1974:63)

In the case of the unnatural 1SG-3PL syncretism above, an analysis involving blocking is readily available. The suffix *-an* could be posited to “mean” just [past] and to be unspecified for number and person. The reason why *-an* would not surface in other paradigm cells is that other forms (*-aan*, *-en*, *-aton*, *-ayan*) exist that are more specific. The distribution of all the forms, therefore, can be straightforwardly stated as the realization of morphosyntactic properties if we assume blocking. Things can get more complicated, however:

	SG	PL
1	ə-kəft	ni-kəftinə
2MASC	ti-kəft	ti-kəfto
2FEM	ti-kəfc	ti-kəftəma
3MASC	yi-kəft	yi-kəfto
3FEM	ti-kəft	yi-kəftəma
Impersonal	yi-kəf ^w cim	

Table 12: Imperfective tense paradigm of Chaha (Semitic) ‘open’ (Völlmin 2017:122)

In the above paradigm, the morphosyntactic distributions of the prefixes *yɨ-* and *tɨ-* are both unnatural. The two formatives cross-cut, furthermore, and thus none of them occurs in a subset of the other. Without recourse to further formal machinery like rules of referral, a way out would be to say that there are in fact two different *tɨ-* in the paradigm which just happen to be accidentally homophonous (see Harbour 2008). This trick would allow each of the *tɨ-* to have a more specific morphosyntactic distribution ([2] and [3FEM.SG]) that would make blocking possible once again if the other prefix *yɨ-* is assumed to be underspecified.

Whatever the plausibility of this particular solution here, one can easily find in natural language cases where blocking is unmistakably not taking place (see e.g. Janda and Sandoval 1984). Observe the following exponence patterns :

	SG	DU	PL
1EXCL	kei:	kei:-nih	kei:-nih-e
2	na:ng	na:ng-nih	na:ng-nih-e

Table 13: Daai Chin (Sino-Tibetan) personal pronouns (So-Hartmann 2009:140)

	'sibling'		'mother'	
	SG	PL	SG	PL
ABS	is	is-na-ba	išu	išu-bo
ERG	is-t-i	is-na-za	iše-t'-i	iše-t'-za
GEN1	is-t-i-s	is-na-za-s	iše-t'-i-s	iše-t'-za-s
LAT	is-t-i-l	is-na-za-l	iše-t'-i-l	iše-t'-za-l

Table 14: Partial paradigms of two Khwarshi (Nakh-Daghestanian) nouns (Khalilova 2009:66)

In Daai Chin personal pronouns, the plural formative *-e* appears in a subset of the cells where the non-singular formative *-nih* does. According to the blocking principle this should not happen and *-e* should prevent the appearance of *-nih*. To avoid this, it is always an analytical possibility to avoid segmentation in these cases (i.e. to leave *-nihe* as an indecomposable plural suffix, see also Section 2.8).

Sometimes, this might seem an elegant solution but some other times there is no way to salvage blocking without doing violence to the data. In Khwarshi, for example, the oblique plural formative *-za*, is clearly segmentable from previous suffixes but is still sometimes present in a subset of the cells of other more general suffixes, e.g. plural (see *-na* in 'sibling') or oblique (see *-t'* in 'mother'). I see no way in which a paradigm like this could be generated in a world where blocking was an inviolable constraint.

The fact that blocking does not always occur does not necessarily mean, of course, that blocking cannot remain an important tendency in the structuring of paradigms. The problem is that examples which are in conflict with blocking accounts are probably simply difficult to find merely from a probabilistic point of view. As rightly pointed out by Pertsova (2011:241), for example, it is indeed a logical necessity, and not an empirical observation, that when two elements are in a subset relation only the more concrete one can block the other one since if the reverse happened we would not even get to see the more concrete exponent.

Despite its problems, blocking is a mechanism which is usually adopted, with one name or another (Superset principle, Elsewhere condition, Panini's principle, remnant syncretism etc.), by every constructivist theory of morphology. The important question to be asked, from the empiricist's perspective, is whether it is a real cognitive principle of language users or is instead, in the light of the above-mentioned ontological and empirical shortcomings,⁹ just a theoretical liberty that formal linguists make use of to describe certain exponence patterns as realizations of morphosyntactic properties.

There are conflicting opinions in the literature. Blevins (2016:214), for example, criticizes at least certain uses of blocking. In his opinion, in some cases when blocking is appealed to,

⁹ Other instances where Paninian blocking seems to leak are found in those exponence patterns where there seems to be a clear default but also a cell without any overt inflectional formative (see e.g. the attributive adjective inflection of Dutch discussed in Pertsova 2011:241). Although theoretical analyses sometimes rely on zeroes blocking overt exponents in those cases I find it intuitively problematic (and it surely opens the door to all sorts of intractable analyses) to defend that an absence can be blocking the presence of an exponent.

Another morphological fact that one should keep in mind regarding Paninian blocking as usually conceived is that there are also many reasonably clear cases of formatives that i) cannot appear together and ii) should be semantically compatible but are by no means in a subset-superset relation. Consider, for instance the incompatibility of dual subject *-k* and plural object *-dár*, and of durative *-tam* and the masculine object *-rár* suffixes in Nimboran (see Inkelas 1993).

“invoking a notion of ‘rule competition’ appears to misconstrue the problem” and may just be a result of the fact that “the statement of the rules overgeneralizes the distribution of the markers that they are meant to describe”. Bauer, Lieber & Plag (2013:636) go much further when they conclude that “blocking is at best a tendency and at worst a myth”. Pertsova (2011:230), by contrast, and even after being critical with the notion of blocking in other important respects argues that “those paradigms that are easily described by appealing to blocking and underspecification appear natural or systematic to us because of the particular cognitive bias for default reasoning we bring to the task of learning associations between form and meaning”.

Given the deep uncertainties surrounding the status of so-called ‘elsewhere’ forms, I will remain agnostic, throughout this thesis as for whether they constitute exponents different from the ones that cannot be captured by blocking. Because of the empirical focus of this research, ‘surface’ distributions will always be trusted over any supposedly underlying ones.

The same thing holds with respect to rules of referral and any other theoretical or formal mechanism. Even if, according to some, “rules of referral are real for speakers and not just thought up by linguists” (Haspelmath & Sims 2010:179), it is my conviction that a typological investigation should not make use of these notions.

2.3 Morphological zeroes.

It is usually taken for granted that the distribution of formatives deserves analysis and explanation in morphology. The offered explanation may be different depending on whether such elements correlate or not with morphosyntactic categories. Morphological zeroes (see e.g. Mel’čuk 2002), however, represent a rather different case in this respect. Concerns about the analysis of unmarked forms are frequently voiced (e.g. Blevins 1995) and disagreement about the interpretation of these forms continues to this day. Consider the following morphological syncretism:

	SG	PL
1	hembu-n-a	hembu-a
2	hembu-a	hembu-w-a
3	hembu-a	hembu-a

Table 15: Orokaiva (Trans-New-Guinea) far past indicative of *hembu* 'walk'
(Baerman et. al. 2005:26 after Healey et al. 1969)

The morphosyntactic distribution of the form *hembua* above is decidedly unnatural, since the syncretism of 3, 1PL and 2SG constitutes a morphosyntactically arbitrary class. Crucially, however, there is no formative whatsoever whose distribution is problematic. That is, both the stem *hembu-* and the suffix *-a* appear in every single paradigm cell and so have natural distributions. The only characteristic of the forms in 3, 1PL and 2SG that distinguishes those cells from others is the absence of an (overt) person agreement suffix like the *-n-* or *-w-* which appear in the other forms. Therefore, the formal identity of the shaded cells in the paradigm of *hembu* may not need to be really “explained” in any way. Specific reference to the cells 3, 1PL and 2SG is not needed to describe the inflectional paradigm in Table 15.

That said, it is hardly controversial to point out that language users are able to assign specific meanings to word forms by virtue of those absences referred to as 'zero morphs'. The knowledge of systematic oppositions within a paradigm often allows language users to interpret absences much like they interpret overt formatives. It is, therefore, a matter for empirical discovery whether or not zero morphs are elements comparable to overt formatives, different elements, or are not elements at all. Given the deeply-rooted uncertainties surrounding zero in morphology (both regarding its status and its distribution) I will, in subsequent research remain agnostic about its nature and I will refer only to overt formatives in the remaining of this thesis.

2.4 Maximal domain of morphemes

One of the questions that remain open regarding morphomic structures (i.e. specifically regarding the perception of a pattern of formal identity as grammatically significant by the language user) is whether morphosyntactic or paradigmatic structure imposes any limit to them. Some linguists (Coats 1973, Jensen 1990, Pertsova 2007:35) have argued that any syncretism which cannot be described by underspecification constitutes a case of accidental homophony. Others allow for systematic structure to exist in the absence of shared features but argue that “there must be some paradigmatic connection” (Blevins 2016:108). Yet others (e.g. Round 2013) believe that morphomic connexions are possible even between paradigmatically unrelated elements such as a verbal affix with meaning X and a nominal affix with meaning Y.

This question (i.e. that of which domain, if any, should be regarded as the broadest within which systematic morphological structure is possible) is related to the acquisition of these structures by the language user. The difficulty to learn or perceive a given formal identity as systematic is likely to increase if independently justified morphological domains are straddled or if syntagmatic differences exist. That is, noting a similarity in morphological behaviour is likely to be harder between a verb and a noun than between two nouns of different inflection classes. Similarly, generalizations across nouns of different classes are probably more difficult than generalizations within a single lexeme's paradigm. Even within a single lexeme's paradigm it is likely that noticing morphological affinities will be easier within narrower domains (e.g. [singular] or [present]) than across those domains.

One of the reasons why the morpheme is such a controversial object of study is that a certain level of contradiction is present in its very definition, a point which is usually not addressed at all. It is quite remarkable that, for us to accept some case as a genuine instance of a morpheme, we require that a given formal identity be at once so-to-say I) 'chaotic' and II) 'systematic'. We are, therefore, demanding two things which are almost antagonistic:

- I) morphosyntactic unnaturalness.
- II) evidence for systematicity.

According to the first criterion, the more different the function or meaning of the different uses of a form, the more morphomic it should be considered. A form which appeared in the 1SG form of the verb and on the 3PL possessor form of the noun would be considered very morphomic indeed according to the first criterion.

According to the second criterion, the more systematic a formal identity is, the more we should regard it as a grammatical unit of some kind. The problem, of course, is that the main source of evidence for systematicity is, in fact, the restriction of a form to some coherent morphosyntactic environment. According to this, the identical marking of 1SG subject agreement on the verb and 3PL possessor on the noun could well be an accident.

Thus, when faced in English with the /s/ forms in *speak-s* vs *ask-s*, we do not usually require further proof of the systematicity and synchronic relevance of the affixal identity. This is, precisely, because formal similarity is accompanied by functional similarity, since both are 3SG forms. However, if someone were to argue that the presence of /s/ in both *speak-s* and *cat-s* is meaningful, we would normally demand some strong additional evidence. Precisely because the two /s/ do not appear to share any functional or semantic component, we would regard these identities as accidental or irrelevant by default and we would in principle, refrain from any unified morphological analysis.

This way of understanding grammar is not a theoretical whimsicality of linguists, of course. On the contrary, I believe it is completely justified. Language users, when making sense of their linguistic input, also use these cues when deciding whether two occurrences of the same form are instantiations of the same element or not. It is a plausible hypothesis that the amount of evidence required to “convince” language users that a formal identity is grammatically relevant is a function of the perceived “distance” between the various uses of the form. A sufficient morphosyntactic distance can probably override even quite robust evidence of formal identity. There is, for example, every reason to believe that the formal identity of the genitive and future markers in Basque which was presented in Section 2.1.2 is grammatically inert synchronically despite their substantial formal overlap.

Naïve speakers of Basque are surprised when this formal identity is revealed to them, which suggests that they are not consciously aware of this quirk of the language. In addition, the distribution of phonologically-conditioned allomorphy *-ko/-go* is no longer identical in its two uses. Thus, for example, after stems ending in /l/, many speakers use *-go* for the genitive (e.g. *Madril-go* 'of Madrid') but *-ko* for the future (*hil-ko* 'will kill'). The different paths taken by these formatives suggest that their formal identity is not cognitively real synchronically.

The fact that speakers of Basque apparently refuse to grant any synchronic import to future/genitive syncretisms does not mean that similar cases cannot be analyzed differently in other languages. Round (2016), for example, proposes that various morphological operations in Kayardild, which can apply to both verbs and nouns with seemingly unrelated meanings, indeed have to be granted synchronic grammatical status in the language. In Kayardild, unlike in Basque, verb-noun affixal identities are recurrent, not limited to an isolated case, which may increase the likelihood of them being attributed synchronic import.

Different word classes tend to inflect for different features. This is likely to make it more difficult for speakers to make generalizations over inflectional patterns in different classes. This is not always the case, however. The phenomenon known as transcategorial polyfunctionality (Stump 2014, Stump 2015:229) unmistakably demands that speakers be able to make unified analyses of nominal and verbal suffixes sometimes. Languages like Tundra Nenets (Ackerman & Bonami 2017), for example, have sets of suffixes indexing person-number combinations in different word classes. The possessor in nouns, the subject in verbs and the object in prepositions are marked with the exact same markers regardless of the word class they attach to. Postulating different homophonous affixes (e.g. a *-da₁* in nouns vs a *-da₂* in verbs, a *-maq₁* in nouns vs a *-maq₂* in verbs etc.) would miss a robust generalization that holds for dozens of other suffixes as well as the common semantic value of the different uses, since both *-da* 'mean' 3SG and both *-maq* 1PL.

Morphological objects can, therefore, definitely straddle the border between different grammatical categories in some cases. Can morphomic elements do so too? That is, can affixes with unnatural morphosyntactic distributions span more than one word class? The fact that different word classes usually inflect for different values, as I mentioned before, is a big

hindrance. However, when different word classes do happen to inflect for the same values, morphomic paradigmatic patterns can indeed be shared by different classes. Consider the following person-number syncretism:

	<i>xo:t</i> 'house' (noun)			<i>we:r</i> 'make' (verb)		
	SG	DU	PL	SG	DU	PL
1	<i>xo:te:m</i>	<i>xo:te:mən</i>	<i>xo:te:w</i>	<i>we:rle:m</i>	<i>we:rle:mən</i>	<i>we:rle:w</i>
2	<i>xo:te:n</i>	<i>xo:tlən</i>	<i>xo:tlən</i>	<i>we:rle:n</i>	<i>we:rlələn</i>	<i>we:rlələn</i>
3	<i>xo:tl</i>	<i>xo:tlən</i>	<i>xo:te:l</i>	<i>we:rləlli</i>	<i>we:rlələn</i>	<i>we:rle:l</i>

Table 16: Khanty (Uralic) possessor and subject inflection (Nikolaeva 1999)

The same unnatural syncretism pattern (2/3DU+2PL) is found in the noun (for possessor inflection) and in the verb (for subject inflection). The same syncretism, in addition, is repeated in other possessee and object numbers (Table 16 shows singular object/possessee), which suggests we are dealing with a systematic pattern here.

Zooming in more, e.g. within a single word class, it is my contention that it becomes gradually easier for language users to spot identical recurrent partials and to integrate formal identities into their grammatical understanding of the language. For example, between different lexemes, formal identity is usually not unexpected and even could be said to be the 'default'. Consider the following two Russian declensions:

	<i>rabota</i> 'work'		<i>mesto</i> 'place'	
	SG	PL	SG	PL
NOM	<i>rabota</i>	<i>raboty</i>	<i>mesto</i>	<i>mesta</i>
ACC	<i>rabotu</i>	<i>raboty</i>	<i>mesto</i>	<i>mesta</i>
GEN	<i>raboty</i>	<i>rabot</i>	<i>mesta</i>	<i>mest</i>
DAT	<i>rabote</i>	<i>rabotam</i>	<i>mestu</i>	<i>mestam</i>
INST	<i>rabotoj</i>	<i>rabotami</i>	<i>mestom</i>	<i>mestami</i>
LOC	<i>rabote</i>	<i>rabotax</i>	<i>meste</i>	<i>mestax</i>

Table 17: Declension of three Russian nouns

It is probably not accidental, therefore, that the oblique plural forms of the different inflection classes of Russian share the same form. This, in fact, was an analogical levelling implemented by language users (cf. Slovene DAT.PL *-am* vs *-om*), so positing homophonous affixes (e.g. in the dative plural: *-am₁*, *-am₂* etc.) would seem to be a misrepresentation.

When considering other formatives, however, the situation becomes different. The suffix *-u* can mark the accusative singular (in *rabota*), and the dative singular (in *mesto*). Should we recognize independent homophonous suffixes *-u₁* *-u₂* because *-u* has different values in different inflection classes? Or should we understand *-u* as Kayardild's (Round 2016) inter-class inflectional formatives (i.e. as a single operation which can map onto different values in different classes)?

The evidence in favour of the second analysis is, intuitively, not too robust (much less so than that for future=genitive in Basque). The pattern is limited to *-u*, which, being one of only five (or six) vowels to choose from in Russian could well be used more than once in declension merely by chance. Much as it happened in Basque with the genitive/future syncretism, therefore, this formal identity may well be inert in synchronic grammar.

In other cases, however, it seems more clear that affixes in different lexemes are somehow “the same thing” despite not having the same morphosyntactic distribution:

	Class I		Class IV		Class V	
	SG	PL	SG	PL	SG	PL
NOM	-∅	-∅	-∅	-ni	-∅	-∅
GEN	-∅	-ni	-∅	-ni	-∅	-∅
LOC	-∅	-ni	-∅	-ni	-∅	-∅

Table 18: Some nominal inflection classes in Nuer (Nilotic)
(Baerman 2012: 470, from Frank 1999)

Consider, for instance, the formative *-ni* across different nominal inflection classes in Nuer. Its distribution often differs from one class to the other and cannot be defined successfully in morphosyntactic terms. One could, as in Russian, posit homophonous suffixes with different distributions. However, the sheer ubiquity of the formative (it appears, with one distribution

or another across more than 20 different classes), as well as the fact that it always appears in the plural, and preferably in the oblique plural, intuitively suggest that positing a few homophonous *-ni* cannot be the right approach. The alternative is, inevitably, that we have just a single formative with a complex morphosyntactic distribution.

As was the case with morphomic identities across grammatical categories (e.g. in Basque or Kayardild), the same unnatural morphological affinity can actually be repeated with several exponents, thus minimizing the likelihood of formal identity being accidental. Consider the following inflection classes:

	'tap head'	'return'	'do'	'pass'	'love'	'hear'	'fall'	'give birth'	'plant'
AOR	tuà	goá	tié-ní	cié	bua	gbà-dì	bà	ma	bù-lì
IPFV	tua-ní	goâ	tie	cié-dí	buà	gbà	baà-lí	ma	bu
PFV	tuà	goá-ní	tié-nì	ciê	bua-dì	gba-dì	bà	ma-lì	bù-lí

Table 19: Some inflection classes in Gourmanchéma (Atlantic-Congo)
(Baerman et al. 2017 after Naba 1994 & Ouoba 1982)

Consider the suffix *-ni*. It appears in a seemingly unmotivated set of contexts in different inflection classes. However, its distribution is matched exactly by that of other formatives like *-di* and *-li*. This provides a strong motivation for language users to actively employ these predictive relations and to internalize them, thus optimizing the resolution of the so-called Paradigm Cell-filling Problem (Ackerman et al. 2009). A speaker of Gourmanchéma coming across the AOR form *tié-ní*, for example, will be able to predict its corresponding IPFV and PFV forms if they have internalized the pattern described here. If they have not, the forms of IPFV or PFV could well be any of *tie*, *tie-ni*, *tie-di* or *tie-li*. In fact, with this system, any affixed form licenses reliable inferences about other cells. An affixed IPFV form, for example, immediately entails unsuffixed AOR and PFV forms. Conversely, an affixed PFV also entails an unsuffixed IPFV.

However interesting morphological affinities across classes may be, the domain within which morphological identities tend to be explored in morphomic literature tends to be smaller. Many researchers, in fact, voice objections to treating morphological affinities beyond and

within the paradigm (or beyond and within a subparadigm) in the same way. Blevins, for example, argues that:

Pairs of elements with no discernible connection, such as the agentive and comparative -er markers in English, are (...) not morphemes. A morphomic pattern can, in principle, involve words, parts of words, or even sequences of words. But there must be some paradigmatic connection between these elements. (Blevins 2016:108)

According to this reasoning, morphological affinities between different word classes, like the ones of Kayardild (Round 2016), or between different inflection classes (e.g. in Gourmanchéma in Table 19) cannot ever be morphomic. Pertsova goes even further in the restriction of the window of opportunity for morphemes when she argues that:

it is plausible that in trying to solve the mapping problem, the learner chunks up the semantic space into smaller subspaces or subparadigms and operates within these smaller spaces first (so that accidental homophony between formatives in different subparadigms may not be so starkly dispreferred). (Pertsova 2011:254)

Similarly, when enunciating his Syncretism Principle, Müller (2005:236) also argued that the null hypothesis for linguists and language learners must be that identity of form implies identity of function, but just within independently justified morphological domains. The impulse to pursue unified analyses of only those formal identities that “share” something apart from just form certainly 'makes sense'. As argued by Bermúdez-Otero & Luís (2016:336) “the ease or difficulty with which a category is discovered may largely depend on the logical relationship between the features that go into the category’s definition”.

The concerns of these various linguists are, therefore, very much justified in my opinion. It is, for example, intuitively plausible that if the Basque formal identity, instead of genitive and future, had involved closer functions, it may have been able to remain a synchronically productive part of the grammar. Regardless how well founded these concerns are, however, in the absence of some concrete ways of implementing these in a sensible way, there is a danger that one simply disregards particular morphological identities on aprioristic arbitrary

reasons or just because they 'do not fit well' into one's own theoretical framework. One could, for example, restrict what counts as an independently justified morphological domain in a way that the possibility of morphomic exponents is rendered impossible altogether. If, for example, the present tense subparadigm, or the singular subparadigm can constitute autonomous domains in this respect, then any formative that occurs inside and outside the domain will simply be analyzed as two homophonous formatives rather than one. In this way, even the most incontrovertible morpheme would be simply 'converted' into two or more morphemes (remember the discussion on Chaha in Table 12). This is clearly not the right approach to investigate morphomicity.

A possible sensible criterion could be the one advocated for by Blevins (2016). There is, in my opinion, a big difference between those formatives whose morphomicity only becomes apparent when equating elements from different paradigms (e.g. Basque, Kayardild, Nuer, Gourmanchéma) and those whose morphosyntactic unnaturalness is already evident within a single lexeme's paradigm and is simply replicated in others:

Gender	Conjugation 1		Conjugation 2	
	SG	PL	SG	PL
I Male	j-	s-	b-	t-
II Female, animate	g-	s-	n-	t-
III Miscellaneous	g-	j-	n-	b-
IV Mass nouns	j-	j-	b-	b-
V Banana, sago tree	j-	g-	b-	n-
VI Arrows, coconuts	g-	g-	n-	n-

Table 20: Conjugations in Burmeso (Isolate, New Guinea)
(Corbett 2009:9, from Donohue 2001:100-102)

As the above paradigm shows, within any conjugation 1 verb's paradigm, a form like *j-* or *g-* can appear, depending on the noun that triggers the agreement, in the singular, in the plural, in both numbers and in none of them. Thus, the contexts where these forms appear within a single paradigm, do not constitute a natural class. The fact that the same identical pattern is found in other lexemes' paradigms, both with the same exponents (in other conjugation 1 verbs) and with others (in conjugation 2 verbs) is just a bonus and not the factor upon which

the purported unnaturalness hinges. To be on the safe side, therefore, I will focus here on these morphemes, i.e. on those which are identifiable within a single paradigm.

This is not meant to imply that morphological relationships beyond the paradigm are irrelevant. It is hardly a far-fetched suggestion, for example, that the systematicity of Gourmanchéma verb class structure may enhance the learnability of the system as a whole. Its nine inflection classes can be arranged into just three classes based on the suffix used: -ni -di or -li and into another three classes based on the paradigmatic distribution of the affix:

	-ni	-di	-li
Suffixed IPFV	'tap head'	'pass'	'fall'
Suffixed PFV	'return'	'love'	'give birth'
Suff. AOR/PFV	'do'	'hear'	'plant'

Table 21: Orthogonality of Gourmanchéma inflectional classes' traits

If the achieved economy (abstracting six categories instead of nine) is sufficient, then these abstractions are maybe worth doing by language users of the language. If that is the case, ni-containing verbs would constitute a 'class of classes' and would be synchronically “the same” at some grammatical level, which is what is usually asked of morphemes.

It is my contention that, if the evidence offered to the language user is sufficiently compelling, grammatical categories can indeed be posited that transcend the borders of inflection classes or word classes. Phrased in other terms, if the optimal strategy for the acquisition of a pattern involves the *ad hoc* creation of a morphomic category beyond the paradigm, this will probably be done. It is, however, extremely difficult for the linguist to assess when this is the case and when morphological affinities are ignored instead.

Because looking into speakers' brains is not an option, an alternative strategy has to be sought to try and discard most instances of “spurious morphemes” like the Basque one. Morphological affinities beyond the paradigm are necessarily weaker than those within a single paradigm. The amount of formal evidence required to “convince” a language user that genitive and future are marked by the same formative must be, therefore, larger than that

required to convince them of some intraparadigmatic affinity. I agree with Joseph (1997:11) when he mentions that “[m]ost generalizations (...) should be recognized as being truly local in nature, that is, they have a restricted scope, and where linguists’ generalizations go astray is in not being sufficiently localized”. Because I cannot investigate here in detail the cognitive status of morphomic patterns, I will strategically restrict my attention to those morphomic patterns which are apparent within the inflectional paradigm of a single lexeme.

2.5 Isolatability and the concept of the ‘stem space’

Although this is all that is usually mentioned, defining a morphome simply as an unnatural set of cells or morphosyntactic values which are systematically syncretic is not enough. Consider, for example, the 1SG.PAST and the 3PL.PAST in German:

	machen 'do'				singen 'sing'			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	mache	machen	machte	machten	singe	singen	sang	sangen
2	machst	macht	machtest	machtet	singst	singt	sangst	sangt
3	macht	machen	machte	machten	singt	singen	sang	sangen

Table 22: Present and past tense inflection of two German verbs

Those two paradigm cells constitute an unnatural class and also behave in the same way morphologically, since the use of the affix *-te* in one of the cells implies its use in the other and the use of some vowel apophony in one cell also implies the same form in the other.

Intuition tells us, however, that we are clearly 'cheating' by analyzing the exponence of the 1SG.PAST and the 3PL.PAST separately from that of the other past tense cells. The other past cells, after all, also share the same quirks across every single lexical item so that there is no reason (i.e. no form in any lexeme) for isolating the cells 1SG.PAST and 3PL.PAST from their neighbouring cells.

In cases like German above, it is intuitively clear that the correct unit of analysis from the morphological perspective is the whole of the past tense cells. However, this is not always so clear. Consider the patterns of formal contrasts in Skou (Sko, New Guinea):

	2PL	1SG	3SG.NF	1PL	3PL	3SG.F	2SG
eat	ang	kang	kang	nang	tang	kang	mang
walk	ha	ha	ka	na	ta	wa	ma
narrate	láng	láng	láng	táng	jáng	wáng	páng
hit	ká	ká	ká	ká	já	wá	bá
utter	lá	lá	lá	lá	lá	wá	pá
get.F	wé	wé	wé	wé	wé	wé	pé
plant	wá	wá	wá	wá	wá	wá	wá

Table 23: Exponence patterns of various Skou verbs (Donohue 2004:219)

In Skou, there are several progressively larger unnatural sets of paradigm cells that are syncretic. The conflated cells highlighted for the verbs 'walk', 'narrate', 'hit', 'utter' and 'get.F' are all sets of cells which, sometimes, do share form to the exclusion of neighbouring paradigm cells. This is a criterion that I will use throughout this dissertation. With this requirement in mind we would say that there are several morphemes in Skou which constitute, like Russian dolls, progressively smaller subsets of each other.

It has to be noted that this criterion is, to some extent, an executive decision. There are plausible alternative ways to think about or operationalize these patterns. Because of the implicative subset relations that unite these various sets of cells, an alternative would be to understand this morphological affinity as a single but 'blurry' morphological unit as it were:

	SG	PL
1		
2		
3.F		
3.NF		

Table 24: Probability of syncretism with the 2PL in Skou verbs (darker = more probable)

The decision to deny morphomic status to any set of cells that never shares a form to the exclusion of neighbouring paradigm cells is not as trivial as the German example in Table 22 suggests. In some cases, cross-paradigmatic evidence can indeed pick out a set of cells (e.g. because they, and only they, always share some form across every single lexical item) without surfacing as a formally identifiable unit in any one lexeme's paradigm. This is the case, for example, of the infinitive and the 2PL imperative in Spanish:

	1PL future	infinitive	2PL imperative	2SG imperative	1PL present Ind
'be'	se-remos	se-r	se-d	se	somos
'go'	i-remos	i-r	i-d	ve	vamos
'have'	tend-remos	tene-r	tene-d	ten	tene-mos
'read'	lee-remos	lee-r	lee-d	lee	lee-mos
'sit'	senta-remos	senta-r	senta-d	sienta	senta-mos

Table 25: Some selected paradigm cells of 5 Spanish verbs

There is no formal element whatsoever in any lexeme that appears in the infinitive and the 2PL imperative cells of the paradigm to the exclusion of all other cells. One can see in Table 25 that in 'go' the stem in the infinitive/2PL.imperative is also used in the future. In 'have', in contrast, it is the 1PL (and 2PL) present indicative that use the same stem as the infinitive and the 2PL imperative. In no lexical item, therefore, does a stem alternant or a formative appear in the paradigm confined to the infinitive and the 2PL imperative.

The generalization that these cells (and only these two cells) behave always in the same way regarding stem alternation is also inescapable, however. This is the reason why they are regarded as forming a so-called 'stem-space' in Spanish (see Boyé and Cabredo-Hofherr 2006). Even if stem-spaces like this one are obviously closely-related to the notion of the morphome and very interesting objects of morphological analysis, they will, for definitional reasons, be excluded from any further consideration here. Because I will be requiring, for morphomhood status, that the set of cells constituting an alleged morphome be formally identifiable within a single lexeme's paradigm, sets of cells like standard Spanish's infinitive + 2PL imperative will not be further examined throughout this thesis.

A clarification note seems appropriate in relation to this, however. The distinction between morphemes and stem spaces is one that could well turn out to be superfluous if the two elements share every single empirical property apart from their definitional one(s). There is, for example, some evidence that, in the same way as morphemes, stem spaces can also constitute cognitively real grammatical entities for language users. This is illustrated, for example, by a very common morphological change in substandard Spanish. For many speakers, the etymological form of the 2PL imperative tends to be replaced by the form of the infinitive (i.e. *sed>ser*, *id>ir*, *tened>tener* etc.). As a result of the change, therefore, the two cells (and only those two cells) become syncretic at the level of the whole word and thus come to form a formally isolatable morphological unit in these speakers' grammar.

It is, therefore, safe to say that, in the domain of stems, there is at most a very thin line between unnatural stem spaces and morphemes. Despite cases like the one just presented in Table 25, the criteria used for stem-space identification and for morpheme identification often converge in practice on the same sets of cells. For example, Boyé and Cabredo-Hofherr's (2006) identification of stem spaces in Spanish yields, among others, the units '1SG Present Indicative and Present Subjunctive', and 'Preterit, Imperfective Subjunctive I and II, and Future Subjunctive'. These are the sets of cells known as the L-morpheme and PYTA respectively in morphomic literature. Be that as it may, in order to both narrow down the object of study of the present dissertation and to avoid a break with established terminology, the two concepts will be kept separate here. Consequently, the requirement will be enforced throughout this dissertation that a morpheme be identifiable within a single paradigm by some form exclusive to it.

2.6 Morphemes and cross-linguistic recurrence

One of the few points where linguists of quite different convictions (e.g. Maiden 2016, Koontz-Garboden 2016) seem to have agreed so far is the claim (or theoretical stance) that morphemes must be typologically unique. That is, for a paradigmatic structure to be truly morphomic, it should not be found to occur in two unrelated languages. The reasoning behind this is that, if something had emerged more than once independently it might be proof of

some extramorphological *raison d'être* or rationale for its synchronic existence, even if we had no idea what this might actually be.

While one may sympathize to some extent with that general line of thought, there are some fundamental problems with it. The first is related to circularity. Thus, we cannot *claim* to have found out that morphemes are typologically unique if we *require* them to be so. That is, we have to be very clear as to whether something is part of the definition of some phenomenon or an empirical finding predicated of it. If we make our definition of morphemehood (or our diagnostics thereof) dependent on typological uniqueness, this precludes any possible further empirical discoveries in this respect. This is particularly undesirable in this case because language users have no access to the cross-linguistic recurrence of a pattern (nor to its historical origin). Because of this, speakers cannot be expected to draw any distinctions along these lines.

Another big problem comes when assessing typological uniqueness. At a sufficient level of granularity, probably every single grammatical category (e.g. the Russian accusative, the English past, the Spanish passive etc.) is unique. Thus, if we require identity with respect to every detail and variable, all morphemes will, indeed, be typologically unique. However, the same as we do with other grammatical entities (consider the long-winded debate on comparative concepts and descriptive categories), this should not be the end of the typological enterprise. We must be allowed to look at specific variables at a time to find that morpheme A and morpheme B are, for example, the same in one particular respect and different in another. This is, essentially, the backbone of Multivariate (Bickel 2010) and Canonical Typology (Corbett 2005).

The typological uniqueness of morphemes has usually been predicated of their paradigmatic distributions as a whole. Maiden (2018b:167), for example, defines the N-morpheme as an alternation such that “the forms of the first-, second-, and third-person singular and of the third-person plural in the present indicative, present subjunctive, and imperative share formal characteristics not found elsewhere in the paradigm”.¹⁰ He insists on the typological

¹⁰ Under close scrutiny it becomes apparent that, in fact, his assessment of whether two morphemes are “the same” or not is not driven so much by synchronic paradigmatic distributional concerns as by etymological (i.e.

uniqueness of such a paradigmatic structure and he makes it clear elsewhere (2018:22) that a morphological opposition of SG+3PL vs 1/2PL is a different pattern, and possibly not even morphomic, he argues, given that it is found in unrelated languages.

At a sufficient level of abstraction, however, the N-morphome is, indeed, made up of SG+3PL cells. The number of tenses that a morphome spans (three in this case), and whether or to what extent a morphome is confined to particular inflectional subdomains (e.g. the present tense in the case of the N-morphome) are obviously relevant but logically independent variables of cross-linguistic variation. An important general finding that has emerged from the present research program and database is that morphomes, like any other grammatical structure or phenomenon in language, are liable to be compared within and across languages and classified as for their relative degree of similarity or dissimilarity.

Because of the aforementioned ontological and diagnostic problems, restricting the attention of the present research to typologically unique patterns would be both arbitrary and pernicious to further empirical discovery. Language users do not have access to the grammatical systems of the world's languages and I therefore see no grounded reason to attribute any special status to those patterns that are only attested once as opposed to those which are attested more than one time. This is likely to be determined merely by the size of our sample of languages or by the current state of language documentation rather than by any inherent property of the patterns themselves. The same can be said of linguistic history. Because cross-linguistic recurrence and diachronic origin are inadequate diagnoses for morphomhood we will simply have to find our definitional criteria elsewhere.

genealogical descent) considerations. This is evidenced by his approach to labelling. Thus, stems appearing in SG+3PL present indicative and in 2SG imperative (but crucially not in the present subjunctive) are taken to be instantiations of the N-pattern (Maiden 2018b:195). The same can be said of alternants involving SG+3PL present indicative, 2SG imperative and ALL subjunctive (Maiden 2018b:194). Even patterns involving 2SG+3SG+3PL present indicative are said to be also instantiations of the N-morphome (Maiden 2018b:227).

It is clear, thus, that an N-morphome (root) is recognized as such when its form is regularly descended from a Latin rhizotonic one, independently, to some extent, of whether it has preserved its original paradigmatic configuration. It cannot surprise us, therefore, that Maiden regards the N-morphome as a typologically unique trait of the Romance family.

Maiden's (and acolytes') approach the morphome constitutes a philological study of the morphological and paradigmatic configurations and reconfigurations of inherited stem allomorphies. This approach is, of course, perfectly valid and highly illuminating. It is, however, an endeavour different altogether from a broader typological one like the present dissertation. In typology, comparisons and assessments of 'sameness' and differences cannot and should not be done from an etymological perspective.

2.7 Morphemes and locus of exponence.

It is fair to say that most of the research around morphemes has focused, to date, primarily on stems rather than on affixal formatives. This may be so because, for many morphological models and linguists, the stem is a locus for lexical and not for grammatical meaning:

Stems do not serve as realizations of properties, though the property set of a word form may determine which stem is selected as the base for inflection. (Spencer 2016:226)

Consequently, it is, for many, not unexpected to find that a particular stem alternant does not have a morphosyntactically coherent distribution (i.e. that it does not “mean” anything grammatically). By contrast, in grammatical formatives this eventuality is unexpected and undesirable from the formalist constructivist perspective. Because of this, all sorts of analyses and formal mechanisms are proposed in these cases to conjure up a coherent morphosyntactic function in suffixes and to transfer it away from stems:

*In German, for example, some verbs show characteristic ABLAUT or UMLAUT patterns, where person and **tense-indicating formatives trigger different vocalisms**. From *tragen* ‘carry’, we get first person singular present *trage*, second person singular present *trägst*, and first person singular past *trug*, each with different stem vowels. (Bickel & Nichols 2007:186, emphasis mine)*

From an atheoretical point of view, however, there is no reason to assume, a priori, that grammatical meaning must be realized exclusively by means of segmentable inflectional formatives. In the particular case advanced by Bickel & Nichols (2007), for example, it seems more sensible to say that the locus for the present/past distinction is to be found, at least partially, in the difference in stem vocalism rather than in affixal material exclusively:

	PRESENT		PAST	
	SG	PL	SG	PL
1	trag-(e)	trag-en	trug	trug-en
2	träg-st	trag-t	trug-st	trug-t
3	träg-t	trag-en	trug	trug-en

Table 26: German verb *tragen* 'carry'

Given that most of the suffixes in Table 26 are (or appear to be) tense-neutral (e.g. *trag-t* vs *trug-t*), saying that the stem alternation pattern is triggered by the suffixes (Bickel & Nichols 2007: 186) does not seem to follow easily from the empirical data.

There is cross-linguistic evidence that stem alternations can sometimes serve as the sole exponent of morphosyntactic distinctions. The verb 'give' in Iha, for example, changes its stem according to person and number of the recipient:

	SG	PL
1EXCL	qpe	qpe
1INCL	-	qpi
2	kewé	kiwi
3	kow	kow

Table 27: Verb 'give' in Iha (West Bomberai, New Guinea)
(Donohue 2015: 413)

It is also not difficult to come up with cases of clearly segmentable affixes failing to encode morphosemantic distinctions. Consider, for example, the distribution of -ni, -di and -li suffixes in Gourmanchéma as was presented in Table 19.

These examples suggest that, unless it is programmatically incorporated as part of their definition, the distinction between stems and affixes is unrelated to the presence or absence of grammatical meaning. Throughout this thesis, therefore, stem or affixal status will not influence the assessment of a pattern's morphomicity.

2.8 Segmentability

A property of prototypical formatives (and also of whole words) is that they are units which are easily segmentable from surrounding elements. That is, in more technical terms, they are islands of invariance surrounded by peaks of unpredictability. A property of all the Spanish 1PL verb forms (e.g. *somos, fuimos, damos, amaremos* etc.) is that their shared form is easily identifiable and segmentable by linguists. It is clearly *-mos* and not *-os* or *-amos* that the 1PL forms all have in common. This formative, in addition, cannot be said to express any other thing rather than 1PL since it appears always in that morphosyntactic context and never in other contexts. Its properties are, thus, not very different from other grammatical units (e.g. a preposition like 'under' or an article like 'the') which have abstract meaning. As argued by Pertsova (2008:15), it is not clear that anything would prevent a child from “using general learning strategies for segmentation and association of forms with meanings to posit morphemic lexical entries” in cases like *-mos*.

Deviations from this unproblematic case are not difficult to find, however. Problems with segmentability and with mutually incompatible segmentations are well known (e.g. Bank & Trommer 2012, Blevins 2016:26-28). Sometimes, the elements which can be identified on transitional grounds alone are relatively clear, as in Wardaman:

	SG	DU	PL
1EX	nga-	yi-rr	
1INC		nga-yi-	nga-rr
2	yi-	nu-	
3	∅-	wu-rr-	

Table 28a: Wardaman (Yangmanic, Australia) intransitive indicative prefixes (Merlan 1994:125)

Despite this apparent segmentability, the morphosyntactic distribution of some of the formatives (e.g. *yi-* or *rr-*) in Table 28a is problematic, which by itself, according to some analyses (see e.g. the approach to segmentation in Trommer & Bank 2017), should cast doubt on the segmentation that yielded those elements in the first place. The advantage for the

language user of a decompositional analysis of these forms (i.e. *yi-rr-*) over the alternative analysis involving undecomposed elements (i.e. *yirr-*) is indeed not clear.

Alternative and mutually incompatible possibilities for segmentation are not infrequent and many discussions have focused on addressing problematic instances. One such case concerns the right segmentation of the velar augment characteristic of the L-morpheme. According to the traditional analysis, forms like Spanish *vengo* or *tengo* are decomposable into the stems *veng-* and *teng-* and the 1SG suffix *-o*. O'Neill's (2015) segmentation, however, identifies *ven-* and *ten-* as the stems and *-go* as the 1SG suffix. In so doing, he is basically relocating the allomorphy from the stem (e.g. *ten-/teng-*) to the suffix (*-o/-go*). The decision to segment in one place or the other (or in both) is arbitrary to a large extent and irrelevant for the present discussion. In either case we are left with a morphological element with an L-shaped paradigmatic distribution which we need to account for.

Despite the irrelevance (for morphomicity) of segmentation in many cases formatives can sometimes be dependent on (debatable) segmentations to exist. Those arising from very unorthodox segmentations are more exposed to being only by-products of a theoretical analysis rather than a grammatical unit in the language. In a similar vein, a given pattern of formal identity will be easier to perceive and learn by language users if it affects elements that are combinatorially treated consistently as whole objects, like Spanish *-mos*, rather than if a formal identity involves forms with an uncertain or a variable combinatorial status:

	Agent non-past		Subject		Subject past	
	SG	PL	SG	PL	SG	PL
1	ʔan-	mu-k-	ʔa-n-	muk-	ʔan-	muk-
2	ka-	ka-	ka-	ka-	ka-	ka-
3	mu-	mu-	ʔa-	ʔa-	∅-	∅-

Table 28b: Agreement prefixes in Xincan (Xincan, Guatemala) (Sachse 2010:233)

In some agreement contexts in Xincan, the third person shares some element of form (/mu/ or /ʔa/) with another paradigm cell. The resulting patterns of affixal identity (i.e. 3+1PL and 3+1SG), however, only ever get instantiated by one form and are dependent on

segmentations (i.e. *mu-k-* and *ʔa-n-* respectively) that do not appear supported by forms in other paradigms. This may therefore not really represent a significant fact about Xincan morphology but might constitute simply a case of accidental partial homophony. Note that if we allowed similar *ad-hoc* segmentations elsewhere one could spot unnatural patterns of morphological identity practically everywhere:

	German 'need'		Spanish 'need'	
	SG	PL	SG	PL
1	brauche	brauchen	necesito	necesitamo-s
2	brauchs-t	brauch-t	necesita-s	necesitái-s
3	brauch-t	brauchen	necesita	necesitan

Table 28c: Two unorthodox segmentations in German and Spanish

Thus, in German, on purely combinatorial grounds, /t/ is a formative (all by itself) in the 3SG and the 2PL but not (or not so certainly) in the 2SG, where the suffix is usually taken to be /st/. Similarly, in Spanish, /s/ is a formative in the 2SG but probably just a fragment of a bigger formative in the case of the 1PL and the 2PL.

Even if, as argued by Blevins (2016), there is no reason to assume that different patterns, incompatible from a constructivist perspective, cannot be simultaneously relevant, the availability of alternative (and better?) analyses to the language user may undermine the status of elements emerging from controversial segmentations like those in Tables 28b and 28c. With this in mind, uncontroversial morphemes should be based upon forms which are easily discriminated, syntagmatically, from neighbouring phonological material. Thus, I will refrain throughout this dissertation from performing noncanonical segmentations like these, and will stick to the choices of the original grammatical descriptions.

2.9 Difficult cases

One of the things that discussion around the morpheme most urgently has to come to terms with is that the distinction between morphosyntactically motivated and unmotivated patterns is not the dichotomous choice that part of the literature seems to assume. Even within tabular inflectional paradigms, where it is usually easier to tell, things that look morphosyntactically unmotivated at first sight may not always be straightforwardly so and various degrees and sources of motivation are often possible.

2.9.1 The problem of the 1PL

As usually represented (i.e. in tabular form) paradigmatic structure seems to be a matter of well-behaved orthogonal features with mutually exclusive values. However this is sometimes just a convenient fiction. For example, in the domain of person, several 'he's (3SG) can indeed be equated with 'they' (3PL); however, several 'I's (1SG) are, if anything, a dissociative identity disorder. That specially 1PL and to a lesser extent 2PL are not straightforward plurals of 1SG and 2SG respectively is well known (e.g. Cysouw 2003, 2005). The 1PL in English, for example, can refer to various groups in which the speaker is always present (e.g. 1+3) but in which the addressee is usually present as well (e.g. 1+2, 1+2+3). What is more, if frequency of use is taken into account, most uses of the 1PL actually include, rather than exclude, the addressee. Despite this, syncretisms involving 1PL and 2SG, or 1PL and 2, are most usually treated as morphosyntactically unmotivated and morphomic without further discussion (e.g. Baerman & Brown 2013, Stump 2015: 128).

Apart from the above mentioned denotative affinity of 2SG and 1PL there are other reasons to doubt that this is the best example of a wholly unmotivated pattern. Although I have argued in Section 2.6 that this would not be considered a definitional factor here, cross-linguistic recurrence might still be revealing. The 1PL/2SG syncretism is relatively common in Papuan languages, for example. It is present, robustly, throughout the Tonda (Yam) and Gorokan (TNG) families, as well as in several individual languages such as Ekagi (TNG), Suki (TNG), and Yessan-Mayo (Sepik):

	Ngkolmpu (Yam) (Carrol 2016:306)		Benabena (Gorokan) (Young 1964: 59)			Suki (TNG) (Voorhoeve 1975)		Yessan-Mayo (Sepik) (Foreman 1973:27)		
	SG	PL	SG	DU	PL	SG	PL	SG	DU	PL
1	w-	n-	-be	-be	-ne	ne	e	an	nis	nim
2	n-	y-	-ne	-be	-be	e	de	ni	kep	kem
3	y-	y-	-be	-be	-be	u	i	ri/ti	rip	rim

Table 29: 2SG/1PL morphological affinities in PNG

As can be seen, the affinity of 2SG and 1PL can affect both agreement affixes (e.g. Ngkolmpu and Benabena) and pronouns (e.g. Suki and Yessan-Mayo) and is present in various genetically unrelated and geographically relatively distant languages.

A still more acute problem arises in those cases where the 1PL shares exponence with the second person as a whole. The motivation to mark 2 and 1PL in the same way seems relatively clear on semantic grounds. In the absence of clusivity, it is these person-number categories and these only that may refer to the addressee. Morphological patterns conflating 1PL and 2 are also not exceedingly difficult to find in unrelated languages:

	Darma <i>ra</i> 'come' (Willis 2007:350)		Mazatec 'lay down' (Jamieson 1988:106)		Aguaruna object agreem. (Overall 2017:243)	
	SG	PL	SG	PL	SG	PL
1	rayu	ransu	fañ-	tsjuñ-	-hu	-hama
2	ransu	ransu	tsjuñ-	tsjuñ-	-hama	-hama
3	rasu	rasu	fañ-	fañ-	-∅	-∅

Table 30: Some 1PL+2 morphological patterns

Morphological identity of 1PL and 2 is found, in the above examples, in whole-word forms (Darma, Sino-Tibetan), as well as in stems (Mazatec, Otomanguan) and affixes (Aguaruna, Chicham, Peru) separately. The shaded cells have a possible reference to the addressee in common. However, because in languages without clusivity the defining feature of the category 1PL is not inclusion of the addressee but of the speaker, this pattern (and the previous one of 2SG+1PL), even if not nearly as arbitrary as those involving comparable

person-number combinations (e.g. 2SG+3[PL]), cannot be described as a natural class in the traditional sense of the term. Thus, the shaded cells in the above paradigms, although they come close, are not reducible to the presence of the feature value 2. I will, consequently take these patterns as morphomic in this dissertation, although with a pinch of salt.

It has to be kept in mind, however, that not all languages categorize the plural person complex in the same way. Languages with clusivity code 1INC, 1EX and 2PL all in different ways. English and other languages without clusivity conflate 1INC and 1EXC, and distinguish those from 2PL. However, the mirror-image of English also exists. Some languages do not have 1 as their definitional criterion for the categorization of the plural complex. If the crucial aspect is not inclusion of the speaker but inclusion of the addressee, languages will code 1INC and 2PL in an identical way and distinguish these from 1EXC (see e.g. Sanuma [Yanomamic, Brazil] below). When some formative spans this addressee-centred plural complex and the 2SG, it may superficially appear that the form has an unmotivated distribution (see Ojibwe [Algonquian] below). However, as has been explained, in these cases there is indeed a necessary and sufficient condition (reference to 2) that accounts for the distribution, which will thus be morphosyntactically motivated here and not morphomic:

	Sanuma non-emphatic pronouns		Ojibwe intransitive prefixes	
	SG	PL	SG	PL
1EXC	sa	samakö	int-	
1INC	-	makö	-	kit-
2	wa		kit-	
3	∅		∅	

Table 31: Some 1INC=2PL paradigms (Cysouw 2003:154-155)

It has to be kept in mind that the use of the label 1INC, (as opposed to, for example, a label like 2INC, which would suggest that the category is somehow a second person which includes the speaker), is a mere convention. This originates probably from the fact that most languages where just one distinction is drawn categorize the complex as 'groups including the speaker' vs 'groups not including the speaker' and not, like Sanuma, as 'groups including the addressee' vs 'groups not including the addressee'. Objectively, however, we have no reason to favour

any of the two choices and therefore our theorizing has to remain neutral in this respect. Cases like Sanuma *makö* or Ojibwe *kit-* will thus not be regarded as any more morphomic than English 'we' or Ojibwe *int-*.

2.9.2 Syntactically licensed morphemes

Traditionally, the term morpheme has been applied exclusively to elements within the realm of morphology. I do not intend to depart from that tradition here. However, whatever we want to call the operations that target unnatural classes in other modules of grammar, we have to come to terms with the fact that these exist. Unlike usually assumed, unnatural classes in syntax, for example, can sometimes be the locus for particular operations or constructions. Take a look at the following sentences from Aguaruna (Overall 2007:443-444):

10a) atafu-**na** yu-a-tata-ha-i
 chicken-ACC eat-HIAF-FUT-1SG-DECL
 'I will eat chicken'

10b) atafu yu-a-tata-hi
 chicken eat-HIAF-FUT-1PL
 'we will eat chicken'

11a) nĩ ðima-ta
 3SG carry.PFV.IMP
 'you(sg) carry him!'

11b) kutji maa-ma-uhumi
 pig kill.HIAF-PAST-2PL
 'you (pl) killed a pig'

12a) tsabau-**na** yu-a-ti
 banana-ACC eat-HIAF-JUSS
 'let him eat a banana'

12b) kutji-**na** maa-aha-mi
 pig-ACC kill.HIAF-PL-RECPAST.3.DECL
 'they killed a pig'

As illustrated by the sentences above, nouns or noun phrases in the object position sometimes take the accusative marker *-na* and sometimes do not. This, however, is not due to any inherent property of the noun or the object itself but depends on the subject. This should, therefore, be described as a syntactic phenomenon. However, the set of subjects that trigger or do not trigger the accusative marking is not a class that would normally be considered natural. This rule seems to separate 1SG and third person subjects on the one

hand (which require the accusative *-na*), from 1PL and second person subjects on the other (which require an unmarked object noun phrase). Cases like these are probably infrequent but by no means unique. Another comparable case comes from Marsalese (discussed in Corbett 2016:82-83, from Cardinaletti & Giusti 2001):

13a) Vaju a pigghiu u pani.

go.1SG to fetch.1SG ART.SG.M bread

‘I go to fetch bread.’

13b) *Emu a pigghiamu u pani.

go.1PL to fetch.1PL ART.SG.M bread

‘We go to fetch bread.’

14a) Vai a pigghi u pani.

go.2SG to fetch.2SG ART.SG.M bread

‘You(sg) go to fetch bread.’

14b) *Iti a pigghiati u pani.

go.2PL to fetch.2PL ART.SG.M bread

‘You(pl) go to fetch bread.’

15a) Va a pigghia u pani.

go.3SG to fetch.3PL ART.SG.M bread

‘(S)he goes to fetch bread.’

15b) Vannu a pigghianu u pani.

go.3PL to fetch.3PL ART.SG.M bread

‘They go to fetch bread.’

As shown by the examples above, this particular syntactic construction is possible for some subjects (SG and 3PL) but not for others (1PL and 2PL). The set of subjects for which the construction is available constitutes an unnatural class.

As Aguaruna and Marsalese illustrate, the syntax can sometimes be sensitive to unnatural classes. These are fine syntactic *rara*, but what interest may they possibly hold for the study of structures which are exclusively morphological? Consider the following:

	Aguaruna object agreement (Overall 2017:243)		Marsalese 'go' present (Cardinaletti & Giusti 2001)	
	SG	PL	SG	PL
1	-hu	-hama	va-ju	emu
2	-hama	-hama	va-i	iti
3	-∅	-∅	va	va-nnu

Table 32: Two (syntactically licenced?) morphomic patterns

If morphemes are defined as elements of form which are independent of other modules of grammar (like e.g. syntax) the above morphological structures cannot possibly be considered morphemic. The previously discussed syntactic constructions in Aguaruna and Marsalese show that the syntax of those languages sometimes does care about (i.e. treats in a coherent way) classes like 1PL+2 or SG+3PL. The distribution of *-hama* and of the stem alternant *va-*, therefore, is not independent from syntax and cannot be said to be unmotivated in that sense. If we assume, like many theoretical models of grammar do, a layered structure whereby pragmatics precedes and motivates semantics, semantics precedes and motivates syntax, and syntax precedes and motivates morphology, these structures would be, indeed, externally motivated.

However, it seems that excluding these elements from the ranks of morphemes would do violence to the whole enterprise. On the one hand, this is not how we usually think syntax ought to work. If anything, in cases like Marsalese, we would like to explain the syntactic phenomenon as triggered somehow by the morphology, rather than the other way around. This is suggested by the fact that the same morphemic pattern (the N-morpheme) is found all over Romance and yet we seldom encounter cases like Marsalese. Thus, we tend to think of these cases more as counterexamples to the principle of morphology-free syntax than as cases of syntactically-motivated exponence. On a more utilitarian note, the amount of research that would be required to spot and discard these cases would be daunting. This means that, in practical terms, excluding those morphological structures that have some extramorphological correlate of this sort is impractical. The (probably few) cases where an unnatural morphological class is matched by an identical unnatural syntactic class will simply be accepted throughout this dissertation as *bona fide* morphemes, even if conceding the problematic nature of these cases.

2.9.3 Gender or morpheme? Other analytical uncertainties

I try throughout this dissertation to define morphemes in an empirically-oriented way, i.e. as something that can be identified in a language on purely distributional grounds and is independent from its subsequent theoretical or formal analysis. But every empirical definition of the morpheme (or any other phenomenon really) is necessarily intertwined with our definitions of other phenomena and, in general, the rules that we have agreed upon in our descriptions of language. The identification of some particular cases as morphomic, therefore, rests entirely on our correct identification of the relevant inflectional features and also on what we think other linguistic phenomena (e.g. gender) can be like. Consider the following agreement patterns:

	Subject		Direct Object		Indirect Object IPFV	
	SG	PL	SG	PL	SG	PL
M	-e	-ib	a-	ya-	-ha	-ye
F	-o	-ib	wa-	ya-	-we	-ye
N1	-e	-o	a-	wa-	-ha	-we
N2	-o	-o	wa-	wa-	-we	-we

Table 33: Gender-number agreement affixes in Mian (Trans-New-Guinea)
(adapted from Fedden 2011: 163)

As represented in the paradigms above, which is also Fedden's (2011) analysis, the gender-number agreement inflection in Mian is clearly morphomic. The shaded affixes can appear, depending on the gender of the noun in question, in the singular, in the plural, and in both numbers simultaneously. This pattern is unnatural. However, there is an alternative analysis, which Fedden entertains and discards as inferior to the analysis implied in Table 33. This alternative would mean construing gender in Mian as based on the simple dichotomy of masculine vs feminine. The neuters that trigger the same agreement as the feminine singular would be, indeed, feminine, and the neuters that share their agreements with masculine singular would be masculine. If we accepted this gender system, the patterns of morphological identity observed in Table would be simply the result of an over-articulated description of the language. If the orthogonalities presented in Table 33 above were

“fictitious” to some extent, or if we do not identify the relevant features and values involved in a given morphological contrast, we cannot be surprised to find that there are features that operate at cross-purposes to the structure we have posited. Consider the following pattern:

	Statements	Questions
1	-ada	-ari
2	-ari	-ada
3	-ari	-ari

Table 34: Perfective positive suffix in Northern Akhvakh (Creissels 2008)

The distribution of the two allomorphs of the perfective positive appears unmotivated as laid out in Table 34. However, the reason behind the very existence of the terms 'conjunct/disjunct' or 'egophoricity' (Floyd et al. 2018) in linguistic literature is that the distribution above is not unmotivated but rather related to the epistemic properties of speech participants in different illocutionary contexts. If we had identified the “correct” feature involved, then, the first person in statements and the second person in questions would indeed pattern together as a natural class in opposition to the other persons.

In Mian, based on the behaviour of agreement targets, there are, indeed, just three classes of nouns judged by their syntactic behaviour: those that co-occur with affixes -e, a- and -ha; those that trigger -o, wa- and -we; and those that appear alongside -ib, ya- and -ye. If we said that Mian has those three genders, there would be no morpheme in the language, as the exponence patterns displayed in Table 33 would be straightforwardly derived from the gender membership of the nouns. As gender (again as usually defined) is a purely syntactic feature, sensitivity to such a feature would never be labelled morphomic.¹¹

¹¹ Note that depending on our definition of 'morphomic' this is not at all unarguable. Membership to one gender or another (e.g. in French or German) is often arbitrary to some extent and, apart from a few small semantic fields, relatively unpredictable on the basis of meaning. Membership to a particular gender, thus, can be very much like a list: an unstructured set of nouns that belong together simply because they occur with the same forms in their targets. A morpheme is also basically a list: a list of lexemes (in the case of inflection classes) or morphosyntactic contexts (in the case of metamorphemes) that only belong together because they share (some) inflectional properties.

The problem, and the reason why such an analysis is rejected by Fedden, concerns the internal composition of those classes. The membership of each gender would be unusual given the most common understanding of what a gender should be like. One of the genders would contain only nouns referring to more than one entity. Another would only have nouns that denote one entity. The last one would contain singular and plural nouns but, depending of which lexical item, only one of them may belong to the class. For most lexemes, therefore, their gender would differ from singular to plural under this analysis. This intertwinedness of gender and number appears to be undesirable from a theoretical perspective. Gender systems that are orthogonal to number and other features are preferred and regarded as more 'canonical' cases of gender (Corbett & Fedden 2016). Because of this, cases like Mian (or like German or Romanian below) in which the classification suggested by the forms deviates from orthogonality are most usually recast in terms of orthogonal features and values with abundant syncretism:

	Masculine	Neuter	Feminine
NOM/ACC.SG	-ul		-a
DAT/GEN.SG	-ului		-ei
NOM/ACC.PL	-ii	-ele	
DAT/GEN.PL	-ilor	-elor	

Table 35: Romanian definite articles (Gönczöl 2007:30)

	Masculine	Neuter	Feminine
NOM.SG	der	das	die
DAT.SG	dem		der
NOM.PL	die		
DAT.PL	den		

Table 36: German definite articles (partial paradigm)

Romanian shows how, to match our definition of gender, or of what gender can be like, values can be proposed even in the absence of autonomous forms. Saying that, for some lexemes, the singular is masculine but the plural is feminine appears to be unacceptable¹² if we

¹² The size of the class seems to make a big difference, however. The noun *arte* in Spanish, like the neuters

conceive of gender as a system of lexical (nominal) classification. This was exactly the same problem found in Mian.

German, in turn, shows the collapse of gender distinctions in the plural. For the same desideratum of orthogonality, however, we don't usually say that *Auto*, for example, is no longer neuter in the plural. We rather say that neuter plural is simply syncretic with masculine and feminine plurals. But what do these analytical choices or uncertainties mean for the purposes of the morpheme? Note that patterns similar to Mian, that offer alternative analyses, are not difficult to find. Consider, again, the case of Burmeso:

Gender	Conjugation 1		Conjugation 2	
	SG	PL	SG	PL
I Male	j-	s-	b-	t-
II Female, animate	g-	s-	n-	t-
III Miscellaneous	g-	j-	n-	b-
IV Mass nouns	j-	j-	b-	b-
V Banana, sago tree	j-	g-	b-	n-
VI Arrows, coconuts	g-	g-	n-	n-

Table 37: Conjugations in Burmeso (Donohue 2001:100-102)

The gender system of Burmeso seems to be strikingly similar to that of Mian. Three classes of nouns can be found in Burmeso according to the forms they trigger in verbal agreement. It is the requirement of gender-number orthogonality that trebles the number of gender distinctions in the language. Some other times it is the interaction of gender with person that appears to lack the desired orthogonality. Consider the following agreement paradigm:

in Romanian, behaves as masculine in the singular but as feminine in the plural and yet linguists don't usually posit a third gender in Spanish. The same can be said about cases like the Russian second locative. An unarticulated principle of 'diminishing returns' seems to be present in the reasoning of most linguists whereby one has to find a balance between the number of values and the number of exceptions.

	SG			PL		
	M	F	N	M	F	N
1	-ha					
2						
3	-bĩ	-bõ	-ha	-bã	-ha	

Table 38: Subject agreement in Barasano (Tucanoan, Colombia)
(Jones & Jones 1991:73-74)

As shown in the paradigm above, speech-act participants in Barasano trigger the same agreement as neuter nouns independently of the actual gender (M or F) of their referent. An identical situation holds in closely-related Tucano (Baerman & Corbett 2013:4). Analyses of these cases where gender and person, or gender and number, appear not to be orthogonal as suggested by the surface forms often rely on positing a default gender value (neuter in this case) that some items take when they do not 'really' have any gender. Accordingly, patterns like these are not usually described as morphologically stipulated or unnatural.

The main point I am trying to make, thus, is that the orthogonality of features may not always appear to hold when one looks at the forms in a paradigm. In these cases it is tempting to favour accounts that rely on assigning extramorphological properties (like gender) to elements (like pronouns or word-forms) that by definition should not be allowed in principle to have them.¹³ Consider by way of example the paradigmatic distribution of the forms -i and -ni below and their most appropriate analysis:

	SG		PL	
	MASC	FEM	MASC	FEM
1	o-ha-ni	o-ha-ni	i-ha-ni	i-ha-ni
2	ti-ha-ni	ti-ha-ni	ti-kehera-ni	ti-kehera-ni
3	to-ha-i	to-ha-ni	to-kehera-i	to-kehera-ni

Table 39: Declarative form of the copula in Kulina (Pano-Tacanan, Brazil)
(Dienst 2014:229)

¹³ Consider standard definitions whereby, for example, “genders are classes of **nouns** reflected in the behaviour of associated words” (Hockett 1958: 231, cited in Corbett 1991:1, emphasis mine).

First and second persons in the above paradigm appear to take in Kulina the same form *-ni* that is used for the feminine nouns. The pattern is very similar to the one in Barasano. Other agreement targets exist in Kulina that behave similarly (e.g. the topic marker, auxiliaries etc.). Exponence patterns like this one could well lead us to question the usefulness of definitions of gender which, like the one in the above footnote, restrict the phenomenon to nouns. If we reject this, then we are free to argue for the idea that first and second person pronouns have an inherent gender in Kulina and that this fact is straightforwardly reflected in the presence of the *-ni* feminine suffix above. However, as mentioned by Dienst (2014:81-82), it is not always the case that first and second person pattern with the third person feminine in the language. Often, speech act participants appear to collocate with masculine forms instead:

	SG		PL	
	MASC	FEM	MASC	FEM
1	o-zepe	o-zepe	i-zepe	i-zepe
2	ti-zepe	ti-zepe	ti-zepe	ti-zepe
3	zepe	zapa-ni	zepe	zapa-ni

Table 40: Possessive paradigm of 'hand' in Kulina (based on Dienst 2014)

In yet other cases, some agreement targets seem to be able to agree with the actual sex of their referent. Adjectives that inflect for gender, for example, behave unremarkably:

	SG		PL	
	MASC	FEM	MASC	FEM
1	hada-i	hada-ni	hada-i	hada-ni
2	hada-i	hada-ni	hada-i	hada-ni
3	hada-i	hada-ni	hada-i	hada-ni

Table 41: Inflection of the adjective 'old' in Kulina (Dienst 2014)

Note that the frequent differential behaviour of speech act participants and the third person with respect to agreement would result in frequent gender mismatches if that is what affixes *-ni* and *-i* 'mean' in all cases:

16a) hada-i o-ha-ni
 old-m 1SG-be-f

16b) hada-ni o-ha-ni
 old-f 1SG-be-f

'I am old' (uttered by a man)

'I am old' (uttered by a woman) (Dienst 2014:81)

All this quirky behaviour, in my opinion, casts doubt on the very analysis of the forms *-ni* and *-i* as gender forms, especially where they do not appear to make much sense in this respect (e.g. in the first and second person).

Previous examples have involved a lack of orthogonality of gender to some other feature like number, case or person. Of course, cases of apparent non-orthogonalities can involve several features at once. Consider cases like Jarawara, Basketo or Daasanach:

	SG		PL	
	MASC	FEM	MASC	FEM
1	o-man-o		man-o	
2	ti-man-o		man-o	
3	man-o	man-i	man-i	

Table 42: Jarawara (Arawan, Brazil) possessor paradigm of 'arm' (Dixon 2004:315)

According to Dixon's analysis, the 3PL pronoun in Jarawara controls feminine agreement because it is thought to originate from a noun meaning 'people', which may well have been feminine originally. In addition, because of the agreement forms they trigger, he conceives of the 1PL and 2PL pronouns as inherently masculine, regardless of the gender (M, F or mixed) of their referents. The agreement pattern in some verbal paradigms in Omotic is similar:

	SG	PL
1	ʔerer-a	ʔerer-i
2	ʔerer-a	ʔerer-i
3FEM	ʔerer-a	ʔerer-i
3MASC	ʔerer-i	ʔerer-i

Table 43: Basketo (Omotic) affirmative converb of 'know' (Hayward 1991:536)

	SG	PL
1EX	-á	-á
1INC		-í
2	-á	-í
3FEM	-á	-í
3MASC	-í	

Table 44: Benchnon (Omotic) medial verb agreement (Rapold 2006:178)

We can see how in Basketo and closely related Benchnon, masculine singular and (most) plurals sometimes trigger the same agreement suffix while first and second singular show the same form as the feminine singular nouns. This has often been interpreted as a sign that “the different persons of discourse (1s, 2s, etc.) have grammatical gender” (Rapold 2006:178). Other than scholarly tradition and the origin of the forms, there seems to be few reasons to prefer such an analysis over one in terms of person-number agreement. For example, the pattern of formal conflation of medial verbs displayed in Table 44 is contradicted by the one found in final verbs:

	SG	PL
1EX	-ù	-ù
1INC		-ù
2	-ù	-ènd
3FEM	-ù	-ènd
3MASC	-èn	

Table 45: Benchnon (Omotic) indicative final verb agreement (Rapold 2006:179)

To stick to the view that this is gender, one would have to propose two different gender systems operating orthogonally to each other (see Fedden & Corbett 2017) or multiply the number of genders to four to take care of the orthogonality (something Rapold indeed suggests [2006:179]). It is unclear that any of these alternatives are preferable to an agreement system with person-number plus syncretism, especially because such a feature is needed in the language to account for the exponence patterns in other paradigms, such as the polar-question agreement suffixes (Rapold 2007:218), which make the same number of distinctions (8) as the pronouns.

It is tempting to interpret the messiness of patterns like these as a sign that the crucial feature or motivation for the pattern has been missed entirely. If we believe this is the case, meaningless features like gender could always be posited to account for the facts. It has to be kept in mind, however, that there may well be no limits to how 'messy' patterns of syncretism can get. One that definitely ranks very high on this scale (while also showing some similarities to the Omotic paradigms above) is Daasanach:

	SG	PL
1EX	seǔ	sieti
1INC		seǔ
2	sieti	sieti
3FEM	sieti	seǔ
3MASC	seǔ	seǔ

Table 46: Subject agreement of 'walk' in Daasanach (Cushitic)
(Baerman et al. 2005:106 after Tosco 2001)

If portrayed in terms of person and number, the two different forms upon which the agreement system is based apply to a heterogeneous list of morphosyntactic contexts. The form used in the masculine singular is also used in the 1SG, 1INC and 3PL. The form used in the feminine singular is the same that is used in 2 and 1EX. Presented in person-number terms, therefore, this pattern appears to be as arbitrary as it can possibly get.

The alternative, as has been suggested in the literature for some of the patterns above, would be to 'trust the forms' and assume that, like in the systems portrayed in Tables 43 and 44, there is a third feature (e.g. gender) which is independent from the ones represented here (i.e. person and number) and which has just two values (e.g. feminine and masculine). In this particular case, comparative evidence from other Cushitic languages like Oromo or Somali might argue against the latter analysis. The Daasanach paradigmatic arrangement illustrated in Table 46 appears to have originated from a full-fledged person-number agreement system in which phonological erosion has resulted in rampant syncretism:

	Oromo (Ali & Zaborski 1990:5-6)		Somali (Saeed 1999)			
	'go' (past)		'bring' (past)		'say' (past)	
	SG	PL	SG	PL	SG	PL
1	déem-e	déem-n-e	keen-ay	keen-n-ay	idh-i	n-idh-i
2	déem-t-e	déem-t-an	keen-t-ay	keen-t-éen	t-idh-i	t-idhaahd-éen
3.F	déem-t-e	déem-an	keen-t-ay	keen-éen	t-idh-i	y-idhaahd-éen
3.M	déem-e	déem-an	keen-ay	keen-éen	y-idh-i	y-idhaahd-éen

Table 47: Agreement affixes of Oromo and Somali (Cushitic)

Leaving the 1PL aside, where clusivity is complicating the picture, the contexts that take the so-called Form B (e.g. *sieti*) in Daasanach are those that take consonantal affixes in more conservative Cushitic languages while those that take Form A (e.g. *seḏ*) are those that take vocalic or zero affixes. In addition, the various ways in which Form B diverges from Form A in Daasanach (e.g. *ces/yes* 'kill.PERF', *guuranna/guurma* 'migrate.IMP', *leeḏi/leeti* 'fall down.PERF' etc.), are also, for the most part, readily interpretable as the outcome of run-of-the-mill sound changes affecting consonants differently in different phonological environments (e.g. *guuram-t-a/guuram-a > *guuranta/guurama > *guuranta/guurma > guuranna/guurma). Speakers of Daasanach, when faced with these sound changes appear to have responded by re-organizing their person-number paradigm into one with only two arbitrarily distributed forms (Sasse 1976).

Reanalyses do take place all the time in the course of diachronic change in language so this origin is no guarantee that the Daasanach system is to be analyzed synchronically in the same terms (i.e. with person and number features) as the agreement systems of Somali or Oromo. We are, however, able to conclude that the system has, at least, the same sound-change-triggered origin as some of the most prototypical morphomic patterns.

Last of all and despite the efforts that here and elsewhere have been devoted to arguing for one of the two alternatives, I believe that analyzing these patterns in terms of gender agreement or conceiving them instead as autonomous morphological syncretisms is not very different in practice. After all, both analyses involve assigning a common abstract property (whether a gender value or a morphological syncretic index) to a disparate set of elements

which are irreducibly list-like. These abstract properties would not have any real meaning but would constitute merely a formal device to capture the (semantically) arbitrary morphological patterns that we observe. This is precisely what formalizations of the morpheme have traditionally involved (e.g. Aronoff 1994, Round 2013). Cases like the ones presented throughout this section will therefore be considered morphomic in this dissertation whenever they meet my definitional criteria for morphomhood otherwise.

The problem of the complex identity of the 1PL, of syntactic unnatural classes, and of gender systems “gone wrong” have been three illustrative cases of the uncertainties we may face in some cases when deciding whether concrete formatives are morphomic or not. These difficulties may persist, even after deliberately narrowing down the object of study to those contexts (tabular inflectional paradigms) whose systematic structure eases the identification of natural and unnatural classes. The remaining analytical uncertainty involved in these and other unforeseen cases will simply have to be accepted at this stage. Further investigation of these borderline cases would be, of course, most welcome.

3 Empirical aspects on the morphome

3.1 What else can be morphomic?

In the cases discussed in Chapter 2 and in almost all the literature on the morphome, it is inflectional formatives which are discussed as the object of analysis. However, not only inflectional forms may have an unnatural distribution in the paradigm. Other morphological phenomena (e.g. syncretism, heteroclisys, defectiveness etc.) can apply differently in different parts of the paradigm and single out morphosyntactically unnatural sets of cells as their domain of application. Some other times, even derivational structures may be thought of as paradigmatically organized and liable to displaying morphomic structures. This chapter, thus, explores the possibility of morphomic phenomena in less obvious domains.

3.1.1 Syncretism

Syncretism and morphomes are intimately linked, since both are concerned with (total or partial) morphological identities. Many of the examples of morphomes that will be presented in this dissertation, thus, will involve whole-word syncretism:

	Present		Progressive	
	SG	PL	SG	PL
1EXCL	ur-o	ur-ciga	ur-ca	ur-ciga
1INCL		ur-cina		ur-cina
2	ur-o	ur-cini	ur-ca	ur-cini
3	ur-o	ur-o	ur-ca	ur-ca

Table 48: Partial paradigm of 'drink' in Mongo Daju (Dajuic, Chad) (Avilés 2008)

As you can see in Table 48 above, syncretism interacts in various ways with the SG+3PL morphomic pattern present in Daju. First, within a given tense (e.g. the present), there is whole-word syncretism of the person-number cells that make up the morphome (i.e. all are *uro*) whereas the cells outside of it are kept distinct (i.e. *urciga*, *urcina*, *urcini*). On the other

hand, the distinction between the tenses (i.e. present vs progressive) is only drawn within the morpheme cells (i.e. *uro* vs *urca*) whereas the cells outside of it are underspecified for tense (i.e. *urciga* vs *urciga*). Different configurations can be found elsewhere:

	Alpago (Zörner 1997)				Standard Italian			
	Indicative		Subjunctive		Indicative		Subjunctive	
	SG	PL	SG	PL	SG	PL	SG	PL
1	'dorme	dor'moŋ	'dorme	dor'mone	'dormo	dor'mjamo	'dorma	dor'mjamo
2	'dorme	dor'me	'dorme	dor'mede	'dormi	dor'mite	'dorma	dor'mjate
3	'dorme	'dorme	'dorme	'dorme	'dorme	dormono	'dorma	dormano

Table 49: Present tense of 'sleep' in two Romance varieties of Italy

The cells constitutive of the N-morpheme have become whole-word syncretic (/ 'dorme/) in the variety of Alpago.¹⁴ Thus, not only person and number but even the category of mood appears to be neutralized within the morpheme in this paradigm. This is not the case outside of the morpheme cells .

There is a different way of exploring the relationship between morphomicity and syncretism, however. If we consider sensitivity to particular features, instead of forms *per se*, morphomic structures may emerge even in quite familiar places:

	Balochi (Axenov 2016:164)				German			
	SG		PL		SG		PL	
	PRES	PAST	PRES	PAST	PRES	PAST	PRES	PAST
1	in	un	an		e	∅	en	
2	ay		it		st		t	
3	t	∅	ant		t	∅	en	

Table 50: Sensitivity of person-number agreement suffixes to tense

¹⁴ This is a carefully chosen example, as other verbs in this variety do not share this syncretism. However, one may wonder whether morphomic affinity may favour the emergence of diachronic whole-word syncretism (consider the typological parallel of Dhaasanach [Table 46] compared to Oromo and Somali [Table 47]).

In both Balochi and standard German, for example, the 2SG and the PL person-number suffixes (an unnatural class) show syncretism between past and present. Syncretism and overdifferentiation, thus, can have morphomic distributions. These might be subject to a tendency to have more distinctions/allomorphs in more frequent cells and less in the less frequent ones (see Table 55 for the frequency of different person-number cells).

3.1.2 Heteroclisis

Similar to syncretism, the paradigmatic distribution of a pattern of heteroclisis may align to a meaning distinction (consider e.g. Czech *pramen* ‘spring’ which declines like a soft-masculine noun in the singular but as a hard-masculine noun in the plural, see Stump 2006:280), or may instead split the paradigm in unnatural ways:

	‘woman’		‘president’		‘philosopher’	
	SG	PL	SG	PL	SG	PL
NOM	žena	ženy	předseda	předsedové	filosof	filosofové
GEN	ženy	žen	předsedy	předsedů	filosofa	filosofů
DAT	ženě	ženám	předsedovi	předsedům	filosofovi	filosofům
ACC	ženu	ženy	předsedu	předsedy	filosofa	filosofy
VOC	ženo	ženy	předsedo	předsedové	filosofe	filosofové
LOC	ženě	ženách	předsedovi	předsedech	filosofovi	filosofech
INSTR	ženou	ženami	předsedou	předsedy	filosofem	filosofy

Table 51: Pattern of heteroclisis of Czech *předseda* ‘president’ (Stump 2006:290)

Czech *předseda* behaves as a hard feminine noun in the NOM, GEN, ACC, VOC, and INSTR cases in the singular, and as a hard masculine elsewhere, i.e. in DAT and LOC singular and in the plural. This, thus, could be described as a morphomic pattern of heteroclisis.

The link between heterocclisis and more traditionally morphomic phenomena (e.g. stem alternations) is well known (see e.g. Maiden 2018b: 55, 220). Thus, particular morphomic stems (e.g. PYTA) may have a particular inflectional class (e.g. non-1st conjugation) associated with them in a way that, when the inflectional class membership of the lexeme elsewhere differs, it leads to heterocclisis. Thus, the PYTA forms of first conjugation *andar* ‘walk’ and *estar* ‘be’ in Spanish take non-first conjugation endings (e.g. *anduv-iste*, *anduv-ieras*, *estuv-iste*, *estuv-ieras*). Sometimes, however, the same pattern of heterocclisis is found even in the absence of any pattern of stem alternation:

	Conjugation I	‘give’	Conjugation II
Infinitive	am-ar	d-ar	corr-er
2SG.PRES.IND	am-as	d-as	corr-es
2SG.PRES.SBJV	am-es	d-es	corr-as
2SG.IPF.IND	am-abas	d-abas	corr-ías
1SG.PRET	am-é	d-í	corr-í
2SG.PRET	am-aste	d-iste	corr-iste
3SG.PRET	am-ó	d-ió	corr-ió
2SG.IPF.SUBJ	am-aras	d-ieras	corr-ieras

Table 52: Some inflectional forms in Spanish

In the Spanish verb *dar* ‘give’, the morphosyntactically arbitrary paradigm subset known as PYTA is singled out by heterocclisis alone, instead of by stem allomorphy, thus constituting a prime example of the connection that may sometimes be found between morphomicity and heterocclisis.

3.1.3 Overabundance and defectiveness

Morphomicity constitutes an affinity in the exponence of a morphosyntactically arbitrary set of paradigm cells. Thus, we would expect that idiosyncratic exponences like overabundance (Thornton 2012) and defectiveness (Baerman et al. 2010), may also be morphomically

distributed in the paradigm. This has been shown to be indeed the case (see e.g. Albright 2003, and Maiden & O’Neill 2010). In this section I will briefly present the issue in connection with the L-morpheme of Spanish.

In the paradigmatic domain of the L-morpheme, near-suppletive stem alternations (e.g. *caer/quep-o*) and velar stem augments (e.g. *pon-er/pon-g-o*) are in competition with non-alternation (e.g. *met-er/met-o*). That is, in verbs of the second and third conjugation, which is where the phenomenon may take place, alternation and non-alternation are common. In those verbs which are frequent enough (e.g. *caer/caigo*, *venir/vengo*, *decir/digo*, *vencer/venzo* etc.), alternation or lack thereof is just lexically stipulated. In verbs which are infrequent but which are of a phonological structure which does not ever show alternation (i.e. those whose stem does not end in a vowel or in /n/, /l/, /s/ or /θ/) there is also no uncertainty. Many infrequent verbs which are derivationally created out of adjectives by means of the suffix *-ecer*, in turn, invariably must include the velar augment (e.g. *engrandecer*, *palidecer* etc.) and so there is also no uncertainty for verbs belonging to this big (+300) class of verbs. The problem arises when the verb is not of this class, is infrequent, and is of a phonological structure which seems like could maybe require some L-morphomic exponence. In some of those cases, normative grammar either prescribes one of the two possibilities (e.g. *mecer* does not alternate according to RAE but *pacer* and *asir* do) or offers two or more correct alternatives (e.g. for *roer*, the forms *roo*, *roigo* and *royo* are all accepted and for *yacer* the same happens with *yazgo*, *yazco* and *yago*).

Despite the recommendations of prescriptive grammarians, the truth is that, whenever there is this uncertainty, speaker choices vary: nonstandard forms like *paza* (without the velar augment) or *mezca* (with the augment) do occur, as well as the prescribed variants *pazca* and *meza*. They constitute cases of overabundance which extend to every cell within the L-morpheme:

	IND	SUBJ	IND	SUBJ
1SG	mezo/mezco	meza/mezca	roo/roigo/royo	roa/roiga/roya
2SG	meces	mezas/mezcas	roes	roas/roigas/royas

Table 53: L-Morpheme overabundance in two Spanish verbs, partial paradigms.

In my opinion, however, the most accurate usage description is that, because of the uncertainty they have to face in these paradigm cells, language-users tend to avoid the forms altogether in those seldom-used verbs which are not as entrenched as more frequent ones. It seems to be the case that, somewhat paradoxically (since they are definitionally opposite phenomena), the border between overabundance and defectiveness is fuzzy here:

	<i>mecer</i> 'rock'				<i>aburrir</i> 'bore'			
	PRES IND	PRES SUBJ	IMPERF	PAST	PRES IND	PRES SUBJ	IMPERF	PAST
1SG	2	0	15	2	212	33	70	93
2SG	3	0	1	0	82	21	21	11
3SG	200	3	205	39	540	100	330	160
1PL	4	0	0	1	32	6	18	7
2PL	0	0	0	0	9	1	0	0
3PL	75	3	96	12	240	32	100	52

Table 54: Token frequency counts in CORPES of two Spanish verbs

As illustrated in the above Table 54, the forms belonging to the L-morphome are exceedingly infrequent in *mecer* and in other verbs where competition can be found between alternants:

	L-morphome-overabundant verbs				Other verbs			
	PRES IND	PRES SUBJ	IMPERF	PAST	PRES IND	PRES SUBJ	IMPERF	PAST
1SG	0,16%	0%	2,11%	0,14%	7,22%	0,76%	1,68%	3,57%
2SG	0,85%	0%	0,33%	0,01%	2,21%	1,18%	0,56%	0,56%
3SG	30,15%	0,33%	29,33%	13,82%	28,43%	3,72%	10,17%	12,24%
1PL	0,58%	0%	0,07%	0,18%	3,23%	0,24%	0,39%	0,58%
2PL	0,01%	0%	0%	0%	0,33%	0,02%	0%	0,01%
3PL	9,40%	0,12%	10,71%	1,70%	13,30%	2,32%	4,20%	3,38%

Table 55: Token frequency proportion in the two groups

As shown in Table 55, the frequency of verb forms within the L-morphome usually amounts to around 15% of the surveyed forms. By contrast, in the case of those verbs¹⁵ with L-

¹⁵ The token frequencies of the verbs *entender*, *entrar*, *tostar*, *ser*, *mentir*, *aburrir* and *perseguir* has been

morphome overabundance, those forms represent less than 1%. Overabundance and defectiveness, thus, affect in the same way all the cells within the L-morphome, which reaffirms the deep morphological affinity of these forms in synchronic grammar, even in the case of non-canonical (zero or multiple) inflectional forms.

3.1.4 Morphomicity in derivation

Because of its greater semantic and formal predictability, it is in the domain of inflection, particularly in conjunction with tabular paradigmatic structure, where one expects the notion of the morphome to be most useful. One could even argue that in the case of prototypical morphemes, the existence of at least two orthogonal dimensions/features in a paradigm is necessary to identify unmistakable cases of morphomicity (i.e. morphological affinities which are morphosyntactically unnatural regardless of any eventual posited feature structure). For this practical reason, the focus of this thesis will be on inflection.

It must be stressed, however, that derivation is by no means incompatible with morphomicity. It is, for example, a crucial part of Latin's third stem, discussed by Aronoff (1994) as a prime example of a morphome. As mentioned elsewhere, the lexicon is full of cases where a resonance does not correspond straightforwardly to any shared semantics (e.g. *deceive*, *receive*, *conceive* etc.). In many cases the formal similarities may be accidental and grammatically irrelevant. In other cases, however, there is evidence that those resounding elements do constitute a grammatical unit, despite the lack of semantic content. Words with those bound stems, for example, may participate together in unpredictable morphophonological processes in word formation (*deception*, *reception*, *conception* etc.). There is also psycholinguistic evidence (Giraudo et al. 2016) that these words even prime one another beyond what the shared form would account for, suggesting, therefore, a deeper grammatical affinity of some sort.

surveyed as the control group and those of *mecer*, *asir*, *yacer* and *roer* for the group of L-morphome-overabundant verbs.

The concept of the morpheme can also be useful, thus, for lexical organization. Exploring, for example, the domain of terms related to ethnicity, Schalchli & Boyé (2017) find evidence for systematic syncretisms like the ones usually described as morphomic:

	Ethnicity	Area	Language	Ethnicity	Area	Language
N	français	France	français	russe	Russie	russe
Adj	français	français	français	russe	russe	russe

Table 56: Some french terms related to ethnicity (Schalchli & Boyé 2017)

The decision to restrict myself in this dissertation to inflectional paradigms is to be understood, therefore, as a convenient way of narrowing down the object of study of the present dissertation, and not as an advocacy for morphomicity or paradigmatic structure being exclusively inflectional phenomena.

3.2 The scale from natural to unnatural

As usually construed (e.g. Bybee 1985:118, Haspelmath & Sims 2010:2, Blevins et al. 2016:275, Booij 2016:104), morphology is the branch of linguistics that studies the covariation of meaning and form in the word. Constructivist models assume that elements of form exist *in order to* express meaning or morphosyntactic distinctions. The architecture of language as a whole is usually posited to proceed from the most abstract components to the more concrete ones (i.e. pragmatics > semantics > morphosyntax > phonology). This hierarchical architecture is explicitly assumed in many models (e.g. In Functional Discourse Grammar: Hengeveld & Mackenzie 2008). In models with this overall architecture, morphology is usually considered post-syntactic (e.g. in Anderson's [1992] A-morphous Morphology, and in Distributed Morphology: Halle & Marantz 1994) so that syntax (also semantics) is usually hypothesized to be morphology-free.

The precedence of meaning over form and the subordinate status of form to the more abstract layers of grammar is implicitly or explicitly assumed by most researchers and frameworks. Distributed Morphology, for example, “asserts that morphs (including zeroes)

realize single syntactic terminals, and so at the point of vocabulary insertion it establishes a one-to-one mapping between exponenda and exponents” (Luís & Bermúdez-Otero 2016:1). Realizational models (as in Matthews 1965) also posit rules of grammar that spell out in surface abstract morphosyntactic properties. Thus, although it seems that these should be just two sides of the same coin, it is often emphasized that it is the abstract grammatical properties that determine form, and not form that signals the grammatical properties.

If, as suggested by this approach, elements of form exist merely to express morphosyntactic distinctions, formal structure is necessarily dependent on meaning and should ideally be completely isomorphic with syntactic and semantic structure. That is, straightforward, one-to-one, biunique mappings are expected between form and meaning. Formal similarity should echo morphosyntactic similarity and conversely, morphosyntactic differences should be signaled by differences in form. Such “canonical” structures are not difficult to find:

	downwards	on-same-plane	upwards
Towards EGO	ter	tek	tem
Away from EGO	jer	jek	jem

Table 57: Teribe (Chibchan, Panama) deictic-directional movement verbs (Quesada 2000: 67)

	SG	DU	PL
1	na	nato	nakare
2	ni	nito	nikare
3	nu	nuto	nukare

Table 58: Suena (Trans-New-Guinea) pronouns, INCL forms excluded (Wilson 1974: 16-17)

	SG	PL
1	təmən	təmdan
2	nəmən	nəmdan
3	gəmən	gəmdan

Table 59: Kusunda (Isolate, Nepal) verb *əm* 'eat', realis (Watters 2006: 60)

From the perspective of Canonical Typology (Corbett 2005, Brown & Chumakina 2013), the above cases can be considered canonical inflectional paradigms (Stump 2015: 35-41). As mentioned by Round & Corbett (2017:54), “[c]anonically, a feature value would be realized uniformly by just one, overt exponent in all contexts, and that exponent would be distinct from all others in the system.”

Every formal element in Tables 57 to 59 follows this ideal and adopts a natural-class distribution. In morphology, morphosyntactic natural classes are those which can be straightforwardly assigned a meaning or morphosyntactic property because, distributionally, they coincide completely with some morphosyntactic feature value or bundle of values. Thus, in Suena pronouns, the formative *-to* appears in every dual pronoun and never outside the dual. Similarly, *-i* appears in all second person pronouns and only there. That structures like the ones of Teribe, Suena and Kusunda exist speaks of the importance of meaning and morphosyntactic features for the organization of linguistic structure, both in the lexicon and in the grammar if these are believed to be different modules (cf. Booij & Audring 2017). The probability of such perfectly isomorphic structures emerging by chance would be infinitesimal and yet they are found frequently across natural languages. It is hardly ever questioned, therefore, that meaning is of the utmost importance in grammar and that morphosemantic values like [plural] or [addressee] are crucial when explaining morphological structure. We must therefore begin by acknowledging that “[t]here is a universal semiotic principle favouring biunique matching of lexical signata and signantia” (Maiden 2011c: 266)

That perfectly isomorphic structures exist does not mean, of course, that they are the only possibility in natural languages. There is disagreement, however, as to whether one-to-one mappings are or not the ones most frequently found across languages:

“the “one meaning - one form” principle is actually used very sparingly.” (Bybee 1985:209)

“[a] biunique relation between meaning and form is the most common relation in inflectional morphology” (Aalberse 2007:114)

To be able to assess these claims one would need a thorough quantitative typological investigation coupled with clear criteria for segmentation (see Section 62), the adoption of an uncontroversial feature inventory and structure, and clear criteria for distinguishing homophony, polysemy and vagueness in meaning. Consensus on these issues is unlikely to be reached in the near future and so I will refrain from making the assessment of these claims one of my goals in the present dissertation. It should suffice by now, and it is hardly controversial, to claim that deviations from biunique mappings are at least not uncommon.

As was mentioned before, morphosyntax is usually assumed to be further up in the hierarchical structure of grammar than phonology. According to constructivist models, morphosyntactic functions pre-exist, so to speak, and are simply realized in surface by exponence rules. But this reasoning is clearly perverse. First of all, as is well known, we deduce whether a morphosyntactic distinction (e.g. tense or number) is present or absent in the grammar of a language precisely by looking for any formal correlates along those lines. In addition, sometimes, a unitary treatment concerning form can lead us to posit a single grammatical unit (i.e. a morphosyntactic feature) even in the absence of any shared extramorphological properties:

Although series are conventionally assigned morphosyntactic labels, such as 'past', 'aorist', 'perfect', etc., the forms in a series often share a common base rather than a set of grammatical properties. (Blevins 2016:90)

There is, therefore, a tendency to overinterpret morphological terms. A good case in point are the various functions of tenses (like for example the Spanish 'imperfect') and of cases (like for example the Latin 'ablative'). This shows that, at least sometimes, formal identity leads linguists to posit grammatical categories (i.e. features or values) for which no further evidence exists. Similarly, if we happen to observe lexeme-dependent formal distinctions with no clear semantic correlate we just posit *ad hoc* features like gender¹⁶ or inflection classes. The result

¹⁶ This need not have a 'semantic core'. See, for example, gender in Uduk, for which Killian (2015: 62) comments: "All nouns in Uduk, including proper nouns, are allocated into one of two possible grammatical genders, labelled as *Class I* and *Class II*. Grammatical gender is not based on biological sex, and assignment into these classes is largely arbitrary. Semantics in fact appears to play almost no role in the choice of which gender a noun is placed in, even with a small semantic group related to humans or animate nouns."

is that, consciously or not, we are building up (bi)uniqueness into our descriptions of morphological systems.

With this *modus operandi*, it is hardly surprising, therefore, that we should find strong parallelisms between formal and morphosyntactic structure. And yet, despite this approach, we do find many cases in which, unlike in Tables 57 to 59 before, the mapping between form and features is not canonical. I will present throughout the next pages a few cases in order of increasing deviation from the biuniqueness ideal presented before:

	SG	PL
1	lahem	lahemi
2	lahesh	laheni
3	lahet	lahen

Table 60: Albanian *lah* 'wash' present non-active (Newmark et al. 1984: 59)

The case of Albanian may seem straightforward, since all the morphosyntactic distinctions are drawn in the formal paradigm. However, there is a non-trivial difference with respect to the examples that were presented before. Unlike in those perfectly isomorphic examples, formal elements in Table 60 do not reflect the assumed morphosyntactic structure. For example, despite the morphosyntactic affinity (i.e. shared person value) of 2SG and 2PL, there is no formal reflection of that affinity. Thus, no element of form can be consistently identified with a morphosyntactic feature value. That is, we cannot identify in Table 60 a marker for [addressee] or for [plural].

We are then forced to make reference not to single features, but to feature bundles. Thus, the distribution and meaning of the suffix *-mi* has to be described as a conjunction of features (first person+plural). This (i.e. cumulative exponence) might be regarded as problematic, given that syntax is sometimes posited to manipulate features but not to have access to specific combinations of feature values (Corbett 2016: 72).¹⁷ The issue boils down to the

¹⁷ In the absence, in languages like Albanian, of morphological evidence for independent features like person and number we may wonder what is the need to assume those categories in the first place. An alternative analysis, though by no means an unobjectionable one, would imply simply 'listening' to the morphology and analyzing each of the 6 morphosyntactic entities in Table 60 as irreducible morphosyntactic objects.

theoretical boundary between syntax and morphology and will not be discussed here further. Another subtle deviation from the canonical isomorphic inflectional paradigm can be illustrated by the Russian past tense inflection:

	SG	PL
M	rabotal	rabotali
F	rabotala	
N	rabotalo	

Table 61: Russian past imperfective forms of the verb 'work'

We can see here that Russian verbs in the past tense agree in gender and number. However, these features do not appear to be orthogonal and, judged by the morphology, gender agreement does not occur in the plural. These cases, where sensitivity to a feature is seemingly lost completely within a certain domain (e.g. in the plural in this case) are usually not considered exceedingly problematic. The form in question (i.e. *rabotali*) is usually considered to be simply unspecified for gender. This means that it is usually considered uninformative regarding gender rather than ambiguous between the different gender values. The form, still has, therefore, a clear single meaning [plural]. The same analysis may be unsuitable for slightly different phenomena in other languages:

	SG	DU	PL
1	wun	an	ñan
2.FEM	ñən	bər	gwur
2.MASC	mən		dəy
3.FEM	lə		
3.MASC	də		

Table 62: Manambu (Ndu, New Guinea) personal pronouns (Aikhenvald 2008: 66)

Manambu personal pronouns, for example, distinguish gender in the second and the third person singular but not in the first. In the dual, moreover, the distinction between second and third person is also missing. We thus cannot say that features such as gender or person are relevant or irrelevant in the domain of a certain number. Finer-grained conditions are required to describe the distribution of forms and sensitivity to a particular feature.

When distinctions are seemingly 'lost' in one domain in comparison with another, there will be one form in one domain corresponding to several forms in another domain. These syncretisms will depend on feature structure to be classified as a natural or unnatural class:

	SG	DU	PL
NOM	rꙗka	rꙗčě	rꙗky
VOC	rꙗko		
ACC	rꙗkꙗ		
GEN	rꙗky	rꙗku	rꙗkъ
LOC	rꙗčě	rꙗkama	rꙗkaxъ
DAT			rꙗkamъ
INS	rꙗkojꙗ		rꙗkami

Table 63: Old Church Slavonic *rꙗka* 'hand/arm'

We can see in Table 63, that in Old Church Slavonic, much like in Manambu pronouns, there are fewer distinctions drawn in the plural than in the singular, and less in the dual than in the plural.¹⁸ In the dual, all three different case forms appear to be vague or underspecified compared to the forms in other numbers. If there were some inherent affinity of genitive and locative, and of dative and instrumental, the forms 'rꙗku' and 'rꙗkama' would still constitute single morphosyntactic objects with a single meaning. However, the problem is exacerbated when the forms the formal distinctions in different domains cross-classify:

	Object suffixes	Subject suffixes	
		Realis	Irrealis
2SG	-o	-lu	-le
1SG		-ie	-fe
3SG	-fo	-lee	-be
2PL	-mo	-mo	-bule
1PL		-ne	-bile
3PL	-te		

Table 64: Kwomtari (Kwomtari-Nai, New Guinea) person agreement (Spencer 2008: 107)

¹⁸ This is correlated to (and probably caused by) the differences in the frequency of use of each number and must constitute a cross-linguistically robust tendency.

Kwomtari, as presented in Table 64, sometimes conflates the values for first and second person plural (see object suffixes), but some other times the values of first and third person plural (see subject suffixes). In both cases there is evidence that suggests systematicity since both patterns (i.e. 1=2 and 1=3) are found two times with different exponents, the former in the singular (-o) and in the plural (-mo), and the latter in the realis (-ne) and irrealis (-bile). These cross-classifying identities render an analysis of these formal neutralizations problematic for morphological models with a rigid and hierarchical feature structure.

There are approaches to morphology, however, which are based on the 'lexicalization' or 'spelling' of "adjacent" features (e.g. geometrical: McCreight & Chvany 1991, nanosyntactic: Caha 2009). These frameworks, because they are less restrictive, would still be able to account for cross-classifying syncretisms like the ones in Kwomtari. Provided that the values are ordered so as to make syncretic forms contiguous (in the case of Kwomtari the order would have to be 2>1>3 as shown in Table 64), a single form could spell out any combination of adjacent values. There are cases, however, that defy any such orderings:

	I	II	III	IV
SG	che:	tósè-gɔ	k'on-dɔ	tòn
DU		tósè	k'on	
PL	che:-gɔ		k'on-dɔ	

Table 65: Kiowa number marking (Wunderlich 2012: 178 after Wonderly et al. 1954)

Kiowa number syncretisms, as presented in Table 65, make it impossible to arrive at any fixed order such that formal identity occurs only between adjacent values. Analyses which rely on morphosyntactic affinities or on covert feature structure as an explanation for syncretism may need, therefore, some extra machinery even for some one-dimensional syncretisms (note that all the morphological syncretisms that have been presented until now occurred between cells that shared at least one value).

Bi- or tridimensional formal confluations, in turn, also vary in the extent to which they can be analyzed as the expression of a well-defined value. Consider the following paradigm:

	SG	DU	PL
1	fecemin	fecohul	fecomun
2	fecem	fecebil	
3	feceb		

Table 66: Amele (Trans-New-Guinea) verb 'see' perfect switch reference (Roberts 1987)

In the paradigm of Table 66, the form *fecebil* conflates both 2 and 3, and DU and PL. Under the right feature structure, however, the distribution of this form can be characterized simply as non-speaker non-singular. It would thus have a morphosyntactically coherent description and may be regarded as a natural and morphemic exponence.

Patterns of formal identity involving L-shaped or T-shaped configurations are more problematic to account for. Note, however, that some of these syncretisms may receive a semantic explanation if the conflated values do indeed have some affinity:

	1SG.O	1PL.O	2SG.O	2PL.O
1SG.S	-	-	-tan	-tadiž
1PL.S	-	-	-tadiž	-tadiž
2SG.S	-samak	-samiž	-	-
2PL.S	-samiž	-samiž	-	-
3SG.S	-samam	-samiž	-tanzat	-tadiž
3PL.S	-samiž	-samiž	-tadiž	-tadiž

Table 67: Erzya (Uralic) subject-object conjugation, partial paradigm (Rueter 2010)

Despite their seemingly haphazard paradigmatic distribution, Erzya *-samiž* and *-tadiž* can be analyzed as markers of argument plurality in conjunction with some person object value. If we say that *-samiž*, for example, means ‘first person object in the presence of a plural argument’, its distribution is successfully described, even if such a description may be understood to do strain exponence rules as usually conceived.

In those cases where the horizontal and vertical axes display completely different feature values, however, analyses based on extramorphological affinity become less plausible. In the Papuan language Benabena, for example, there is a paradigmatic pattern (affecting stem alternants and the allomorphy of certain other elements) whereby the singular and the first person forms behave in the same way:

	SG	DU	PL
1	bu-ʔohube	bu-ʔohuʔibe	bu-ʔohune
2	bu-ʔahane	bi-ʔehaʔibe	bi-ʔehabe
3	bu-ʔehibe	bi-ʔehaʔibe	bi-ʔehabe

Table 68: Verb 'go' in Benabena, past tense (Young 1964:48)

This category (i.e. SG and/or 1) is labeled 'monofocal' by Young, whereas the other cells are labeled 'polyfocal'. The terms, of course, suggest the possibility of a semantic affinity of some sort between the cells. Regardless of the merits of this specific analysis, these L-shaped patterns do seem to appear occasionally in other areas of language. Carstairs-McCarthy (1998) for example, notes that terms with disjunctive meanings (X or Y), although infrequent, are sometimes possible in lexical semantics, provided that the two values intersect and their conjunction (X and Y) can be referred to by the same name:

	Clambering	No clambering
Upward	climb	climb
Not upward	climb	-

Table 69: Meaning features of *climb* (Jackendoff 1985)

Jackendoff (1985) also explains how the (or his) use of the verb 'climb' is appropriate to describe actions involving motion upwards and/or performed with the use of limbs. If grammatical formatives behave regarding meaning in the same way as lexemes, morphosyntactic distributions like that of Benabena's 'monofocal' stem could indeed count as well-defined in a single lexical entry and need not be necessarily morphomic. Of course, given the rarity of these semantics, one could well proceed in the complete opposite way and argue that the semantics of 'climb' are morphomic.

Be that as it may, L- or T-shaped patterns can (and often do, see Section 4.1.3.1) arise in one step from 'natural' morphosyntactic distributions by means of 'natural' morphosyntactic or semantic extensions. This suggests that they are “not as far” from them in this respect. Since naturalness is, as shown throughout this section, a scalar dimension, morphological patterns can easily be found which are a bit further away from the isomorphic ideal:

	SG	DU	PL
1	-onji	-ontae	-ontone
2	-onji	-onji	-ontifi
3	-i	-onji	-ontifi

Table 70: Same-subject non-future medial verb agreement in Safeyoka (Angan, New Guinea) (West 1973:10)

in Table 70, the suffix *-onji* appears in all non-plural forms except in the 1DU and 3SG. Patterns like these are, thus, two steps away from a morphosyntactically natural distribution. The morphosyntactic contexts where *-onji* appears constitute, however, still a contiguous region in the paradigm space since all its cells are connected by changes of just one feature value at a time. This fact is crucial in some models of morphological exponence like McCreight & Chvany's (1991) geometrical approach.

Other patterns of morphosyntactic distribution, however, are problematic even for these models. Exponents which display a 'diagonal syncretism' (see Table 71 below) do not occupy a contiguous morphosyntactic space of the paradigm:

	SG	PL
ILL	maddja	maddjid
LOC	maddjest	maddjin
COM	maddjin	maddjuvui' m
ABE	madditää	maddjitää

Table 71: Skolt Saami (Uralic) *maadd* 'base', partial paradigm (Feist 2011:146)

The difficulty of capturing the unnatural distribution of an exponent increases with the number of disjoint contexts in which it appears. In addition, as was mentioned before (Section

2.2), it may also make a difference (and it is at any rate more problematic in theoretical analyses relying on defaults and blocking) whether or not an exponent's distribution is interlocked with that of another unnaturally-distributed exponent. Consider, for example, the following paradigm (see also Daasanach in Table 46):

	SG	DU	PL
1	-ve	-ve	-pe
2	-pe	-ve	-ve
3	-ve	-ve	-ve

Table 72: Subject agreement in Yagaria, partial paradigm
(Stump 2015: 128 after Haiman 1980)

These cases, where formatives have a distribution completely orthogonal to the assumed morphosyntactic feature structure, and where descriptions based on mechanisms like blocking also fail, are as far as one can get from the isomorphic ideal that many theoretical approaches to morphology start from or assume. They are, therefore, troublesome¹⁹ for many formal models, especially for those refusing to grant any independent status to morphology. As a consequence, not all linguists would interpret in the same way²⁰ the data which have been presented throughout this section. However, the fact that these patterns (Tables 46 and 72), although infrequent, are possible in natural languages, seems to suggest that form-function isomorphism is, if anything, a cross-linguistic tendency, and not the only possible organizational principle for (inflectional) morphology. Isomorphism, thus, might constitute a

¹⁹ Maybe as a result of their problematic nature, some deviations from the isomorphic form-meaning ideal have attracted the interest of morphologists and typologists and have been given specific dedicated terms such as syncretism, deponency, defectiveness, overabundance, morpheme etc.

²⁰ Biuniqueness is sometimes 'enforced' by linguists even where the empirical facts do not favour a one-to-one mapping interpretation. For example, in those cases where the distribution of a formative cannot be accounted for in plain morphosyntactic terms, its underlying distribution or meaning is often hypothesized to be different from the one we see in surface. It can be either a superset, in those cases where 'blocking' supposedly takes place, or a subset, in those cases where rules of referral are allegedly operating. However, as argued, for example, by Blevins (2016:214), and despite the widespread use of those devices in formal models of morphology, there is not enough evidence that these paradigmatic readjustment rules are real. They may be largely formal machinery simply aimed at aligning formatives with morphosyntactic properties.

Also because of the expectation that form must be subordinate to function, many analyses have been devoted to trying to find some (at times obscure) semantic affinity between homophonous formatives (e.g. Bittner 1995, Leiss 1997) or between the various uses of unitary morphological objects such as cases (Jakobson 1984).

tendency for paradigmatic organization which can be overridden under the right circumstances. An exhaustive typological study of those cases is likely to provide valuable information about the nature and principles of the morphology.

3.3 Independence of phonology

In the definitions of 'morpheme' that circulate in the literature, it is not seldom that one finds reference to the phonological component of language. O'Neill (2013:221), for example, reports that one of the most usual senses of 'morpheme' refers to “meaningless stems which show identical patterns of allomorphy and which cannot be reduced to any coherent **phonological**, semantic or syntactic generalization” (emphasis mine). Disagreements on whether some particular (stem-alternation) pattern should be considered morphemic or not (e.g. Anderson 2011 vs Maiden 2011b) have also sometimes revolved around the independence of that pattern from concrete phonological environments.

Morphemes, however, are precisely about form, so applying the criterion that a morpheme has to be independent from phonology is much more difficult than could be thought initially. Consider, for example, the following verbal paradigms:

	Russian <i>peč'</i> 'bake'		Spanish <i>plegar</i> 'fold'	
	SG	PL	SG	PL
1	pek-u	peč-ëm	'pljego	ple'gamos
2	peč-ěš'	peč-ëte	'pljegas	ple'gajs
3	peč-ët	pek-ut	'pljega	'pljegan

Table 73: Stem alternation patterns in a Russian and a Spanish verb

In the above subparadigm of Russian *peč'*, the distribution of k vs č as the last consonant of the stem is perfectly correlated to the presence of vowel u vs ë (/jo/) in the following suffix. In Spanish, the use of a vowel /e/ or diphthong /je/ in the stem is correlated as well to the absence or presence of stress in that particular syllable. Furthermore, from a historical perspective those are indeed the phonological contexts that were responsible for the stem

alternations displayed by these verbs. As a consequence, many researchers and analyses present these patterns of stem alternation as phonologically conditioned, which in the view of many implies that they would not possibly be morphomic (although see Maiden [2017] for a different opinion).

To decide whether the alternations in Table 73 are morphomic, it has to be kept in mind that in the synchronic (i.e. living, productive etc.) phonological or phonotactic rules of Russian and Spanish there is absolutely no evidence for a rule which transforms /k/ into $\sqrt{t_c}$ before /jo/ (consider words like Russian *likjor* 'liquor') or which turns /e/ into /je/ in the presence of stress in Spanish (both can be found in stressed and unstressed positions). Adjudicating these patterns to the phonological component, for example by positing diacritics that diphthongating /e/s will have but others lack, does not appear to do much more than recapitulate the historical phonological changes that gave rise to those patterns. As a synchronic analysis, this approach is unsuitable, in my opinion, and mainly an *ad hoc* strategy that does not get us any closer to understanding the phenomenon. The only outcome of these approaches, as far as I can see, is to shrink (or eliminate) the domain of morphology at the cost of enlarging that of phonology.

Trying to explain the distribution of *pek-* vs *peč-* as being determined by that of the suffixes -u vs -ë (or *vice versa*, for that matter) is also simply transferring the burden of the explanation to some other place in the system. There is no reason why one of the two patterns would require an explanation while the other one would not. The same thing happens with diphthongization in Spanish. Explaining the paradigmatic distribution of the N-morpheme (e.g. of /je/) by deriving it from stress ignores the fact that the location of stress is unpredictable in the language. As pointed out by Esher (2015), the paradigmatic distribution of rhizotony in the Spanish paradigm, is, therefore, not a phonological matter but a morphological one. Knowing the paradigmatic distribution of rhizotony is not enough either, as different verbs (even of comparable phonological and phonotactic profiles (e.g. *podar/podo* 'prune' vs *poder/puedo* 'be able to') behave differently as for whether they undergo diphthongization or not.

If we are to remain as close as possible to the empirical data and avoid making problematic assumptions, all we can note in cases like the Spanish and the Russian ones is that there is a perfect correlation between the distributions of two different formal elements which do not need to occur together synchronically but do so in these paradigms. The existence of a correlation, in these cases, could well point to more and not to less morphomicity for these morphological patterns. The morphological affinity assumed by the N-morpheme, after all, is reproduced in a verb like *plegar* not only once, but twice, with two different exponents (i.e. presence vs absence of a glide /j/, and presence vs absence of rhizotony).

When one goes beyond a simple description of formal distributions, however, the situation becomes more complicated. It is difficult to ascertain, for example, whether or to what extent these morphological correlations (e.g. between diphthongization and stress in Spanish) are synchronically active or merely constitute a perpetuation of the context that historically originated the alternations. Disagreements are ubiquitous in this respect. Bermúdez-Otero & Luís (2016), for example, argue for the synchronic relevance of the correlation. They offer evidence of dialectal analogical developments rendering 1PL.SUB *puédamos*²¹ (compared to standard *podámos*). O'Neill (2011), however, argues that we cannot infer causation from this. All we can note, in his opinion, is a correlation between stress and stem vowel, which, although they change together in *puédamos*, do change separately in other dialects (see for example Benasque Aragonese in Table 145).

According to the division of labour between phonology, morphology and syntax which will be adopted throughout the present dissertation, non-automatic alternations like the ones of Spanish and Russian above will not be considered phonological processes. Accordingly, morphological patterns will not be excluded from the ranks of morphemes just because they are coextensive to some phonological environment. That said, clear-cut cases of automatic phonological determination do exist, of course. When some formal alternation is the result of a phonological process that is synchronically active in the language it will not be considered an object of analysis for morphology. This is the case of the following stem alternation:

²¹ This change would still leave the direction of causation possibly undetermined (i.e. is it the diphthong which requires stress or is it stress that requires the diphthong), but would constitute evidence that the correlation between stress and diphthongization is not synchronically spurious.

	SG	DU	PL
1	nə-mraj-iyəm	nə-mre-muri	nə-mre-muru
2	nə-mraj-iyət	nə-mre-turi	nə-mre-turu
3	nə-mre-qin	nə-mre-qinat	nə-mre-qina

Table 74: Declension of the adjective *mraj*- 'lucky' in Alutor (Chukotko-Kamchatkan) (Kibrik et al. 2004:84)

As explained by Kibrik et al. (2004:84) the alternation aj/e is phonologically determined in Alutor. The sequence /aj/ always becomes /e/ syllable-finally and the sequence ajC is not allowed in the language. Because of this, the alternation above will be considered here a matter for phonological and not morphological analysis.²² The paradigmatic patterns that result from automatic phonological processes, therefore, will not be discussed here.

Another issue that has to be settled in relation to the independence of morphemes from 'form' as a whole is the following: It has sometimes been argued in the literature (e.g. O'Neill 2011, Nevins et al. 2015) that, in order for something to qualify as a morpheme, one needs to find that a pattern of formal identity is independent of its actual formal instantiation. A representative expression of that sentiment is the following:

*the clearest and most predictive aspects of the L-morpheme theory says that it is about an abstract relation of complete identity between these cells of the paradigm **without any reference to their phonological form or phonological naturalness.*** (Nevins et al. 2015:8, emphasis mine)

The reasoning appears to be the following: to be sure that an unnatural morphological identity is systematic and not an instance of accidental homophony, morphologists have

²² Note that the non-morphomic character of even these patterns is not self-evident. Some diachronic developments suggest that language users sometimes do acquire phonologically derivable patterns redundantly. In Vinzelles Occitan, for example, (see Morin 1988), an apparently stress-determined (allomorphic) stem alternation (e.g. 'love' 1SG.PRES.IND 'amə vs 2SG.PRES.IND e'ma:) was apparently not analyzed as such by language users since, when they analogically leveled stressed within the present tense, the allomorphy was preserved (i.e. 1SG.PRES.IND 'amə vs 2SG.PRES.IND 'ema:). Similarly, research in East Kiranti (Herc forthcoming-a) suggests that phonologically-derivable patterns of stem alternation are redundantly acquired, since they show otherwise unexpected diachronic resilience and influence in affixal allomorphy.

usually required that the identity be repeated across various different forms. Because of this, the pattern of formal identity in these cases has sometimes been conceived and formalized as independent of the actual forms involved. This is, in my opinion, a *non-sequitur*.

Patterns of morphological identity, I believe, are hardly ever independent of its particular instantiations. This is intuitively sensible, since it is forms (i.e. concrete forms) that reveal morphological structure to the language users in the first place. It could be thought, admittedly, that in the most extreme cases (i.e. given enough variation and unpredictability in form) a pattern of formal identity could plausibly be generalizable (e.g. in wug tests) to formally unattested cases. Consider again the case of Daasanach:

	SG	PL
1INC	-	A
1EX	A	B
2	B	B
3FEM	B	A
3MASC	A	A

Table 75: Daasanach verb form allomorphs (Baerman et al. 2005:106 after Tosco 2001)

Every Daasanach verb has two forms, distributed in the way indicated above. The formal differences between the forms are varied: *seḏ-sieti* 'walk', *kufi-kuyyi* 'die', *guurma-guuranna* 'migrate', *yes-ces* 'kill' etc. If the formal differences between the so-called Form-A and Form-B were totally unpredictable in the language (which they are not) this would mean that both forms would simply need to be memorized for every single lexeme. If this were the case, any wug-forms (e.g. *mefu* vs *pala* for 1SG and 1PL.EX) might well lend themselves to be mapped into morphosyntactic values by adhering to the pattern above despite the total novelty of the alternation involved.

However, most cases of morphemes (and most morphemic oppositions too for that matter) are not instantiated with such a wide array of forms. Consider, for example, the case of the Romance N- or L-morphemes. The number of forms associated with each of the patterns is usually relatively small. The Spanish N-morpheme, for example, is instantiated always by

diphthongization (either o/u>ue or e>ie). Verbs with formal alternations along those lines, therefore, can easily be classified as 'N-morphomic' whereas other kinds of alternations, because of their exceptional character within the system, would face a greater difficulty in fitting into the N-morphome.

Consider, for instance, the history of the Spanish verb *llevar* 'take'. In Old Spanish, the verb was a diphthongizing one *levar-lievo*, in line with hundreds of other verbs in the language. At some point, however, a sound change occurred whereby /lje/ > /le/:

	<i>levar</i> 'lift/take'		<i>levar</i> (after sound change)	
	IND	SUBJ	IND	SUBJ
1SG	lievo	lieve	llevo	lleve
2SG	lievas	lieves	llevas	lleves
3SG	lieva	lieve	lleva	lleve
1PL	levamos	levemos	levamos	levemos
2PL	levades	levedes	levades	levedes
3PL	lievan	lieven	llevan	lleven

Table 76: Old Spanish verb *levar* in two different stages

After the sound change /lje/ > /le/, the former monophthong-to-diphthong alternation (e-ie) was replaced by a consonant₁-to-consonant₂ alternation (l-λ). A formal alternation that was present in hundreds of other verbs was thus replaced by one which was formally unique in the language. As a result, and despite the great frequency of use of the verb, the alternation became unstable and was eliminated from the language soon after it arose. The stems *lev-* and *llev-* spread from their former niches into the rest of the paradigm. The ensuing two lexemes (i.e. *llevar* and *levar*) eventually specialized into different meanings, maybe to avoid complete synonymy (see Carstairs-McCarthy 2010):

	<i>llevar</i> 'take'		<i>levar</i> 'lift' (an anchor)	
	IND	SUBJ	IND	SUBJ
1SG	llevo	lleve	levo	leve
2SG	llevas	lleves	levas	leves
3SG	lleva	lleve	leva	leve
1PL	llevamos	lleveamos	levamos	levemos
2PL	llevais	lleveis	levais	leveis
3PL	llevan	lleven	levan	leven

Table 77: Modern Spanish outcomes

The history of this verb shows that, at least sometimes (I would argue most of the times), the actual formal instantiation of a morpheme does matter a great deal. If a lexeme does not have the 'right' formal alternation, language users may fail to associate them to others, even in the face of an identical paradigmatic distribution.

The history of Spanish verbs with stem-vowel raising alternations also bears witness to the same “inseparability” of a paradigmatic pattern and its formal instantiation. In medieval Spanish, a number of verbs in the third conjugation with stems containing a mid-vowel in the infinitive (e.g. *pedir* ‘request’, *cobrir* ‘cover’) had the corresponding high vowel in many parts of their paradigm (e.g. in the 1SG present indicative *vido*, *cubro*):

	PRES.IND	PRES.SBJV	IPF	PAST	IPF.SBJV I	IPF.SBJV II	FUT	COND
1SG	vido	vida	pedía	pedí	pidiera	pidiese	pediré	pediría
2SG	vides	vidas	pedías	pediste	pidieras	pidieses	pedirás	pedirías
3SG	vide	vida	pedía	pidió	pidiera	pidiese	pedirá	pediría
1PL	vedimos	vidamos	pedíamos	pedimos	pidiéramos	pidiésemos	pediremos	pediríamos
2PL	vedís	vidáis	pedíais	pedisteis	pidierais	pidieseis	pediréis	pediríais
3PL	viden	vidan	pedían	pidieron	pidieran	pidiesen	pedirán	pedirían

Table 78: Distribution of the high vowel stem in Spanish rising verbs

Both the e/i and the o/u alternating verbs followed the same paradigmatic template shown in Table 78. It is, however, revealing, that, while the e/i alternation has been preserved robustly into the modern language, the o/u alternation has largely disappeared:

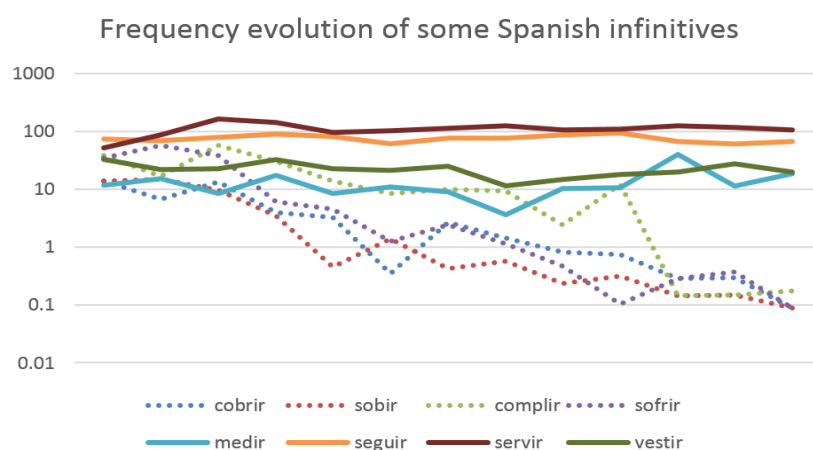


Figure 1: The demise of the o/u alternating verbs in Spanish

Figure 1 shows the frequency (in hits per million words) of various infinitive forms in CORDE between the years 1490 and 1610. As the graph shows, whereas the e/i alternating verbs have been preserved, o/u alternations have been lost to paradigm levelling. Largely in the 16th century, the high-vowel stem was generalized throughout the paradigm (i.e. *cubrir* ‘cover’, *subir* ‘go up’, *cumplir* ‘fulfill’, *sufrir* ‘suffer’). As Figure 1 shows, the differential diachronic treatment of e/i and o/u alternations is remarkable even in verbs with similar (and relatively high) token frequencies. Phenomena like these suggest that, even if some formalizations of the morpheme have involved dissociating paradigmatic distributions from their concrete exponents, this is sometimes²³ largely a convenient fiction.

This, I believe, explains experimental results like the ones reported in Nevins et al. (2015) where they presented speakers of Portuguese with wug-verbs that showed formal alternations (p-f, t-s, k-x) unparalleled in the Portuguese verbal system. Their results showed that language users usually did not extend the wug-alternations by adhering to the distribution of stem alternants in L-morpheme verbs. Because the formal alternations that were presented to the Portuguese speakers did not match those of the L-morpheme verbs in their language, they did not know what to make of a completely alien alternation. This, I

²³ Sometimes one does come across developments which seem to demand that patterns have an existence of and by themselves independently of any particular form(ative). Some suppletive alternations (e.g. Fr. *vais* vs *allons*), for example, were innovated on the basis of patterns they share little formal similarity with. Another interesting example (discussed in Maiden 2018b:208) is found in the variety of Romance spoken in Maragatería, where the vowels in the verb ‘play’ have been reversed compared to their distribution in Spanish. Compare Maragatería *jugo jugas juga jugamos juegais jugan* to Spanish *juego juegas juega jugamos jugáis juegan*.

believe, is not very surprising. In the same way as the history of *llevar*, and of o/u alternating verbs, it reminds us that morphomic paradigmatic patterns (probably also morphemic and “regular” patterns, see Albright 2002 and Albright & Hayes 2003) are most likely not independent from their actual formal instantiations.

3.4 Economy and the morphome

Deciding between alternative analyses of a phenomenon is sometimes extremely difficult. In the simplest case, an analysis that covers 100% of the facts is to be preferred to another one that does not. However, once two different analyses or formalizations cover the facts perfectly it is difficult to decide which one, if any, is “better” or more cognitively plausible. Discussion in these cases revolves usually around matters of 'elegance' or 'economy'. However, there is hardly any consensus as to how these notions are to be understood and therefore whether they favour one analysis or the other in specific cases.

In this section I will compare how different analyses fare in unnatural exponences of various degrees of complexity. This will help us assess whether different systems favour different analyses or whether the same rules of the game should be used at all times. Concretely, we will assess how recourse to Paninian blocking or to autonomous morphological rules can impact the descriptive length of different systems. Consider first the following inflectional patterns from Yagaria:

	Interrogative			Indicative			Subordinate			Coordinate			Apodosis		
	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL
1	-ve	-'-ve	-pe	-e	-'-e	-ne	-ma	-'-ma	-pa	-ga	-'-ga	-na	-hine	-'-hine	-sine
2	-pe	-'-ve	-ve	-ne	-'-e	-e	-pa	-'-ma	-ma	-na	-'-ga	-ga	-sine	-'-hine	-hine
3	-ve	-'-ve	-ve	-e	-'-e	-e	-ma	-'-ma	-ma	-ga	-'-ga	-ga	-hine	-'-hine	-hine

Table 79: Allomorphy of Yagaria mood affixes (Stump 2015: 128, after Haiman 1980)

A total of 8 other moods have been omitted from the paradigm above. These show the same patterns of syncretism as the moods displayed in Table 79. They have been left out for the

sake of brevity and also because they involve the same formal alternations as some of the moods above. In addition, as shown above, a glottal stop appears before every dual form (and only in dual forms) and has therefore been left out from the rest of the discussion because it will not make a difference between different analyses. In a mapping that cannot rely on autonomous morphology, nor on blocking and defaults, the descriptive length of the system above would be considerable:

1SG/DU ²⁴ .INTER > -ve1	1SG/DU.IND > -e1	1SG/DU.SUB > -ma1	1SG/DU.COORD > -ga1	1SG/DU.APOD > -hine1
2DU/PL.INTER > -ve2	2DU/PL.IND > -e2	2DU/PL.SUB > -ma2	2DU/PL.COORD > -ga2	2DU/PL.APOD > -hine2
3.INTER > -ve3	3.IND > -e3	3.SUB > -ma3	3.COORD > -ga3	3.APOD > -hine3
1PL.INTER > -pe1	1PL.IND > -ne1	1PL.SUB > -pa1	1PL.COORD > -na1	1PL.APOD > -sine1
2SG.INTER > -pe2	2SG.IND > -ne2	2SG.SUB > -pa2	2SG.COORD > -na2	2SG.APOD > -sine2

In an analysis where Paninian blocking is permissible (but where morphology cannot have its own rules beyond this one) the descriptive length of the system would be reduced:

Superset Principle

1PL.INTER > -pe ₁	1PL.IND > -ne ₁	1PL.SUB > -pa ₁	1PL.COORD > -na ₁	1PL.APOD > -sine ₁
2SG.INTER > -pe ₂	2SG.IND > -ne ₂	2SG.SUB > -pa ₂	2SG.COORD > -na ₂	2SG.APOD > -sine ₂
INTER > -ve	IND > -e	SUB > -ma	COORD > -ga	APOD > -hine

The same as in an analysis with autonomous morphology but without blocking:

1SG/DU > μ	2DU/PL > μ	3 > μ	1PL > λ	2SG > λ
μ INTER > -ve	μ IND > -e	μ SUB > -ma	μ COORD > -ga	μ APOD > -hine
λ INTER > -pe	λ IND > -ne	λ SUB > -pa	λ COORD > -na	λ APOD > -sine

Last of all, obviously, the descriptive length of the system would be reduced most if we could make use simultaneously of the machinery of Paninian blocking and of autonomous morphological rules:

²⁴ Combinations of values like 'singular' and 'dual', 'dual' and 'plural', 'first' and 'second' person, or 'second' and 'third' person will be considered natural semantic classes for the purposes of the exponence rules here. It must be noted, however, that this fact (i.e. the existence of a non-flat feature structure) will be helping us reduce the number of rules needed but is, in turn, an additional element of complexity that should not be taken for granted.

Superset Principle 1PL > λ 2SG > λ

INTER > -ve IND > -e SUB > -ma COORD > -ga APOD > -hine

λINTER > -pe λIND > -ne λSUB > -pa λCOORD > -na λAPOD > -sine

Let's take a look now at a somewhat less complex exponence pattern from the variety of Nivkh (Isolate) spoken in the east of the island of Sakhalin:

	Non-future						Future					
	Narrative		Distant		Coordinating		Narrative		Distant		Coordinating	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	-t	-t	-tot	-tot	-ta	-ta	-n	-n	-non	-non	-na	-na
2	-r	-t	-ror	-tot	-ra	-ta	-r	-n	-ror	-non	-ra	-na
3	-r	-t	-ror	-tot	-ra	-ta	-r	-n	-ror	-non	-ra	-na

Table 80: Nivkh converb inflection (Gruzdeva 1998: 56, Nedjalkov & Otaina 2013: 40-42)

The exponence of the Coordinating and Distant converbs differs predictably from that of the Narrative (addition of -a and addition of -oC respectively, where the quality of C is decided on the basis of the previous suffix). Because they are straightforward one-to-one mappings they will be the same regardless of the analysis and will not be considered in ensuing discussion. Without any machinery whatsoever, the exponence mappings would have to be stated as follows:

2/3SG > -r 1SG.NFUT > -t₁ PL.NFUT > t₂ 1SG.FUT > -n₁ PL.FUT > -n₂

With Paninian blocking but without independently morphological rules:

Superset Principle

2/3SG.NFUT > -r₁ 2/3SG.FUT > -r₂ NFUT > -t FUT > -n

With independent morphological rules but no blocking:

1SG > λ PL > λ
 λNFUT > -t λFUT > -n 2/3SG > -r

Independent morphological rules and blocking, unlike in Yagaria, would never apply together profitably in this system. We can see how for this particular pattern, of intermediate complexity, morphological machinery does not result, unlike in Yagaria, in a great simplification of the exponence mappings. Consider last of all the simplest unnatural pattern of syncretism, one that is not repeated with any other formatives. This is the case, for example, of the following diagonal syncretism:

	SG	PL
NOM	viessu	viesu-t
ACC/GEN	viesu	viesu-id
ILL	viessu-i	viesu-ide
LOC	viesu-s	viesu-in
COM	viesu-in	viesu-iguin
ESS	viessu-n	viessu-n

Table 81: North Sami (Uralic) *viessu* 'house' (Hansson 2007)

Leaving aside consonant gradation and the exponence of those cases that are not involved in the syncretism (again these mappings will be the same regardless of the analysis) we would have to devise the following exponence rules in an analysis with no blocking and no autonomous morphology:

LOC.SG > -s LOC.PL > -in₁ COM.SG > -in₂ COM.PL > -iguin

If we allowed blocking but not autonomous morphological entities:

Superset Principle

LOC.SG > -s [] > -in COM.PL > -iguin

If we had to make use of autonomous morphology instead to capture the syncretism, the mapping operations would be the following:

COM.SG > λ LOC.PL > λ
 λ > -in LOC.SG > -s COM.PL > -iguin

I have been showing throughout this section that the relative economy (measured in number of mapping operations)²⁵ of the different analyses and formal mechanisms depends on the degree of complexity (e.g. allomorphy) of the system where they are applied. The following Figure 2 summarizes this fact:

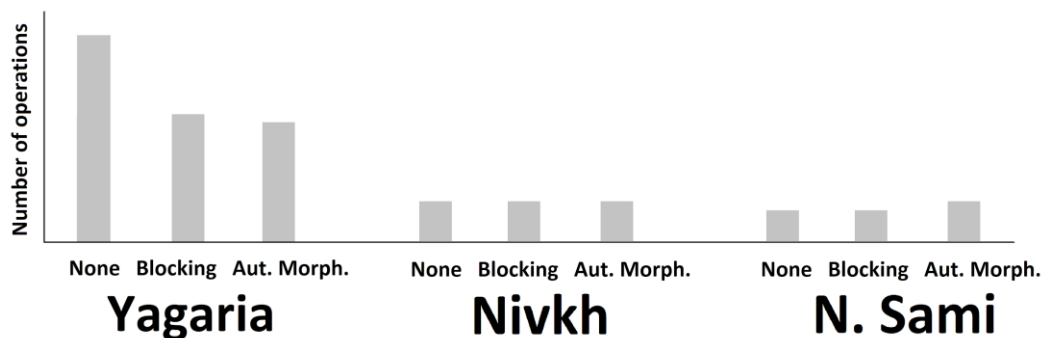


Figure 2: Comparison of the realizational economy of different analyses

Figure 2 shows how the economy effect of incorporating an autonomous morphological component is felt only in the inflectional systems of greater complexity (e.g. Yagaria). We can see how in the simplest, one-off cases of unnatural syncretism (North Sami), an autonomous morphological analysis seems to be actually more uneconomical than the competing alternatives. This is the reason why I will set as a minimum requirement for morphomic status in Chapter 5 that a pattern be instantiated with at least two different exponents.

Alongside these considerations of economy one could entertain the 'elegance' of the analyses as a separate factor. Those that have to resort to separate lexical entries or mappings for systematically homophonous elements could well be considered less elegant than those where distributional systematicities are acknowledged in the formalism. Under this criterion, some of the earlier analyses would be inelegant (in red below):

²⁵ It is not evident by any means that this is the “right” measure of realizational economy. One could think of alternative ones, for example the number of characters needed to represent the full set of rules.

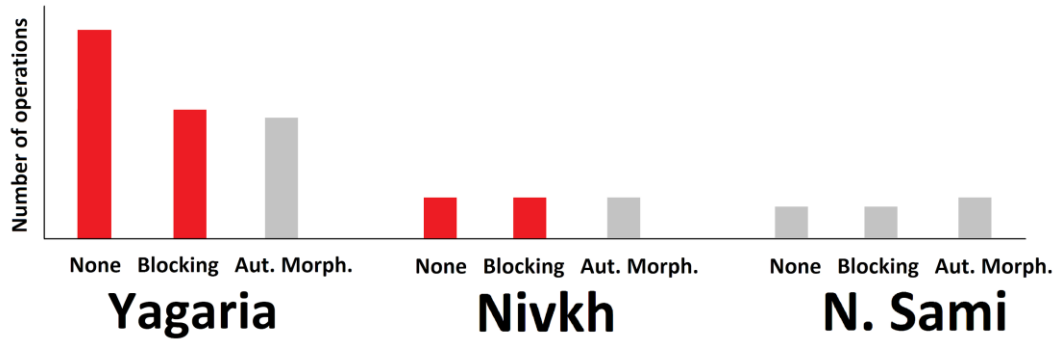


Figure 3: Comparison of the realizational economy and elegance of different analyses

All this being said, it has to be recognized that there is absolutely no consensus in the discipline concerning what should count as more 'elegant' or more 'costly'. The operations that the two figures above count and simply add up are of very different types and we just do not know how/whether the costs of a competing-rule resolution operation can compare to those of a straightforward content-to-form mapping operation. We also have no reason to suppose that all the operations of the same kind should be equivalent in this respect. It has to be acknowledged, thus, that we have absolutely no idea as to how/whether these considerations of formal economy and elegance of the analysis map onto the language users' cognitive representations or onto actual psycholinguistic processing costs.

If we believe that language change can be used as a window into cognitive architecture, the little evidence we do have concerning the above patterns actually seems to point toward the relative insignificance of the matters that we have discussed throughout this section. Judged by the figures above, for example, there would be little reason to pursue an autonomous morphological analysis of the North Sami syncretism and yet it appears that in some dialects the pattern analyzed here has been spread to new contexts with different formatives (see Hansson 2007), which seems to suggest that language users did analyze the syncretism in those terms at some point. It remains to be understood (even imperfectly), therefore, how the factors that this section has dealt with, as well as others like frequency or segmentability, guide the cognitive representations or analyses of language users. This is the reason why a typological investigation like this dissertation cannot rely on them for the identification of its object of inquiry.

4 Morphemes in diachrony

Synchronic states are often explained in science with reference to diachrony. This is probably unsurprising, since, in the words attributed to biologist and classicist D'Arcy Thompson "everything is the way it is because it got that way". Also in linguistics, language change is often taken to be one of the main sources for true explanation.

The case of morphemes is a bit exceptional in that, here, diachrony has come to be almost embedded into the very definition of the phenomenon. Morphemes (also morphemes, see Wurzel 1989:29) have come to be often defined as a "cognitively real" unit in the minds of language users. However, because we have basically no access whatsoever to the inner cognitive representations of language in the mind, language change has come to be used in their stead as a diagnostic of when a putative morpheme is real or not. Thus, if a given set of paradigm cells behaves in an internally homogeneous way in processes of analogical change, so the reasoning goes, then it must be cognitively real in the minds of speakers. If no such evidence exists, then the forms at stake may just be stored in the lexicon and constitute merely "diachronic junk" with no synchronic grammatical import.

As argued elsewhere (Section 2.6), however, it is not only impractical but also unreasonable to define or diagnose a synchronic grammatical phenomenon in diachronic terms. Consequently, diachrony and morphomhood will be regarded as independent here, which will allow us to scrutinize and typologize the different ways in which morphemes may arise and disappear from a language. This will be the purpose of the present section.

4.1 The emergence of morphemes

4.1.1 Sound change

The morphologization of sound changes or their effects is probably the first thing that comes to mind when one thinks of the possible diachronic sources of morphemes. This is the

ultimate²⁶ origin of most of the morphemes which have been discussed in the literature more frequently (e.g. the N-, L- and U-morphemes of Romance). The label ‘sound change’, however, can refer to somewhat different processes of morpheme emergence. Sometimes, as in the classical Romance morphemes mentioned above, sound changes, in conjunction with different phonological environments, can create a formal alternation where there was formerly none. Consider also the following cases:

	Pre-Jabuti		Jabuti	
	SG	PL	SG	PL
1	*tʃabä	*hi-tʃabä	habä	hi-rabä
2	*a-tʃabä	*a-tʃabä	a-rabä	a-rabä
3	*tʃabä	*tʃabä	habä	habä

Table 82: The verb ‘get tired’ in two stages in Jabuti (Macro-Je) (Pires 1992:45-46)

	Pre-Old High German		Modern German	
	SG	PL	SG	PL
1	faru	farem	fahre	fahren
2	*faris	faret	fährst	fahrt
3	*farit	farant	fährt	fahren

Table 83: The verb ‘drive’ in two different stages of German (Braune & Reiffenstein 2004)

	Ancient Greek ²⁷		Modern Greek	
	SG	PL	SG	PL
1	‘e-dēsa	e-’dēsamen	‘e-desä	‘desame
2	‘e-dēsas	e-’dēsate	‘e-deses	‘desate
3	‘e-dēse(n)	‘e-dēsan	‘e-dese	‘e-desan

Table 84: Aorist past tense ‘tie’ in different stages of Greek (Holton et al. 2012)

²⁶ Of course morphomic patterns may be subsequently replicated and reinforced analogically, but this is done often on the basis of the original alternations created by regular sound change.

²⁷ The antepenult syllable placement of the stress in the past tense in Greek, together with the shapes of person-number suffixes, meant that sometimes the past tense prefix was stressed and sometimes unstressed. More concretely, this augment, as it has been called traditionally, was stressed in the SG and 3PL cells and unstressed elsewhere.

In Jabuti, an originally non-alternating stem is split into two different stems as a result of sound changes involving intervocalic voicing plus certain subsequent point and mode of articulation changes. In German, anticipatory distant vowel assimilation to a following /i/ (i.e. i-Umlaut) creates stem-vowel apophony where there was formerly none. In Greek, in turn, a past tense vocalic prefix is deleted in unstressed pretonic contexts. Despite their differences, all of these conditioned²⁸ sound changes have resulted in forms with unnatural morpho-syntactic distributions.

The three cases presented above are, thus, similar in this respect but also show non-trivial differences in others. For example, the phonological environment that gave rise to the alternation is still in place in Jabuti (and arguably in Greek) but has been erased in German. Although the forms can be said to be completely morphologized in both cases (because the formal alternations are no longer synchronically productive phonological processes), only in the latter case (i.e. in German) can the alternation potentially become informative and participate fully in the system of formal contrasts in the language. Note, in this respect, that the sound-change-triggered alternation /a/ vs /e/ has now become the only formative distinguishing 3SG and 2PL present in many German verbs like *fahren* above.

Despite their differences, both in Jabuti, in Greek, and in German, sound change has generated from scratch an alternation between two formerly identical forms. I will call this type of morphome origin the **formal divergence** scenario. The research undertaken here for the compilation of the morphome database (Chapter 5) has demonstrated this to be a very common origin of morphomes (see the ones of Ayoreo, Daasanach, French, Kele, Iraqw, Saami or Spanish for morphomes of comparable diachronic origin).

²⁸ Conditioned sound change takes place when some segment or sequence behaves differently in different phonological environments. Of course this is opposed to unconditioned sound change, where every single occurrence of a segment changes into something else. To put an example, in the history of Spanish, /j/ changed to /x/ everywhere. Although I know at present of no example of a morphome arising from an unconditioned sound change, this is entirely possible logically iff the segment's new pronunciation is the same as that of a pre-existing phoneme. This could then result in an accidental homophony between formerly distinct word forms which could later be interpreted as systematic and grammatically meaningful by language users (see Tables and).

Sound changes, thus, can produce formal variation in various ways. Starting from a situation without any formal alternation (i.e. A~A), this can happen when a phonological change modifies the quality or quantity of the form in certain contexts (i.e. A~B, see German above) or differently in different contexts (i.e. B~C, see Jabuti above). Sound changes can also generate a contrast between a form A and its absence (i.e. A~∅, see Greek above).

In an orthogonal contrast to this one, sound-change-generated morphomic structures also show variation with regard to another aspect. The sound change(s) that give rise to them can take place in different *loci* with respect to the morpheme. Phonological change can target the paradigm cells constitutive of the morpheme or can instead target their complement set. These two scenarios are not mutually exclusive since, sometimes, the sound changes that create a morpheme may happen both in the morpheme cells and in their complement set.

A well-known but particularly appropriate example of this last scenario is the L-morpheme of Romance. Its emergence can be traced back to two independent sound changes. One involved the palatalization of velars before front vowels (see *nascere* below) and the other the palatalization of non-labial consonants before /j/ (see *medire*). Because front vowels and yods were in complementary distribution in the paradigm (e.g. 'do': *fak-jo*, *fak-is*, *fak-it* etc.), the contexts where the two changes occurred were the exact opposite of each other, which means that they gave rise to the same pattern of stem alternation:

	Old Spanish <i>nascere</i> 'be born'				Portuguese <i>medir</i> 'measure'			
	Indicative		Subjunctive		Indicative		Subjunctive	
	SG	PL	SG	PL	SG	PL	SG	PL
1	nas[k]o	nas[ts]emos	nas[k]a	nas[k]amos	meço	medimos	meça	meçamos
2	nas[ts]es	nas[ts]edes	nas[k]as	nas[k]ades	medes	medis	meças	meçais
3	nas[ts]e	nas[ts]en	nas[k]a	nas[k]an	mede	medem	meça	meçam

Table 85: Two verbs illustrative of the Romance L-morpheme (Herce 2019a:113)

Note that the shaded cells of *nascere* are those where palatalization (i.e. *naskes*>*nastses*) did not happen whereas those of *medir* are those where palatalization (i.e. *metjo*>*meço*) did happen. Regardless of their origin, the shaded cells became the odd-ones-out (i.e. a minority

alternant) within their paradigms, which is probably the reason, as I have argue in Section 5.1.3, why these cells, rather than their complement, are the ones which are taken in the literature to constitute a morpheme. See also the case of Svan in Section 5.2.59 for another morpheme with a similar diachronic origin. For morphemes created by sound change(s) in the morpheme cells, see those of Chinantec and Pite Saami, and for morphemes created by sound change in the morphome’s complement cells see Luxembourgish and Wutung.

These cases where sound change creates morphemes by generating formal variation or alternations from scratch (i.e. AA>AB) contrast to other cases where sound change leads to a **formal conflation** (i.e. AB>AA) instead. In Livonian, for example, comparative evidence suggests a systematic morphological syncretism between 1SG and 3SG derived from a sound-change-generated formal conflation which became analogically extended:

	Estonian				Livonian			
	PRES		PAST		PRES		PAST	
	SG	PL	SG	PL	SG	PL	SG	PL
1	tapan	tapame	tapsin	tapsime	tapab	tapam	tapiz	tapizm
2	tapad	tapate	tapsid	tapsite	tapad	tapat	tapist	tapist
3	tapab	tapavad	tapis	tapsid	tapab	tapabəd	tapiz	tapist

Table 86: The verb ‘kill’ in two Finnic languages (Baerman 2007a)

Comparison with other closely-related languages like Estonian suggests that, as a result of the regular loss of word-final /n/, two formerly distinct word forms (1SG.PAST and 3SG.PAST) became formally identical in Livonian. This initial accidental formal conflation was analyzed as grammatically relevant by language users and was subsequently extended to the present, where the two forms would not have become syncretic by regular sound change. The accidental formal merger of formerly distinct forms as a consequence of sound change is, therefore, another possible source of unnatural syncretisms.

Another quite revealing example of this type of morpheme emergence can be found in the history of Scandinavian. The infinitive and the 3PL present forms were different in Proto-Germanic. However, sound changes (consider the loss of various final unstressed vowels, the loss of word-final -n etc.) made the two forms fall together by Old Norse:

	Proto Germanic 'drive', infinitive: *faraną				Old Norse 'drive', infinitive <i>fara</i>			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	*farō	*faramaz	*fōr	*fōrum	fer	fōrum	fōr	fōrum
2	*farizi	*farid	*fōrt	*fōrud	ferr	farið	fórt	fóruð
3	*faridi	*farandi	*fōr	*fōrun	ferr	<i>fara</i>	fór	fóru

Table 87: Indicative Mood inflection of 'drive' in two stages of Germanic (Zoëga 1910)

This arbitrary phonological identity, however, was actively preserved in diachrony and even extended occasionally to other forms. Thus, *preterito presentia*, for example, because of their use of etymologically past forms in the present, would never have developed a syncretism of 3PL.PRES and INF (consider the paradigm of *eiga* in Table 88). However, because of the overwhelming whole-word syncretism of these two paradigm cells across the lexicon, some *preterito presentia* acquired this morphological trait analogically by borrowing the 3PL.PRES -u suffix of these verbs into the infinitive. Thus, for example, *skulu* 'owe/have to' (also *munu* 'will') was not only the 3PL.PRES but also the INF form already in Old Norse:

	'own', infinitive: <i>eiga</i>				'owe', infinitive <i>skulu</i>			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	á	eigum	átta	áttum	skal	skulum	skylda	skyldum
2	átt	eiguð	áttir	áttuð	skalt	skuluð	skyldir	skylduð
3	á	<i>eigu</i>	átti	áttu	skal	<i>skulu</i>	skyldi	skyldu

Table 88: Indicative inflection of two preterite-present verbs in Old Norse (Zoëga 1910)

Other preterite-presents like *eiga* (see above) or *vita* ‘know’ did keep the “mismatch” between an infinitive in *-a* and a 3PL.PRES in *-u* somewhat longer and into Old Norse. However, this small group of non-conforming verbs has been slowly brought in line with the majority of verbs in the daughter languages (e.g. Icelandic has nowadays *eiga/eiga* and *vita/vita*, see Jörg 1989).

That it is the infinitive form that is extending into the 3PL present (and not merely the 3PL present suffix *-a* spreading from other verbs into the preterite-presents) is suggested by some of these analogical replacements like the one in the verb *mega* ‘must’ in Faroese (Lockwood 1977), whose earlier 3PL *mugu* is being replaced by *mega* and not by **muga*, which is all that a cross-paradigmatic analogy would probably afford.

It might be interesting to note, even if this is somewhat tangential to the present discussion, that the direction of influence, or in more theoretical terms the directionality of this “directional syncretism”, appears to have shifted in the history of the language. While early changes like INF **skula>skulu* suggest that the INF form is taken from the 3PL, later changes like Faroese 3PL *mugu>mega* suggests the opposite, i.e. that the 3PL form is taken from the infinitive. It might be speculative to venture an explanation here for this change of direction, but it would not surprise me if it had to do with the frequency of the two cells in different periods. If the relative frequency of preterite-present infinitives had been progressively increasing in the Scandinavian languages (maybe as these verbs became less idiosyncratic and more like others), the relative strength of the infinitive in these analogical changes may have increased enough over time to allow it to “take the upper hand” over the 3PL form.

4.1.1.1 The paradigmatic distribution of zero

The arbitrary nature of the linguistic sign, promulgated most famously by Saussure, has been one of the most celebrated axioms of linguistics. Although onomatopoeia, phonaesthemes and other phenomena are known not to conform to this arbitrariness, the core areas of grammar (e.g. the expression of morphosyntactic values in inflection by particular forms) are supposed to do so. Consequently, it could initially seem that cross-linguistic regularities

should not be expected in the domain of sound-change-generated morphemes in general. If every form-meaning association is equally possible (e.g. 2PL=/i/, 2PL=/pu/, 2PL=/ar/, 2PL=∅ etc.) one could well think that tendencies should not arise.

However, more abstract principles for form-meaning relations (like ‘constructional iconicity’, whereby more meaning should correspond to more form) have also been entertained in parallel for a long time. Thus, it was also found after Saussure that the relation of form to meaning is subject to a very important trend whereby an inverse correlation holds between frequency and length of expression. Thus, put simply, more frequent words and meanings tend to be shorter. This is known as Zipf’s Law. Although it is only exceptionless at the level of the whole language system, it still allows for probabilistic predictions for more concrete objects. Thus, Zipf’s Law allows us to predict that, in a randomly selected language, the word for great-grandfather will very probably be longer than the word for father.

These coding asymmetries are also relevant in the expression of grammatical information and categories (see Haspelmath to appear). Thus, 3 will tend to be shorter or unmarked compared to 2, and SG will tend to be shorter or unmarked compared to PL. This means that zero (as opposed to non-zero) will often adopt distributions that appear to make morphosyntactic sense:

	SG	PL		SG	PL		SG	PL		SG	PL		SG	PL
1	∅									∅			∅	
2	∅									∅				
3	∅			∅	∅		∅			∅	∅		∅	∅

Table 89: Some frequency-expected²⁹ distributions of zero

²⁹ See the frequencies provided in Table 55 for the approximate relative frequencies of the different person-number cells. In general: 3SG>3PL>1SG>1PL>2SG>2PL. That is, as is common knowledge in the discipline, ‘singular’ is the most frequent number value and ‘third’ is the most frequent person value. Because of this, inflectional patterns where SG, 3, 3SG, and SG+3 are zero/unmarked are expected from a Zipfian perspective. The fifth pattern in Table 89 (1SG+3) is also not unexpected, given that zero characterizes there the 3 most frequent person-number combinations.

The first three zeros above (i.e. SG, 3 and 3SG) adopt values which are usually considered to be well-formed and plausible for the meaning side of lexical entries. By contrast, the last two morphosyntactic distributions (SG+3PL, and 3+1SG) count as morphosyntactically unnatural.

The present research on morphemes has revealed that a paradigmatic configuration of zero vs non-zero is a relatively common source for overt exponents. These, maybe unlike zero, do need to be acquired in some way³⁰ and may fulfill the criteria for morphomhood that I have set in this dissertation (see Section 2.3). Thus, run-of-the-mill sound changes can and do transform zero-vs-affixed configurations into morphomic A-vs-B configurations. One such case (Jabuti in Table 82) was already presented. This one did conform to one of the paradigmatic distributions of zero that are supposed to be comparatively common due to Zipf’s Law. However, and because of the relative “flakiness” of zero, all sorts of morphomic patterns are attested to derive from zero vs affixed:

	Russian ‘name’			Wutung ‘be here’ (Marmion 2010:305)			English ‘do’	
	SG	PL		SG	PL		SG	PL
NOM	imja	imena	1	punga	nua	1	du:	du:
DAT	imeni	imenam	2	mua	punga	2	du:	du:
INSTR	imenem	imenami	3	mua	mua	3	dʌz	du:

Table 90: Partial paradigms showing zero-derived alternations in various languages

All the formal alternations above (i.e. /ja/ vs /en/ in Russian, /p/ vs /m/ in Wutung, and /u:/ vs /ʌ/ in English) go back ultimately to non-alternating paradigms where a single form appeared everywhere. The darkest shaded cells must have been at some stage characterized by zero, opposed to overt affixes in the other cells. In Russian, the paradigmatic locus of zero

³⁰ The absence of formatives can of course be significant within a paradigm in the sense that absences do participate in the system of morphological oppositions in a language. However, I believe it is unreasonable to expect absences to be carriers of morphological information on a par with overt affixes. Although morphologists often allow (or force) zero to participate in exponence rules in the same way as other affixes (e.g. blocking other overt formatives, see Pertsova 2011), zero cannot be expected to be subject to the same rules, morphosyntactic constraints and generalizations as other forms because it is not a form at all. Speakers, thus, do not need (and arguably cannot have) lexical entries for zero and do not have to learn the paradigmatic distribution of different absences in any unified, congruent way. This is the reason why morphemes in this dissertation have been defined over overt formatives, and never over zero or whole-word syncretism by itself.

“made sense” since it characterized the most frequent number-case cell. In Wutung (Sko, New Guinea), the paradigmatic distribution of zero seemed to be more arbitrary. In English, the distribution of zero could well be said to be completely unexpected from a Zipfian perspective (it classifies indeed as a typological *rarissimum*, see Plank & Filimonova 2000), as the marked form (i.e. 3SG) is actually the most frequent cell.

Be that as it may, in all three cases, the former zero-marked cells have acquired overt forms synchronically. Sometimes (e.g. in Wutung), the former zero-marked cells are the conservative ones and have preserved a (lexical) form lost elsewhere (mua < *m-pua). Some other times, (e.g. in Russian), it is the affixed forms that are conservative in that, in that position, the stem was “protected” from changes that affected the unmarked cells: imja < jimę (Proto-Slavic) < *inʔmen (Proto-Balto-Slavic) < *h₁nóm̃ (PIE) (Derksen 2007:212).

The present section has been dealing with the different ways in which zero-marked forms can become overtly marked *as a result of sound change*. However, forms that share simply a morphological zero can also be singled-out as formally identical by language-users, which may give rise, to some extent, to a ‘morphological niche’ (Aronoff 2016) for the introduction of new forms by means of analogical change (see Section 4.1.3.2) or grammaticalization (see Section 4.1.5 and Bantawa in 5.2.6).³¹ Although special reference to zero will not be made in those sections, it is something to be considered in those other diachronic sources as well.

Throughout this section, I have presented the different attested ways in which morphemes can emerge as a result of sound changes in different morphological or phonological configurations. This general origin has been found to be very common. The discussion in this section has also shown, however, that this general type is internally diverse. Morphemes, thus, can originate from sound changes in quite different ways. Most important among these are: A) the effect of the sound changes (i.e. sound changes can generate morphemes by a.1)

³¹ Bantawa and Athpariya show us how particular formatives can intrude into those (and only those) paradigm cells that are characterized by zero. Zero-marked cells, therefore, despite not sharing any overt morphology (and thus not meeting the definitional requirements for morphomhood that I have set in the present dissertation), can also sometimes provide a template for the distribution of incoming morphological elements. This suggests that they can have at least some morphomic properties under the right circumstances. It would be a matter for future research to find out to what extent the properties of zero-based morphological affinities resemble the ones of overtly-marked morphemes.

creating formal contrasts that did not exist before, or a.2) neutralizing formal contrasts that did exist before. B) The locus of the sound change (i.e. sound changes can generate morphemes by b.1) applying in the cells of the morpheme, or b.2) applying in the morpheme's complement cells. The following Table 91 summarizes this typology and provides examples of the different attested types:

Sound change effect		Sound change locus	
Formal merger	Formal divergence	Morpheme cells	Complement cells
Livonian (Table 86) Old Norse (Table 88)	Jabuti (Table 82) Saami (Table 220)	Chinantec (Table 160) Saami (Table 220)	Wutung (Table 90) French (Table 130)

Table 91: Types of sound-change-related morpheme origins

4.1.2 Semantic drift

Another, relatively well-known source of morphemes is the semantic drift and disintegration of formerly natural classes. This is the origin of the renowned PYTA morpheme of Romance (see e.g. Maiden 2001). The Latin verbal system was generally quite well behaved in the sense that, apart from the well-known 'third stem', most formal distinctions correlated quite straightforwardly to meaning differences. One of the most robust formal and semantic distinctions was the one around aspect. Observe the following Latin verb forms:

	IPFV	PFV
PAST.IND	faciēbat	fēcerat
PAST.SBJV	faceret	fēcisset
PRES.IND	facit	fēcit
PRES.SBJV	faciat	fēcerit
FUT.IND	faciet	fēcerit

Table 91: 3SG forms of 'make/do' in various Latin tenses (Maiden 2011a)

As the above Table 91 shows, one stem (*fac-*) appears in imperfective tenses and another one (*fēc-*) in the perfective ones. This is, therefore, a natural/morphemic alternation. As

Maiden (2011) explains, many of these tenses and their forms have been preserved in some of the modern Romance languages. The semantic and syntactic uses of the tenses, however, have been subject to various seemingly capricious changes. Consider, thus, the descendants in Spanish of the above tenses and their semantic content as reflected by their label below:

hacía IPFV.IND	hiciera IPFV.SBJV
none	hiciese IPFV.SBJV
hace PRES.IND	hizo PRET.IND
haga PRES.SBJV	hiciera FUT.SBJV
none	

Table 92: 3SG forms of ‘make/do’ of various Spanish tenses

The set of tenses that could be classified as perfective in Latin (shaded in Table 92 above) can no longer be assigned any common semantic or syntactic trait in contemporary Spanish. In terms of aspect, these tenses can be perfective or imperfective. In terms of tense, they can be past, present, or future. In terms of mood, they can be indicative or subjunctive. There is, thus, not a common thread of meaning or function extending across this set of tenses in modern Spanish in contradistinction to the others. However, maybe surprisingly, this arbitrary set of tenses have faithfully preserved their inherited morphological affinity. This makes morphological structures like this one morphomic.

For reasons related to feature-value orthogonality (to be presented in Section 5.1.2), morphemes like Spanish PYTA, i.e. so-called (Smith 2013) TAM morphemes where the morphological allegiances concern whole tenses, have not been included in the morpheme database of Chapter 5. This makes it difficult to assess the relative prevalence of semantic drift in the creation of morphemes cross-linguistically. My overall impression is that semantic drift might be comparatively rare as the force responsible for single-handedly creating morphomic structures. Although it is not at all uncommon for formerly semantically motivated forms to become opaque,³² this may happen in different ways (see Section 4.1.3).

³² For example a realis vs irrealis distinction becoming morphomic in Sye (Crowley 1998), a past vs non-past distinction becoming morphomic in Northern Talysh (Kaye 2013), etc.

4.1.3 Analogy

Analogy is a term used so widely in linguistics and which is used to mean so many different things that it is impossible to explain it at any length within the confines of a small section (for more specialized treatments see e.g. Blevins & Blevins 2009, and Gaeta 2010). The term will be used here as a cover term for all morphological and paradigmatic changes driven by language users' failure to acquire and replicate accurately some aspect of a language's grammatical system. I take these changes to be copying errors that take place predominantly in low frequency inflectional areas precisely because they are chiefly due to insufficient input. Analogy, thus, happens when language users, based on the input available to them, induce a grammatical system that differs slightly from the one of their elders. Analogy is usually taken to be a simplifying force in language: infrequent forms, categories or distinctions are lost, lexical idiosyncrasies give way to general rules etc. In the context of the present discussion, I will distinguish two types of analogical processes that may result in morphomic structures: morphosyntactically-motivated and formally-motivated analogies.

4.1.3.1 Morphosyntactically motivated analogy

I define morphosyntactically motivated analogy here as the change, usually in an infrequent cell or set of cells in the paradigm, whereby the original form is replaced by another one borrowed from a neighbouring cell (i.e. from a cell with which it has a particularly close morphosyntactic relationship in the form of shared content). It may appear intuitively contradictory for morphosyntactically-driven analogies to be able to result in morphomic patterns. However, they may do so when they happen in some of the values of a natural class and not in others. Consider the following partial paradigms:

Slovene <i>človek</i> 'man' (Baerman et al. 2005:175, Herrity 2000:49)			Gévaudan Occitan <i>cantar</i> 'sing' (Ronjat 1930, Camproux 1958)		
	DU	PL		PRETERITE	IMP.SUBJ
NOM	človeka	ljudé	1SG	cantère	cantèssie
ACC	človeka	ljudí	2SG	cantères	cantèssies
DAT	človekoma	ljudém	3SG	cantèt	cantèssie
INS	človekoma	ljudmí	1PL	cantession	
GEN	ljudí		2PL	cantessiat	
LOC	ljudéh		3PL	cantèrou	cantèssou

Table 93: Natural syncretisms resulting in unnatural morphological patterns

In the paradigms in Table 93, some comparatively infrequent cells (GEN.DU and LOC.DU in Slovene, and 1PL.PRET and 2PL.PRET in Occitan) have changed their etymologically expected forms, which have been replaced by morphosyntactically related ones from other close values. This is exactly how normal morphosyntactically driven analogy works. Tense or number values may be lost everywhere at the same time, but sometimes they can also start to break down at their weakest links first. The analogical changes in Table 93 (see also Biak in Section 5.2.10), should probably be understood as manifesting the loss of number and tense distinctions in some (infrequent) contexts. The particularity by which this process results in the morphologically unnatural distribution of some forms here (the stem *ljud-* in Slovene and the formative *-ss* in Occitan) is that the extended forms are formally marked as belonging to a broader (natural) set of forms.

Morphemes may originate by morphosyntactic analogy both from morphemic (i.e. natural class distributed) formal elements, like in the cases that have been presented above, but also from morphomic (i.e. unnatural class distributed) forms. Morphosyntactic analogical processes, thus, can modify the paradigmatic extension of morphomic structures without bringing forms back to the realm of morphemes. Consider the following change:

	Pre-Wambisa		Wambisa	
	SG	PL	SG	PL
1	muuka-ru	muukĩ	muuka-ru	muukĩ
2	muuki-mi	*muuki-mi	muuki-mi	muukĩ
3	muukĩ	muukĩ	muukĩ	muukĩ

Table 94: Possessive inflection of *muuka* ‘head’ in Wambisa (Peña 2016:467)

The tendency to level plural forms is morphosyntactically understandable and documented in various different languages.³³ The formal levelling within the natural class ‘plural’, however, does not result in a natural morphological pattern in Wambisa because of the preexisting syncretism of 3SG and 3PL. The present study suggests that developments of this kind are not uncommon. Although it might be difficult to go beyond impressionistic claims in this respect, it looks like analogical changes operating on morphomic structures seem usually oblivious to their morphomic status and not particularly aimed at bringing the forms in line with a natural class. See the morphemes of Nen (Table 208) and Servigliano Italian (Table 106) for other morphomic structures that have been shaped by morphosyntactically driven analogy but have stayed morphomic.

4.1.3.2 Analogy motivated by form

Whereas the previous analogical processes capitalized on the semantic and/or the morphosyntactic proximity of the source and target values (e.g. GEN.PL>GEN.DU, 3PL>2PL), the analogical changes that will be presented here have a very different *raison d’être*. In this case, the motivation for the change has to be found in the morphological similarity of the source and target forms. Although this has not received as much attention as it should, it is well-known (see e.g. Burzio’s 2001 ‘gradient attraction’) that formal similarity may result in

³³ This tendency seems to be particularly strong when the 3PL becomes syncretic with one of the other two plural cells, like in Dutch (where 1PL and 3PL came to be characterized by the suffix *-en*, which later spread to the 2PL) or Old English (where 2PL and 3PL came to be marked with the *-aþ*, which later spread to the 1PL).

still more similarity. Thus, two forms whose only common property is that they are morphologically similar may become more systematically similar or identical even in the absence of shared content. Consider the following case (also dealt with in Table 24):

	'go'	'be'	'say'	'come'	'sing'
Participle	ido	sido	dicho	venido	cantado
3SG Future	irá	será	dirá	vendrá	cantará
Infinitive	ir	ser	decir	venir	cantar
2PL Imperative	id	sed	decid	venid	cantad
2PL Present	vais	sois	decís	venís	cantáis
2SG Imperative	ve	se	di	ven	canta
3PL Present	van	son	dicen	vienen	cantan
1SG Past	fui	fui	dije	vine	canté

Table 95: A selection of word forms in different Spanish verbs

There is a very widespread analogical change in non-standard Spanish whereby the etymologically expected form for the 2PL imperative (e.g. *venid* < *venīte*) is replaced by the infinitive form (e.g. *venir* < *venīre*). Thus, in many varieties and idiolects and despite linguistic prescription, the form *ir* replaces *id*, *ser* replaces *sed*, *decir* replaces *decid*, and so on. This analogical change, and the resulting unnatural whole-word syncretism it produces, is motivated by the preexisting morphological affinity between the two paradigm cells.

Infinitive and 2PL imperative (and no other cell beyond these two) share their stress, theme vowel, and stem-related properties *in every single lexical item*. As a result, there is perfect formal predictability between these two cells because they always differ only in their last consonant, which is *-r* in the infinitive and *-d* in the 2PL imperative forms. Thus, the preexisting formal similarity of these two word forms has provided the motivation for the analogical change described here and for the whole-word unnatural syncretism that it established. Systematic stem identity has thus resulted in affixal identity.

Changes like these, where formal affinity in the stem provides the motivation for the formal identity of affixes, seem not to be infrequent. See the diachronic insights on the morpheme of Girawa (Section 5.2.22) for other morphomic structures with similar origins. The locus of the formal similarity that provides a motivation for formally-driven analogy, and direction of the formal influence, however, can also be the opposite. Thus, the formal similarity or identity of affixes can provide a motivation for the extension of this formal affinity to the stem. Observe the following analogical developments also in Spanish:

	‘die’	‘put’	‘make’	‘come’	‘sing’
Participle	muerto	puesto	hecho	venido	cantado
3SG Future	morirá	pondrá	hará	vendrá	cantará
Infinitive	morir	poner	hacer	venir	cantar
2PL Imperative	morid	poned	haced	venid	cantad
2SG Imperative	muere	pon	haz	ven	canta
3PL Present	mueren	ponen	hacen	vienen	cantan
3PL Past	murieron	pusieron	hicieron	vinieron	cantaron
Gerund	muriendo	poniendo	haciendo	viniendo	cantando

Table 96: A selection of word forms in different Spanish verbs

In some non-standard varieties of Spanish, the stem of the gerund is replaced by the stem used in the so-called PYTA (the 3PL past is provided in Table 96 as a representative of these cells). Thus, *poniendo* changes to *pusiendo* analogically, and *haciendo* changes to *hiciendo* (Pato & O’Neill 2013). The motivation for this change has to be found in the suffixal similarity of the gerund and many of the PYTA cells. Both are namely characterized by a tonic suffix /je/ directly after the root. The association of the PYTA root and /je/ is also seen clearly in the fact that PYTA roots always co-occur with this formative, even in otherwise first conjugation verbs (e.g. *est-a-r* vs *estuv-ie-ron*, *d-a-r* vs *d-ie-ron*, compare to regular *cant-a-r* vs *cant-a-ron*). Thus, the tonic suffix /je/ selects always the PYTA root except in the gerund forms of some verbs like ‘put’ and ‘make’. By extending the former perfective root to the gerund, thus, these analogical changes remove this exception. Note, however, that in doing so, a systematic

morphological identity has been created between cells that have no particular morphosyntactic affinity.

Morphemes, thus, can and do emerge from more-or-less accidental formal similarities between morphosyntactically unrelated paradigm cells or sets of cells. In the history of Persian, for example, we find another analogical change in which an affixal formal similarity provided the motivation for an analogical change that established systematic stem identity between morphosyntactically unrelated cells.

As explained by Kaye (2013:118), older Iranian languages had a morphosyntactically natural system of verb stem alternation whereby past tenses and past participles shared form in opposition to non-past forms of the verb. The past tense forms were characterized by a dental extension/suffix to the stem. This is so because synthetic past tenses had grammaticalized from periphrases originally involving the PIE participle in *-ta*.

Parallel to this we have the form of the infinitive suffix, which in Old Persian, for example, was *-tanaiy*. This form was unrelated to the past tense morphology just described and the stems of one and the other were sometimes different (e.g. *kṛta-/čartanaiy* 'die'). However, the accidental formal resemblance of the infinitive and the past tense forms provided the motivation for the systematic analogical extension of past morphology to the infinitive (e.g. *čartanaiy* > *kerdan* in Middle Persian). Thus, in the daughter languages, infinitives and past tenses pattern together and constitute a morphomic class for the purposes of exponence (e.g. Middle Persian *pursīd* 'asked' vs *pursīdan* 'to ask', Parthian *pursād* vs *pursādan*). This morphomic affinity has been preserved also into modern descendants like Persian (see Bonami & Samvelian 2009:28) and Balochi (Axenov 2006). The formal alternations of PAST/INF and other wordforms have also become quite diverse in synchrony (e.g. in Balochi *and-/andit-* 'laugh', *kap-/kapt-* 'fall', *ill-/išt-* 'put', *band-/bast-* 'close', *kan-/kurt-* 'do', *ra-/šut-* 'go' etc. (see Axenov 2006), so that the non-accidental (i.e. morphomic) nature of the affinity is beyond any doubt.

Previous examples have all involved formally-driven analogy of some forms on the basis of other forms within the same paradigm. Although it might be less frequent, there is of course nothing that makes inter-paradigmatic formal analogies impossible. Consider, in this respect, the case of Acazolco Otomí (discussed by Baerman et al. 2017:13), which analogically changed the form of a suffix to match that of what was originally a different suffix (i.e. one with a different distribution and occurring in a different inflectional class). Because of the present dissertation’s executive decision to restrict the analysis to intra-paradigmatic formal affinities (see Section 2.4), cases like this will not be further discussed here.

4.1.3.3 A note on the motivation of analogy

Although the analogical changes in the previous two sections have been neatly classified as either form-driven or morphosyntactically-driven, many analogical changes involve both forces to some extent. Consider, for example, the following syncretism:

	Expected		Attested	
	SG	PL	SG	PL
NOM	-a	-an	-a	-an
ACC	-on	-on	-an	-an
GEN	-an	-ena	-an	-ena
DAT	-an	-um	-an	-um

Table 97: Weak masculine declension endings in Old English (Bazell 1960:3)

It seems clear that both the formal similarity between the source and target form (i.e. *-on* vs *-an*) and the morphosyntactic affinity between the cells must both have been factors that facilitated or motivated the analogical change. Thus, classification into the two types of analogy that have been identified here is not to be understood as mutually exclusive.

4.1.4 Pattern interactions³⁴

Another way in which morphemes can emerge in a language is by means of the conflict or interaction between different patterns of allomorphy distribution. These patterns can be morphomic or morphemic. For straightforward predictability relations to hold between pairs of cells in a paradigm, it is helpful for forms to be distributed in the same way across lexical items. This could be thought of as the *raison d'être* of morphomic patterns. When two different patterns cross-cut each other in the paradigm, however, this predictability is jeopardized. This leads sometimes to analogical developments by which existing forms change their original paradigmatic configurations or by which new incoming forms intrude into the paradigm by adopting a distribution that is new in the language. Consider the case of Romance L- and N-morphemes:

	Spanish 'understand'		Spanish 'put'		Ansootano Aragonese 'come' (Barcos 2007)	
	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	entiendo	entienda	pongo	ponga	bjengo	bjengaj
2SG	entiendes	entiendas	pones	pongas	bjen(e)s	bjengas
3SG	entiende	entienda	pone	ponga	bjene	bjenga
1PL	entendemos	entendamos	ponemos	pongamos	benimos	bengamos
2PL	entendéis	entendáis	ponéis	pongáis	beniθ	bengaθ
3PL	entienden	entiendan	ponen	pongan	bjenen	bjengan

Table 98: N- and L-morphemes and their paradigmatic distribution

Because of the cross-cutting distributions of the N- and L-patterns in the paradigm, they give rise to four different areas in the paradigm (see the paradigm of 'come' in Table 98) depending on which (or whether any) of the two patterns applies in a given cell. These four sets of cells are the ones where stems will be always internally identical but may be externally different. They do have, therefore, some morpheme-like properties in that they afford formal predictions and may, because of this, provide a niche or template for other (incoming) forms. Consider the paradigms below:

³⁴ This section draws heavily on the data and arguments in Herce 2019.

	Lags Romansh 'let, cause' (Maiden 2018b:108)		Bolognese 'go' (Maiden 2012)		Felechosa Asturian 'bring' (Maiden 2012)	
	IND	SUB	Indicative	Subjunctive	IND	SUB
1SG	lafel	lafj	va:g	va:ga	trao	traa
2SG	lais	lafjes	vε	va:g	traes	traas
3SG	lai	lafj	va	va:ga	trae	traa
1PL	fεin	fεjen	andain	andannja	traemos	trifamos
2PL	fεis	fεjes	andε	andedi	traes	trifaes
3PL	lain	lafjen	van	va:gen	traen	traan

Table 99: Some morphological patterns arising from morpheme interactions

In the Lags Romansh paradigm above, there is a stem alternant *lai-* which lacks the stem-final consonant /j/ and has /i/ instead. This form is believed (see Maiden 2018b:108) to have originated in the SG imperative and to have spread subsequently to these other cells. The SG imperative, 2/3SG indicative and 3PL indicative constitute the set of cells that belong to the N-morpheme (i.e. were rhizotonic in Latin) but not to the L-morpheme. Thus, it is the smallest morphomic niche to which forms originating in the SG imperative can spread.

The Bolognese paradigm above shows how the form /g/ characteristic of the L-morpheme does not appear in the 1PL and 2PL subjunctive where it would be expected. It is relatively common for L-morpheme roots to be expelled from these cells, thus becoming confined to the set of cells that belong to L and N simultaneously. In Felechosa Asturian, in turn, we find a special root (taken from PYTA) being introduced in 1PL and 2PL subjunctive. These are the cells that participate in the L- but not in the N-morpheme.

The analogical processes described above illustrate how the different swaths of the paradigm that originate from cross-cutting formal elements (see Table 98) may become morphomic in their own right by providing a perfect predictability island in the paradigm within which stem identity can be taken for granted. The paradigm areas where either or both morphemes apply can also be singled out, however, as the domain of allomorphy:

	Verb 'have' in Old French		Verb 'measure' in Spanish		Verb 'have to' in Savognin (Maiden 2018b:213)	
	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	tieng	tiegne	mido	mida	stó	stóptga
2SG	tiens	tiegues	mides	midas	stóst	stóptgas
3SG	tient	tiegne	mide	mida	stó	stóptga
1PL	tenons	tiegniens	medimos	midamos	duágn	stóptgan
2PL	tenez	tiegniez	medis	midais	duéz	stóptgas
3PL	tient	tiegnent	miden	midan	stón	stóptgan

Table 100: Another morphomic pattern arising from morpheme interactions

In the above paradigms, formal elements have spread to all the cells that participate in the N- and/or in the L-morphome. Consider, thus, the Old French verb 'have' in Table 100. Regular sound change would have resulted in diphthongization (i.e. /e/>/je/) in the N-morphome cells and palatalization (i.e. /n/>/ɲ/) in the L-morphome cells. These two forms should have, therefore, cross-cut each other like the formatives in Ansotano Aragonese in Table 98. However, the diphthong has spread analogically into 1PL and 2PL subjunctive and has thus come to characterize all the cells where N and/or L apply. The diphthong did not spread beyond this set of cells which acted, therefore, as a niche for that particular form.

In Old French, a form characteristic of the N-morphome was generalized to this particular superset of cells. Something else happened in Spanish *medir* 'measure'. Rising (i.e. /e/>/i/) is the result, in Ibero-Romance, of anticipatory assimilation of mid vowels to a following yod (i.e. *metjo>mido, *metimus>medimos). This yod is precisely what created some of the formal alternations known as the L-morphome. Rising, thus, would have occurred, initially, in just those cells. In Spanish, however, like in Old French before, a single vowel has been generalized to the same N+L superset. In this case, however, it is the vowel that originally characterized the L-morphome.

The last example of how this set of cells can act as a morphological class in Romance is the paradigm of Savognin *duéir*. As Maiden (2018b:213) explains, these N- and/or L-morphome cells is the paradigmatic domain where suppletion occurs in this verb. Stem allomorphy is

present in these cells in the paradigms of other lexemes as well and this fact provides a niche or template for the distribution of other formal elements in the paradigm.

4.1.5 Grammaticalization

Because of the prevalent theoretical stance in the literature that morphemes should be typologically unique, and arise in typologically unique ways (see Section 2.6), grammaticalization processes have not been usually mentioned as a possible source for morphemes. This is so, of course, because the phenomenon of grammaticalization is characterized precisely by its cross-linguistic generality and unidirectionality. If one, as I do throughout this dissertation, remains open to the possibility of there being cross-linguistically recurrent morphemes and cross-linguistically recurrent pathways of morpheme emergence, then one finds that run-of-the-mill grammaticalization processes can and often do result in synchronically unmotivated morphological affinities.

Although usually this is not explicitly discussed, not all linguists have subscribed to the idea that morphemes must be typologically unique by definition. Thus, Stump (2015:134) discusses the case of a morphological affinity in Noon (Atlantic-Congo) which he presents as a textbook example of a morphomic structure. This morpheme involves the use of the same morphology for the expression of the passive voice and of 3PL subject agreement. From a diachronic perspective, this affinity is unsurprising. It is well-known (e.g. Heine & Kuteva 2002:236, Siewierska 2010) that 3PL is often a source for passive morphology, frequently via other intermediate functions like impersonal. Furthermore, the same morphological quirk is found in other unrelated languages like, for example, in Kven (Uralic) (Söderholm 2017).

As other linguists before me (e.g. Lichtenberk 1991), I believe that, even if/when various functions or meanings are historically related (by means of a grammaticalization channel), there need not be any synchronic property shared exclusively by these different uses. This would leave the end-product of many of these grammaticalization paths purely morphomic. Similarly to this particular affinity of 3PL.SUBJ=passive, one could offer other cross-

linguistically recurrent patterns of change like instrumental>ergative³⁵ (Palancar 2001), or 1SG.OBJ>antipassive (Bickel & Gaenszle 2015). The mere fact that these (e.g. ergative and instrumental) are most usually described as different cases/functions with homonymous exponents, rather than as a single case/macrorfunction with various uses, suggests that this same intuition is widely shared.

Morphological vestiges of grammaticalization processes can be relatively common, like the ones mentioned above, or more idiosyncratic. In Lango (Nilotic), for example, as described by Noonan (2011:91-92) the verbal system contains three aspects (perfective, habitual and progressive). Consider the partial paradigm below:

	Perfective	Habitual	Progressive
1SG	àgíkò	àgíkô	ágìkkò
2SG	ìgíkò	ìgíkô	ígìkkò
3SG	ògíkò	ògíkô	àgìkkò

Table 101: Partial paradigm of Lango ‘stop sth’, infinitive: gìkkò (Noonan 2011:92)

As Table 101 illustrates, there is a special morphological relationship between the infinitive and the progressive aspect forms in the language. In a way similar to the affinity between the infinitive and the past tenses that we described for Balochi and other Iranian languages in Section 4.1.3.2, the morphological affinity of the infinitive and the progressive is not derived in Lango from any aspect of these forms’ semantic or syntactic behaviour. It is simply a morphomic trait in the paradigmatic organization of the language.

As explained by Noonan (2011:91), the presence of this trait in the language is due to the fact that the progressive originated in Lango, as in other languages, from a periphrastic construction. This involved the verb *ya* ‘be in a place’ plus the infinitive (observe the similarity to constructions in other languages like non-standard German ‘ich bin am Arbeiten’). The

³⁵ In the Australian language Wambaya (Nordlinger 1998:83-84), for example, the ergative and instrumental functions are marked in the same way, with 4 allomorphs each (-ni, -nu, -ji, yi) distributed in identical phonological and morphological environments.

conventionalization of that construction in Lango to express the progressive meaning and the later univerbation of the construction into a single word are straight-forward grammaticalization-related developments which, however, have left their mark in the synchronic paradigmatic organization of the language in the form of a morphological identity of infinitive and progressive. Notice, however, how similar grammaticalization processes have resulted in very different morphological affinities in other languages (e.g. of infinitive and future/conditional in Romance), which proves that these affinities are no less arbitrary and morphomic, than those between infinitive and past tense arising from formally-driven analogy in Iranian (see Section 4.1.3.2), or between the infinitive and 3PL arising from sound-change in Scandinavian (see Section 4.1.1).

Because of the way syntax behaves most usually, the morphology that emerges from the accretion of formerly separate words tends to be relatively well-behaved in that it usually characterizes natural classes or a whole tense or set of (related) tenses. Because PYTA-style TAM morphemes have not been included in the database of Chapter 5, morphemes emerging from grammaticalization may be underrepresented here. It is, however, definitely not the case that syntax is always only sensitive to natural classes (see Section 2.9.2) or that univerbation processes can only ever occur in natural classes. Consider Athpariya:

	SG	DU	PL
1EXCL	khat-naʔa	khat-ciciŋa	khad-itiŋa
1INCL		khat-cici	khad-iti
2	a-khat-yuk	a-khat-cici	a-khad-iti
3	khat-yuk	khat-cici	u-khat-yuk

Table 102: Athpariya ‘go’, intransitive positive non-past (Ebert 1997:163)

As Schackow (2016:230-231) explains, Athpariya *-yuk* goes back ultimately to a lexical verb *yurŋ*, which meant ‘be’ or ‘stay’. This verb, thus, and others in other Kiranti languages, must have grammaticalized into the so-called ‘tense markers’ we find synchronically in the 2/3SG and in the 3PL in Athpariya. The fact that the univerbation happened in these cells only must be related to the fact that those were the cells that lacked suffixes (or had a ‘zero suffix’ if you will) originally. See e.g. Bantawa (Doornenbal 2009:391), or Puma (Sharma 2014:424).

4.1.6 Borrowing

The borrowing of morphological forms or patterns between languages is a common force in language change. Because of their very particular characteristics, however, morphemes (at least of the kind analyzed here) seem to find themselves almost always in the “worst end” of the borrowability scale. In the analysis of which factors favour or hamper borrowability, the literature on language contact (e.g. Matras 2014, Seifart 2015, Kossmann 2015 etc.) comes to the following conclusions regarding the relative ease with which morphology is borrowed: lexical>grammatical, derivational>inflectional, segmentable>unsegmentable, simple-meaning>complex-meaning. Because of the properties of morphemes as defined here (i.e. they are grammatical, inflectional, complex-meaning structures), they would constitute morphological entities that are not usually borrowed.

There seems to be also an emergent consensus (Carlin 2006, Kossmann 2015) that the borrowing of morphology is particularly common when (bilingual) language users feel the need for a particular morphological distinction present in one of their languages but absent from another. As mentioned by Kossmann (2015) “this stands to reason: there is no clear functional explanation for the transfer of an isolated morpheme to express something that is already expressed. However, the bilingual speaker confronted with different categorizations in the two languages (s)he uses, may wish to express the same categories in the two languages.” Because of this, language users of Slovene Romani borrowed a 2PL suffix from South Slavonic to reintroduce the 2SG/2PL distinction that had disappeared from their language (see Kossmann 2015). Similarly, Mawayana (Arawakan) speakers borrowed a 1PL exclusive pronoun from Waiwai (Cariban) to be able to convey clusivity distinctions (see Carlin 2006). These functional motivations for borrowing seem to be impossible in the case of morphemes which are, by definition, ill-suited for the transmission of meaning.

Probably because of the aforementioned reasons, I have not been able to find any incontrovertible examples of morpheme borrowing. There are, however, cases that come very close indeed, both with respect to matter and pattern borrowing. With respect to the former, for example, Maiden (2018b:101) mentions the case of a Sardinian variety (Campidanese) where one can find classically L-morphomic patterns like the following:

	Indicative	Subjunctive
1SG	'tengu / 'tɛnju	'tɛnga / 'tɛnja
2SG	'tɛnis	'tɛngas / 'tɛnjas
3SG	'tɛnit	'tɛngat / 'tɛnjat
1PL	tɛ'nɛus	tɛn'gaus / tɛn'jaus
2PL	tɛ'nɛis	tɛn'gais / tɛn'jais
3PL	'tɛnint	'tɛngant / 'tɛnjant

Table 103: Present tense paradigm of Campidanese Sardinian *tɛnni* 'have' (Lepori 2001)

As Table 103 illustrates, we find that alongside the regularly expected forms like 'tɛnju, forms with the characteristically L-morphomic velar augment (i.e. 'tɛngu) are also attested. This /g/ is not etymological in this verb, although it is also not etymological in the paradigms of 'have' in many other Romance languages like Spanish *tener/tengo*. In most of these cases it is assumed that the presence of /g/ here is due to the analogical influence from other verbs (e.g. *decir/digo*) which would indeed have had the form as a result of regular sound change. What is remarkable about the presence of this form in Sardinia is, however, that, unlike in every other Romance language, velars were not subject to the palatalizations that generated *decir/digo*-type alternations elsewhere in Romance. The formative /g/ as an exponent of the L-morphome is, thus, completely foreign to Sardinian and must have been necessarily borrowed from another Romance language like Italian or Catalan.

This is undoubtedly a very interesting morphological development. However, it probably falls short of the 'borrowed morphome' we are after in this section. This is so because stem alternations with this same L-pattern configuration in the paradigm do occur in the language natively with other forms and verbs. Although velar consonants /k/ and /g/ (also /n/ for that matter) were not subject to palatalization in Sardinian, /t/ and /d/ were, yielding /ts/ and /dz/ respectively. These forms are the regularly expected L-morphome exponents in the island and have actually spread analogically (e.g. also to the verb 'have' in other varieties, e.g. 'tɛndʒo/'tɛnɛs in Nuorese, see Pittau 1972). Only the formative /g/, and not the L-morphome as such, thus, can be said to have been borrowed into Campidanese Sardinian.

A case where a morphological element has been borrowed into another language along with its arbitrary distribution in the donor language may be found in Resígaro (Arawakan). There is a classifier suffix *-ba* in Bora (Boran, Brazil) which is used mainly for fruits, logs, and drinks. This formative has been borrowed into Resígaro along with its seemingly arbitrary semantic extension in the lexicon (see Seifart 2015:519). Although this could be seen as a case of simultaneous matter-*cum*-pattern borrowing of a morphomic element, it is clear that we are dealing here with a lexical, and not a paradigmatic pattern like the ones that this dissertation deals with primarily.

Concerning pattern-only borrowing of morphological categories, there are very striking cases that one could offer of whole inflectional systems being restructured to match the categorial distinctions of another language. One of the clearest cases is found in Tariana:

	Baniwa (Arawak)	Tariana (Arawak)	Tucano (Tucanoan)
Non-topical non-subject	-	-	-
Topical subject	-	-naku -nuku	-re
Allative	-ʒiku	-se	-pɨ
Superessive	-naku	-se	-pɨ
Orientative	-hre	-se	-pɨ
Ablative	-(hi)ʔe	-se	-pɨ
Perlative	-wa	-	-

Table 104: Morphological realization of some semantic functions in three Amazonian languages (adapted from Aikhenvald 2002:102-104)

As explained by Aikhenvald (2002:102-104), the typically Arawakan system (see Baniwa) for indicating different spatial relations has been replaced in Tariana by a typically Tucanoan system. No forms were borrowed, however, but only the patterns. One of the former spatial suffixes became a marker for topic while another one was extended to cover the functions of the general spatial marker common in Tucanoan languages. The grouping of some (allative ‘to’, superessive ‘on’, orientative ‘towards’, and ablative ‘from’) but not all (consider the perlative ‘through’) spatial relations under a single morphological realization could well be considered semantically arbitrary to some extent.

Another example of a largely arbitrary morphological pattern that straddles linguistic borders can be found in the languages of Vanikoro. As explained by François (2009), the three languages spoken in this island show quite substantial lexical differences (suggesting a relatively deep genealogical separation) but share a very idiosyncratic pattern of syncretism:

	Teanu (Oceanic)			Lovono (Oceanic)			Tanema (Oceanic)		
	SG	DU	PL	SG	DU	PL	SG	DU	PL
1EX	ne-	ba(i)-	pe-	ka-	ba(i)-	pe-	na-	ba(i)-	tu-
1INC	-	la(i)-	le-	-	sa-	kape-	-	ja-	la-
2	u-	ba(i)-	pe-	ku-	ba(i)-	pe-	go-	ba(i)-	tu-
3	i-	la(i)-	le-	ki-	sa-	se(pe)-	i-	ja-	la-

Table 105: Irrealis subject agreement prefixes in the languages of Vanikoro (François 2009)

As the paradigms in Table 105 illustrate, the morphological syncretisms within the plural complex in the languages of Vanikoro are, basically, at odds with any plausible semantic or morphosyntactic feature value or constellation of values. Syncretisms like 1EX/1INC vs 2/3 would be straightforward. Even syncretisms like 1INC/2 vs 1EX/3 could be derivable as the expression of +2 (i.e. addressee) vs a default. This is, therefore, the only two-way syncretism of the plural complex that appears to make no sense whatsoever in extramorphological terms. This quirk, however, is shared across the three languages, often with quite different formatives (e.g. *la(i)-* vs *sa-* vs *ja-*, or *pe-* vs *tu-*). The languages are, indeed, genetically related, so this pattern of syncretism might be derived from a common ancestor, which is the reason why this also may fall short of the ‘borrowed morpheme’ I have been looking for in this section. However, the continued presence of this typologically unique morphomic pattern in the three languages even in the presence of significant morphological changes must surely be, at least partially, due to language contact.

4.1.7 Mixed origins

The previous sections have presented evidence of how morphemes can arise in a language in quite a few different ways: due to i) sound changes (4.1.1), ii) semantic drift (4.1.2), iii) morphosyntactic or form-driven analogy (4.1.3), iv) morphological interactions (4.1.4), v) grammaticalization (4.1.5), and maybe even through vi) language contact (4.1.6). I have therefore attempted to present so far clear examples of morphemes that have emerged due to the single-handed action of only one of these forces. The history of many (maybe of most) morphemes, however, is instead a combination of several of the above-mentioned diachronic processes either simultaneously or, usually, at different stages. Consider, for example, the following cases:

	<i>pote</i> 'can'		<i>di</i> 'say'		<i>ae</i> 'have'	
	IND	SUBJ	IND	SUBJ	IND	SUBJ
1SG	pottso	pottso	diko	diko	aggjo	aggjo
2SG	poi	poi	ditʃi	ditʃi	ai	ai
3SG	pɔ	pottsa	ditʃe	dika	a	aggja
1PL	putimo	putimo	ditʃimo	ditʃimo	aimo	aimo
2PL	potete	potete	ditʃete	ditʃete	aete	aete
3PL	pɔ	pottsa	ditʃe	dika	a	aggja

Table 106: Present tense paradigms of three Servigliano Italian verbs (Camilli 1929)

As other Romance varieties, the palatalization of various consonants before front vowels led in Servigliano to stem alternations in the verbal paradigm (i.e. diko/diki > diko/ditʃi). Because of the phonological profile of Latin suffixes, the alternations singled out the 1SG indicative and the subjunctive forms of the present as those with a different stem from the one found elsewhere. This is the original distribution of the L-morpheme as created by sound change. In Servigliano, however, and in other Italian varieties, morphosyntactically-driven analogical processes involving the loss of mood distinctions in the first and second person have modified the original paradigmatic distribution of the inherited alternations. The morpheme's current paradigmatic extent is thus the result of both i) sound change, and iii) morphosyntactically-driven analogy.

There is, obviously, a large number of different combinations of forces that may result in a particular morphomic pattern synchronically. Many other examples could, thus, be offered of morphemes having a complex diachronic origin. As for the ones in the present database (Chapter 5), the morphemes of Aragonese (Section 5.2.3) and Palantla Chinantec (5.2.14), for example, must have involved both first i) sound change and later iv) pattern interactions. As for morphomic structures discussed elsewhere, the Northern Talysh verbal morphemes discussed by Kaye (2013), for example, involved both iii) formally-driven analogy, as well as the subsequent v) grammaticalization and univerbation of verbal periphrases involving the infinitive. Given the amount of cases that I have assembled, it seems that complex diachronic origins may well be the rule rather than the exception in morpheme emergence.

4.2 The loss of morphomic structures

Earlier sections have dealt with the various ways in which morphemes may arise in a language. Even though these structures are usually taken to be quite stable in the literature,³⁶ it is obvious that, just as any other grammatical trait, morphemes can also disappear from a language. This section will present the different ways in which this may happen.

4.2.1 Loss of productivity and gradual erosion

As soon as a class or category ceases to be productive and incorporate new members regularly, it can be said to be, in some sense, already on its way out from a language. In the absence of new recruits, and provided sufficient time goes by, any class would eventually vanish due to the relentless trickle of “desertions” that it would undoubtedly suffer. Note,

³⁶ The validity of these claims is not clear to me at this point. If we found out that the average life-expectancy of a morpheme is 2000 years, for example, or 3174, it would still not be obvious at all whether that is “a lot of time” or not. Stability is a relative concept, so two millennia are a long time in human timescales but not at all in geological terms. Language evolution is likely to fall between these two. Thus, whether morphemes are relatively stable or not should be answered, I believe, by comparing them to a number of other linguistic traits or forms: the durability of different morphemes, the rate of lexical replacement of different lexical items, or the duration of other grammatical traits like ergative alignment, pro-drop, clusivity etc.

however, that categories can remain largely unproductive for extremely long periods of time before they eventually disintegrate.³⁷ During this time, of course, they may remain part of the grammar and subject to their own rules and organizational principles, which means that they cannot be dismissed lightly as uninteresting or “irregular” phenomena.

Many of the most heavily studied morphemes (Romance PYTA, and L- and N-morphemes) can be described as being at this stage to some extent. They are, thus, largely unproductive but nevertheless “living” morphological categories in many Romance languages (e.g. in Spanish). Some of these morphemes (L- and N-) have probably never been truly productive morphological categories (in the sense that new lexemes did not display them by default). They were probably always losing members, therefore, ever since they first appeared in the language as morphological entities. Some other morphemes like PYTA, by contrast, were completely productive morphological categories at some point. Formal distinctions were regularly made in Latin (e.g. adding a suffix /w/) to mark the perfective tenses. Losing morphological productivity (i.e. a category’s applicability to incoming words) can thus be understood as the first step towards the disappearance of a morpheme.

Because of the long periods of time that unproductive categories can continue to exist in a language (see Footnote 37), it is difficult to find an example of a morpheme that has disappeared exclusively due to the constant eroding effect that lack of productivity brings about.³⁸ It is, however, always a prerequisite, and often a force that is actively involved.

³⁷ Consider, for example, the Germanic strong verbs. The proportion of the verbal lexicon that the class contains has dwindled over time but, two millennia after they ceased to be productive, strong verbs have kept a firm presence in the grammar of most Germanic varieties.

³⁸ Nevins et al (2015) attempt to show experimentally that the L-morpheme is “dead” in Romance and that it died largely because of this. There are a number of problems with their design of the experiment and their interpretation of the results. Most important, in my opinion, is the fact that, even when a pattern is not easily generalizable by language users to new forms (this may well be true actually of most morphemes), it can hardly be said to be “dead”, as it continues to provide a template for the distribution of the forms that do obey the morphemic pattern. This is nowhere clearer than when language users fill-up the complete paradigms of verbs that only ever occur with certain values (e.g. 3SG and nonfinite forms in the case of weather verbs). Language users, when questioned about the 1SG or 1PL present forms of e.g. *llover* ‘rain’, have no doubt in offering *lloveo* and *llovemos* respectively. These cannot be memorized forms, of course, since they never appear in natural speech. The forms are created online, analogically from other verbs with the same formal alternations.

In lieu of an example where a formerly productive morpheme becomes unproductive and gradually decreases its presence in the lexicon until it is completely extinguished, I will present a few examples of this relentless migration of lexical items “deserting” a morphomic pattern. These will hopefully illuminate the reasons why particular lexical items may change their inflection by letting go of a morphomic alternation, and generalize a single form throughout the paradigm.

The Spanish N-morpheme is a relatively robust morphomic pattern, as it appears overtly in over 300 verbs (Herce Calleja 2016). The general trend, however, is for this class to lose members gradually over time. Cases of verbs abandoning the class are more numerous than cases of verbs acquiring an N-morphomic exponence analogically. The verbs that undergo paradigm levelling to become regular are usually found among the relatively infrequent lexical items. This suggests that it is at least partially a matter of insufficient input. If an N-alternating verb (e.g. *mentar/miento* ‘mention’) does not appear frequently enough in its two stems, speakers may simply never learn that it was supposed to have two forms in the first place. When this happens, because of the smaller frequency of use (a proportion of around 1 to 3) of the N-morpheme cells compared to its N-complement set of cells, the surviving alternant is usually the latter (i.e. *mentar/mento* in the case of this verb).

Another verb that is increasingly found without diphthongization is *degollar* ‘cut someone’s throat’. Thus, the N-morpheme verb *degollar/degüello* is being increasingly replaced by a non-alternating *degollar/degollo*. A similar levelling, one which is more underway (both diphthongization and lack thereof are prescriptively acceptable), is that of *asolar/asuelo* ‘destroy’ changing to *asolar/asolo*. Less frequently, however, it can also be the diphthong form that is spread to the rest of the paradigm as when *amoblar/amueblo* changes to *amueblar/amueblo*. The reason for the different directionality of the levelling in different verbs has to be found, I believe, in the synchronic affinity (or lack thereof) of these verbs with their etymologically related nouns *suelo* ‘ground’, *cuello* ‘throat’, and *mueble* ‘piece of furniture’. In the case of the first two, the related verbs *asolar* and *degollar* have become

divorced from their source nouns.³⁹ In the case of the later, the connection to *mueble* remains evident to the Spanish language user, which can steer the levelling into the preservation of this synchronic connection.

Apart from a low token frequency, another factor that may lead to a lexical item losing an alternation is the concrete forms involved in the alternation. As explained in Section 3.3, the formerly alternating verb *levar/lievo* split into two non-alternating verbs *llevar/llevo* and *levar/levo* as a result of the sound change /lje/ > /λe/, which transformed a typical N-morphomic alternation e/je into an exceptional one l/λ. This must have made it more difficult (although definitely not impossible, as witnesses its Romanian suppletive cognate: *iau iei ia luăm luați iau*) to identify e.g. *levar* and *levo* as forms of the same lexeme, which motivated the split and the analogical filling-out of the missing forms. Developments like this, and the diachronic formal convergence that morphemes often display, speak against taking a morphome's applicability to novel forms (see Footnote 38) as the (only) factor to assess whether a given pattern is "living" or "dead". Dichotomous taxonomies like this one are probably too coarse-grained, in any case, to capture a pattern's vitality in the grammar in any meaningful way.

4.2.2 Loss of morphosyntactic categories

Another, quite more abrupt way in which morphemes can disappear from a language is the loss of whole morphosyntactic categories. In the course of normal language change, whole natural classes of cells (usually characterized by comparatively infrequent values like DU, SBJV, PAST etc.) can be lost seemingly "in one fell swoop". When this happens, this will inevitably erase (any part of a) morpheme that fell inside the lost swath of the paradigm. Consider the following examples:

³⁹ In the case of *asolar/suelo*, the loss of the synchronic connection is to be found in the semantic drift of the verb *asolar*, which used to mean 'throw to the ground' before but now means simply 'destroy'. In the case of *degollar/cuello* the loss of a synchronic affinity must be due to the formal discrepancy /k/ vs /g/ produced by intervocalic voicing, which occurred only in the verb.

	Spanish <i>decir</i> 'say'		Pantesco (Loporcaro et al. 2018:297-298)		
	Indicative	Subjunctive	ɔ:tiri 'can'	'di:ɾɪ 'say'	've:ɾɪɾɪ 'come'
1SG	digo	diga	'pɔt:su	'di:kɔ	'veɾ:ɔ
2SG	dices	digas	'pɔ	'di:ʃɪ	've:ɾɪ
3SG	dice	diga	'pɔ	'di:ʃɪ	've:ɾɪ
1PL	decimos	digamos	pu'te:mɔ	di'ʃe:mɔ	vi'ne:mɔ
2PL	decís	digáis	pu'ti:ɾɪ	di'ʃi:ɾɪ	vi'ni:ɾɪ
3PL	dicen	digan	'pɔn:ɔ	'di:ʃɪnɔ	've:ɾ:u

Table 107: Present tense of some verbs in two Romance varieties

In Pantesco Italian, as well as in other southern Italian varieties, the present subjunctive fell out of use in the language and eventually disappeared.⁴⁰ Without this tense, the earlier L-morpheme stems (with classically L-morphomic exponences like /ts/, /k/, and /ŋ/) have become confined to a single cell in the paradigm, which can never be morphomic as defined in this dissertation. Something similar can happen in the case of TAM morphemes like PYTA:

	Latin	Portuguese	Galician	Somiedo Asturian	French	Alpago Italian	Nuorese Sardinian
PAST.IND	fēcerat	fizera	fixera	fi'jera	-	-	-
PAST.SBJV	fēcisset	fizesse	fixese	-	fɪt	'fese	-
PRES.IND	fēcit	fez	fixo	'fiʃu	fit	-	-
PRES.SBJV	fēcerit	fizer	-	-	-	-	-
FUT.IND	fēcerit						

Table 108: 3SG forms of 'make/do' of former perfective tenses

The set of tenses that was perfective in Latin (and therefore was characterized by the perfective stems that gave rise to PYTA) is quite faithfully maintained in western Romance varieties (see e.g. Portuguese above). As one moves (west) along the Romance dialect continuum, however, less of these tenses have been preserved: three are preserved in Galician, two in Somiedo Asturian (see Cano González 1981) and in French (although different

⁴⁰ See Servigliano Romance in Table 106 for an intermediate variety which has lost this tense (or has merged it with the indicative) only in the non-3 forms.

ones), only one in Alpage Italian (see Zörner 1997), and none in Nuorese Sardinian (see Pittau 1972). In the last two varieties, the PYTA TAM morpheme is and can logically be no more.

4.2.3 Sound change

Most of the processes identified in Section 4.1 as potential creators of morphemes can also participate in their disappearance or in their change into a different pattern. In this line, a force that may be involved in the demise of a morphemic pattern is sound change. In the same way as sound changes can introduce alternations into formerly non-alternating paradigms, they can also disrupt pre-existing morphemic patterns. Consider the following:

	Macerata (Maiden et al. 2010)				Standard Italian			
	'sleep'		'feel'		'sleep'		'feel'	
	SG	PL	SG	PL	SG	PL	SG	PL
1	'dɔrmo	dur'mimo	'sɛndo	sin'dimo	'dɔrmo	dor'mjamo	'sɛnto	sen'tjamo
2	'durmi	dor'mete	'sindi	sen'dete	'dɔrmi	dor'mite	'sɛnti	sen'tite
3	'dɔrme	'dɔrme	'sɛnde	'sɛnde	'dɔrme	'dɔrmono	'sɛnte	'sɛntono

Table 109: Present indicative of two cognate verbs in two Italian Romance varieties

The regularly expected distribution of the N-morpheme (illustrated in Table 109 by Italian) has been disrupted⁴¹ in various Italian varieties as a result of sound change. In Macerata, for example, an anticipatory assimilation of the stem vowel to a following /i/ in the verbs shown in Table 109, has modified the paradigmatic domain of occurrence of the classically N-morphemic open-mid vowels. Something similar has happened in French:

⁴¹ In line with the *modus operandi* in the rest of this dissertation, morphemes here are defined over their paradigmatic distribution. Thus, the morphemes of Italian and Macerata above are considered different (albeit cognate) morphemes. Thus, the change in Macerata involves the disappearance of the SG+3PL morpheme (and the emergence of another one). This is the reason why it has been presented in this section (i.e. a section about morpheme disappearance). This way of thinking or talking about it is entirely a narrative convenience and one could just as easily have expressed it as the Italian-type morpheme becoming a Macerata-type one.

	Indicative		Subjunctive	
	SG	PL	SG	PL
1	ɾvǎ	ɾvənɔ̃	ɾvɛn	ɾvənʝɔ̃
2	ɾvǎ	ɾvəne	ɾvɛn	ɾvənʝe
3	ɾvǎ	ɾvɛn	ɾvɛn	ɾvɛn

Table 110: Present tense of French *prendre* 'seize'

Sound change in French treated some segments differently when they were word-final and when they were intervocalic. Thus, an older ɾvɛn\# became ɾvǎ , while in ɾvɛn-ə the stem was “protected” from that change and managed to remain ɾvɛn . In the same way as in the case of Macerata in Table 109, the result of this sound change is that the N-morphome exponent $/\epsilon/$ is evicted from part of its original paradigmatic distribution. The integrity of the N-morphome is, thus, broken in French and in Macerata Italian by later sound changes.

4.2.4 Analogy

As in the case of morphome emergence, analogical forces of various kinds can also be the decisive ones behind the loss of a morphome or its change into a different pattern. Some of the cases presented before (Wambisa in Table 94, and Servigliano in Table 106), already constituted examples of a morphomic pattern being analogically changed into a different one. This section will elaborate on the possible analogical changes to a morphomic pattern.

4.2.4.1 Analogical change into a natural class

Received wisdom in morphomic literature has it that “the death of morphomic patterns does not arise through alignment of alternation patterns with coherent functional or phonological determinants of their distribution” (Maiden 2018b:6). As a general trend in Romance this seems like it might be largely true.

There are a few exceptions, however. One of them is the retreat of the PYTA root to a single tense (most usually the preterite) in some varieties of Aragonese (see e.g. /tu'βemos/ vs /te'nešemos/, /su'pjemos/ vs /sa'pešemos/, /ki'sjemos/ vs /ke'rešemos//, /estu'βjemos/ vs /es'tasemos/ etc. in Panticosa, see Nagore Lain 1986).

Another case of a Romance morpheme retreating into a natural class can be found in Gallo-Romance, where the L-morpheme has abandoned the 1SG indicative, thus becoming confined to the present subjunctive. Consider the following paradigms:

	'know'		'be worth'		'be able to'	
	Indicative	Subjunctive	Indicative	Subjunctive	Indicative	Subjunctive
1SG	'sabu	'saʃe	'valu	'vage	'pwo	'puske
2SG	'sabes	'saʃes	'vales	'vages	'pwos	'puskes
3SG	'sabe	'saʃe	'vow	'vage	'pwo	'puske
1PL	sa'ben	sa'ʃen	va'len	va'gen	'pwen	pus'ken
2PL	sa'be	sa'ʃe	va'le	va'ge	'pwe	pus'ke
3PL	'sabun	'saʃen	'valū	'vagen	'pwō	'pusken

Table 111: Present tense conjugation of three Seyne Occitan verbs (Quint 1998)

Cases like these are sometimes explained by Maiden (2018b) not so much as a fall-back to morphosemantic distributional criteria but in alternative ways. For example, for Aragonese, he suggests the retreat of PYTA to rhizotonic cells (all of which must have occurred initially in the preterite) followed by its subsequent analogical extension to the rest of the preterite cells. In the case of the Occitan development in Table 111, he attributes the change at least in part to the effects of sound changes, that is, to the different treatment in Gallo-Romance of the 1SG indicative suffix -o, which is often subject to deletion unlike subjunctive present suffix -a. When they become word-final, some stem final consonants devoice in the 1SG present indicative, thus breaking stem identity with the present subjunctive.

Even though these, as probably most analogical changes, must be conditioned by a multiplicity of factors, there is, I believe, little reason to not consider the alignment to morphosemantic values one of the motivations (maybe even the most important one) of these morphological changes. Beyond the morphomic literature on Romance, in fact, the alignment of formatives to natural classes (in order to perform functional roles) has usually been considered relatively common (see e.g. Wurzel 1980). Consider some of these well-known analogical changes:

	Pre-Old High German		Old High German		Early Old Eng.		Late Old Eng.	
	SG	PL	SG	PL	SG	PL	SG	PL
NOM	*lamb	*lamb-ir-u	lamb	lamb-ir	fōt	fēt	fōt	fēt
ACC	*lamb	*lamb-ir-u	lamb	lamb-ir	fōt	fēt	fōt	fēt
DAT	*lamb-ir-a	*lamb-ir-um	lamb-e	lamb-ir-um	fēt	fōtum	fōte	fēten
GEN	*lamb-ir-as	*lamb-ir-o	lamb-es	lamb-ir-o	fōtes	fōta	fōtes	fēte

Table 112: Declension of OHG ‘lamb’ (Wurzel 1980:445-448) and OE ‘foot’ (Fertig 2016:436)

Germanic offers some well-known examples of morphological forms changing an inherited unnatural distribution into a natural one in order to perform morphosemantic roles. Some other times, as in Occitan above, there are confounding factors (in the form of formatives which already have the target natural distribution). In this way, some changes into a natural class might also be partially explained as formally-motivated analogies. Cases like the ones in Table 112, however, show that morphosemantic values can also act as templates for the distribution of formatives *even in the absence of suitable formal templates*. Older Germanic languages were extremely fusional and thus, before the emergence of *-ir* and Umlaut plurals, no formatives existed that marked PL exclusively, only number-case suffixes like e.g. DAT.PL -um. No form, thus, could have acted as a model or attractor for these other forms.

Analogical changes like the ones in Table 112, thus, demonstrate that morphosemantic factors can be a force involved in the demise of morphosyntactically unnatural patterns. The reason why this is not observed exceedingly frequently in morphemes may be just derived from our definition of the phenomenon. If we require, for a morpheme to be recognized as such, that the same unnatural pattern be repeated *with various different forms*, then we are

effectively selecting cases where formal niches are particularly robust. Consider the changes we would be requiring:

	<i>salir</i> 'exit'		<i>caber</i> 'fit' (actual form)		<i>caber</i> 'fit' (hypothetical)	
	IND	SUB	IND	SUB	IND	SUB
1SG	salg-o	salg-a	kep-o	kep-a	*kab-o	kep-a
2SG	sal-es	salg-as	kab-es	kep-as	kab-es	kep-as
3SG	sal-e	salg-a	kab-e	kep-a	kab-e	kep-a
1PL	sal-imos	salg-amos	kab-emos	kep-amos	kab-emos	kep-amos
2PL	sal-is	salg-ajs	kab-ejs	kep-ajs	kab-ejs	kep-ajs
3PL	sal-en	salg-an	kab-en	kep-an	kab-en	kep-an

Table 113: Present paradigms of Sp. *salir* and *caber* (phonetic form)

If, for example, as hypothesized in Table 113, the stem alternation pattern in *caber* became aligned to morphosyntactic distinctions, this would not constitute a simplification of the system in any intuitive way. In the case of morphomic affinities which, like the one in Table 113 above, are repeated with different exponents, changing one exponent will not get rid of the morpheme. Transitioning into a semantic niche might be particularly difficult in the case of morphemes because this coextensivity of exponents would mean either i) that various different formatives in many different verbs would have to be changed analogically *at the same time*, which seems an unlikely development given the stealth that usually characterizes language change, or ii) that the morphological exponence would need to get worse, i.e. less predictable, before it can get better again.⁴² That is, the change in the paradigm of *caber* hypothesized in Table 113 would actually constitute a complexification of the system (a deleterious mutation, so to speak), which does not quite fit with the simplifying role that is usually attributed to analogical change in general (Sturtevant 1947).

⁴² "Using a metaphor borrowed from evolutionary biology (e.g. Dawkins 1996), it is as if Spanish were trapped in this respect in a suboptimal summit from which it cannot reach a more optimal design (i.e. form-function isomorphy) because it would have to get worse (i.e. less adapted because of the drop in predictability) before it could get any better." (Herce forthcoming-b)

4.2.4.2 Analogical change into an unnatural class

Much like it was shown in the case of morpheme emergence, not all analogical processes result in more one-to-one form-function relations. Some of the cases presented in Section 4.1.3 already illustrated how both natural and unnatural classes could be changed into a different unnatural pattern by means of morphosyntactically-driven analogical changes. Since this is, I believe, clear by now, I will focus here instead on the analogical disintegration of Romance PYTA (or its change into another, different pattern, if you will) across a number of varieties. Consider the following:

	Sicilian 'have' (Maiden et al. 2010)		Italian 'cook'		Oscos Galician 'put' (Maiden 2018b:76)	
	PRET	IPF.SBJV	PRET	IPF.SBJV	PRET	IPF.SBJV
1SG	'appɪ	a'vissɪ	'cossɪ	cuo'cessɪ	'puʃɛŋ	po'ŋɛsɛ
2SG	a'vijʃɪ	a'vissɪvʊ	cuo'cestɪ	cuo'cessɪ	po'ŋitʃɛ	po'ŋɛsɛs
3SG	'appɪ	a'vissɪ	'cosse	cuo'cesse	'puʃo	po'ŋɛsɛ
1PL	'appɪmʊ	a'vissɪmʊ	cuo'cemmo	cuo'cessɪmo	po'ŋɛmos	po'ŋɛsɛmos
2PL	a'vijʃɪvʊ	a'vissɪvʊ	cuo'ceste	cuo'ceste	po'ŋɛstes	po'ŋɛsɛðɛs
3PL	'appɪrʊ	a'vissɪrʊ	'cossero	cuo'cessero	po'ŋɛroŋ	po'ŋɛsɛŋ

Table 114: Remnants of PYTA root in various Romance varieties

As the above Table 114 illustrates, stress in the root and the PYTA allomorph often coincide in Romance (even if their actual paradigmatic distribution may differ from one variety to another). This can be found even in varieties where PYTA has “morphemicised” by aligning to tense distinctions. In Sassarese Sardinian (Maiden et al. 2010), for example, a stressed PYTA root is sometimes found all through the preterite (e.g. 1SG *'fesi*, 2PL *'fesidi*) and an unstressed non-PYTA root has spread to all the forms of the continuant of the Latin pluperfect subjunctive (e.g. 1SG *fa'dzissia*, 2PL *fa'dzissiaddi*). Many varieties have, thus, trimmed the inherited distribution of perfective root allomorphy and made rhizotony and the PYTA root (both purely morphological properties) paradigmatically coextensive (see Esher 2015 and Maiden 2018a).

The developments of Table 114, therefore, illustrate another possible motivation for the loss of morphemes in a language. The “fall” of PYTA has come about diachronically largely as a result of its redistribution in the paradigm to fit the template provided by a different morphological form (i.e. stress). The analogical matching of the distribution of two formerly independent morphological traits or formatives (i.e. modifying the paradigmatic distribution of root allomorphy to become identical to that of rhizotony) constitutes an interesting development which is in line with the overall findings of morphomic literature in other domains (Herce 2019).

In many varieties, therefore, rhizotony and the special root have become part of the same allomorphic phenomenon synchronically. Although this story seems largely correct, explaining all these morphological changes and patterns exclusively as a retreat of PYTA to the rhizotonic cells of the paradigm may not be sufficient.⁴³ The paradigmatic distribution of rhizotony itself has also been subject to changes in its evolution from Latin. In Italian and Sicilian, for example, the 3PL has become rhizotonic, and in Italian and Asturian, the 1PL has changed to become arhizotonic. These changes also demand an explanation.

Although largely speculative, a greater paradigmatic parallelism of preterite rhizotony to present tense rhizotony could explain most of the analogical changes mentioned: for example, the impulse to make the 1PL preterite arhizotonic (like the present), and the motivation for making the 3PL preterite rhizotonic (also like the present). The drive to make root allomorphy coextensive with stress could also be understood as an analogical influence from the morphology of the present tense and might explain why, as Esher wonders (2015:522-523), “stem allomorphy is aligned with stress placement rather than stress placement being aligned with stem allomorphy”. Consider the following forms:

⁴³ Accounting for synchronic segmental properties by reference to suprasegmental traits, or for the latter by reference to the former (or to morphological specification exclusively) is obviously not very enlightening.

	Present subjunctive of 'die' (actual and expected forms)		Preterite of 'make' (actual forms)		Preterite of 'make' (expected forms)	
	SG	PL	SG	PL	SG	PL
1	'muoia	mo'riamo	'feci	fa'cemmo	'feci	'fecemo
2	'muoia	mo'riate	fa'cesti	fa'ceste	fe'cesti	fe'ceste
3	'muoia	'muoiano	'fece	'fecero	'fece	fe'cero

Table 115: Paradigmatic extension of rhizotony in Italian present and preterite

In Table 115 above it is shown how the analogical changes in the domain of stress placement in Italian have increased the parallelism of preterite and present. Something similar happens in the distribution of stem allomorphy:

	Present subjunctive of 'die' (actual and expected forms)		Preterite of 'make' (actual forms)		Preterite of 'make' (expected forms)	
	SG	PL	SG	PL	SG	PL
1	'muoia	mo'riamo	'feci	fa'cemmo	'feci	'fecemo
2	'muoia	mo'riate	fa'cesti	fa'ceste	fe'cesti	fe'ceste
3	'muoia	'muoiano	'fece	'fecero	'fece	fe'cero

Table 116: Paradigmatic extension of apophony in Italian present and preterite

The paradigmatic coextensiveness of stress and root allomorphy is regularly expected in the present as a result of sound change. In the preterite, however, it would not have been expected and has been analogically introduced. Overall, therefore, the analogical changes observed in the paradigmatic configuration of PYTA reveal how formally-driven analogy to other forms in the same subparadigm (i.e. rhizotony) and/or templates from other subparadigms (i.e. the N-morpheme) can both result in the loss or paradigmatic reconfiguration of an inherited morphomic pattern.

4.2.5 Mixed causes

As the discussion in the previous section has already begun to suggest, the story of most morphomes' demise usually involves a combination of factors, rather than one motivation exclusively. For example, Maiden (2018b:288) discusses the following pattern:

	'sing'		'save'	
	SG	PL	SG	PL
1	tsãtə	tsãtã	sawvə	sovã
3				
2	tsãta:		sova:	

Table 117: Present tense of two Gartempe Romance verbs (Maiden 2018b:288)

As Maiden explains, stem allomorphs like sawv- (vs sov-) are the descendants of rhizotonic (i.e. N-morphomic) forms. In most of Gallo-romance, 2SG=2PL and 1PL=3PL syncretisms in non-alternating verbs are a result of regular sound changes. In the case of verbs with stem alternation, whole-word syncretisms should not have resulted. However, the consolidation of the sound-change-triggered syncretisms at the morphosyntactic level motivated the levelling of the form of the stem inside these newly-emerged paradigmatic cells. Thus, the N-morphome stem alternant changed its etymological distribution and became confined to the 1SG=3SG cell. Sound change and analogy, thus, both conspired to get rid of the N-morphome in Gartempe Romance.

This case constitutes but one example of morphomes' demise as a result of several different forces. Although the different motivations have been discussed separately in this section for convenience, in reality it is most of the times it is a combination of factors that is responsible for a morphome's demise in a language (e.g. in the case of the L-morphome in Occitan in Table 111 probably sound changes, formal analogies, and form-function isomorphy).

4.3 Discussion

The emergence and disappearance of morphomic patterns in a language show important parallelisms. Largely the same forces have been identified as potential motivators of both morpheme creation, morpheme change, and morpheme loss. This is not really surprising: it merely indicates that anything that leads to changes in an inflectional paradigm is a potential creator and/or destructor of (both morphomic and morphemic) morphological patterns. In the roughest terms, grammaticalization and sound changes introduce formatives and morphological alternations into the paradigms and language users have to deal with them. They will try to find a rationale or purpose for the distribution of inflectional forms in order to recreate faithfully the grammatical system that was handed down to them. If they fail, analogy will occur. Because it is driven by language users' necessity to use language productively even when they may be unsure about what an actual form should be (what has come to be known as the paradigm cell-filling problem), analogical change is one of the (if not the single most) important sources of evidence regarding the nature and organization of morphological architecture and its cognitive representations.

The most important contribution of the present research in this respect has been the identification of two quite different organizational principles in the domain of inflectional morphology. One is meaning. The other one is form itself. Both can provide the niche, template or domain for sub-word units. The majority of morphological models and linguists assume as self-evident that meaning is the most relevant factor when accounting for morphological forms. The reader is, thus, likely to need little convincing that this factor is of the utmost importance. That forms can by themselves serve a similar role is much less clear and has not been studied as extensively. This discussion section will be devoted largely to the presentation and discussion of some individual cases of form-derived morphological niches and of cases where form-derived templates take the upper hand over morphosyntactic or semantic ones.

Romance is well known for this in the literature. In various Romance varieties, formerly independent lexical items (e.g. Latin *ambulāre* and *vadēre*) are combined into a single suppletive paradigm following the same pattern as the formal alternations generated by

regular sound changes (e.g. the vowel apophonies associated with rhizotony). Such developments are well-known, so evidence from other language families will be presented here instead. Although not nearly as widely discussed, Luxembourgish, for example, as well as other Germanic languages, can also provide some beautiful examples of the power of forms to act as templates or niches for other forms. Consider the following OHG paradigms and their descendants in Luxembourgish:

	<i>faran</i> 'drive'		<i>wësan</i> 'be'		<i>kweman</i> 'come'		<i>mahhōn</i> 'make'	
	SG	PL	SG	PL	SG	PL	SG	PL
1	faru	farem	bim	birum	kwimu	kwemem	mahhōm	mahhōm
2	feris	faret	bist	birut	kwimis	kwemet	mahhōs	mahhōt
3	ferit	farant	ist	sint	kwimit	kwemant	mahhōt	mahhōnt

Table 118: Present tense of four Old High German verbs (Braune & Reiffenstein 2004)

	<i>fueren</i> 'drive'		<i>sinn</i> 'be'		<i>kommen</i> 'come'		<i>maachen</i> 'make'	
	SG	PL	SG	PL	SG	PL	SG	PL
1	fueren	fueren	sinn	sinn	kommen	kommen	maachen	maachen
2	fiers	fuert	bass	sidd	kënns	kommt	méch	maacht
3	fiert	fueren	ass	sinn	kënnt	kommen	mécht	maachen

Table 119: Present tense of the same four verbs in Luxembourgish (Schanen 2004)

In the history of Germanic, a vowel was sometimes fronted or raised before an /i/ in the next syllable (see Table 83). In the verbal paradigm, this happened in the 2SG and 3SG of some verbs (see e.g. *faran*), which gave rise to an alternation pattern opposing 2SG/3SG to the other forms of the present. These regular (i.e. sound-change-created) stem alternations, however, have been used as a template for the distribution of other formal differences. They have acted, diachronically and in processes of analogical change, as “islands” that favour internal homogeneity. Formal differences are, thus, pushed, if anything, to the borders between these sets of cells.

In the verb ‘be’, for example, we observe how Luxembourgish analogically establishes stem identity within a set of cells where several different roots were found before. The earlier 3PL form seems to have served as a model for the rest of the cells. In the verb ‘come’, the stem-final bilabial nasal is able to assimilate in place of articulation to a following alveolar only in 2SG/3SG. The peer-pressure for stem identity within the complement set of cells makes it impossible for the 2PL to assimilate in the same way.⁴⁴ In the case of the verb ‘make’, we see how an alternation between 2SG/3SG and the rest of the cells is sometimes analogically introduced into verbs that would not have had the alternation etymologically.

One of the most striking examples of a formal alternation pattern providing the niche for other formatives is found in the Kiranti language Yakkha (Schakow 2016). In this and in other East Kiranti languages, verbs have two stems, one of which (usually longer) occurs before suffixes beginning with a vowel, while the other one occurs before consonants. Consider the non-past tense paradigms of transitive (Table 121) and intransitive (Table 120) verbs in Chintang, a closely related language, for an approximated illustration of the system ancestral to these East Kiranti languages:

	SG	DU	PL
1EXCL	thap-maʔã	thap-cekeŋa	thab-ikiŋa
1INCL		thap-ceke	thab-iki
2	a-thap-no	a-thap-ceke	a-thab-iki
3	thap-no	u-thap-ceke	u-thap-no

Table 120: Paradigm of Chintang ‘come level’ non-past, intransitive (Paudyal 2013:86)

	SG	DU	PL
1EXCL	pid-ukuŋ	pi-cokoŋa	pid-ukumma
1INCL		pi-coko	pid-ukum
2	a-pid-oko	a-pi-coko	a-pid-ukum
3	pid-oko	u-pi-coko	o-pid-oko

Table 121: Chintang ‘give’ non-past, transitive, 3SG patient (from Paudyal 2013:294)

⁴⁴ Consider also the opposition, in modern German, of 3SG *ha-t* and 2PL *hab-t* ‘have’ (both from Old High German *habet*) for a comparable development.

Formal alternations are thus found on the right edge of the stem in these languages depending on the vocalic (shaded) or consonantal (unshaded) nature of the following segment. Although some alternations have become a bit more opaque synchronically (e.g. haks-V/haŋ-C, hops-V/hom-C) most are phonologically predictable or straightforward (e.g. chept-V/chep-C, thur-V/thu-C, ab-V/ap-C) in that they involve the simplification of (often illicit) consonant clusters, or intervocalic voicing. In any case, the shaded cells, and their complement set, share nothing but a common stem in these phonologically conditioned formal alternations. Observe, however, the situation in Yakkha:

	SG	DU	PL
1EXCL	am-meŋna	am-meŋciŋha	ab-iwaŋha
1INCL		am-meciya	ab-iwaha
2	am-mekana	am-meciŋha	ab-iwagha
3	am-meʔna	am-meʔciya	ŋ-am-mehaci

Table 122: Paradigm of Yakkha ‘come’ non-past, intransitive (Schackow 2016:243)

	SG	DU	PL
1EXCL	tund-waŋna	tum-meŋcuŋna	tund-wamŋana
1INCL		tum-mecuna	tund-wamna
2	tund-wagana	tum-mecugana	tund-wamgana
3	tund-wana	tum-mecuna	n-dund-wana

Table 123: Yakkha ‘understand’ non-past, transitive, 3SG patient (Schackow 2016:244)

As the paradigms above illustrate, the shaded vs the unshaded paradigm cells in Yakkha have acquired inflectional suffixes in common. Thus, a suffix *-wa* now characterizes the shaded cells and a suffix *-me* characterizes the unshaded ones. As Schackow (2016:230-231) explains, these suffixes go back ultimately to lexical verbs, which grammaticalized into the tense markers we find synchronically.⁴⁵ An utterly morphosyntactically unnatural stem alternation

⁴⁵ There is still today in the language a verb *wa-ma* that means ‘sit’, ‘stay’ or ‘live’. The verb *meʔ-ma*, in turn, has cognates in closely related languages (e.g. in Bantawa) where they mean ‘do’ or ‘cause’.

pattern, thus, has provided the niche for the incoming present tense suffixes, which thus adopt the exact and only paradigmatic configuration that could have possibly preserved the *status quo* (i.e. unchanged stem alternation patterns and preservation of phonological conditioning of the alternation).

I would like to conclude this discussion with a cautionary note about the correct analysis of those analogical morphological changes that result in morphosyntactically natural formal distinctions. Consider, for example, the following analogical change in Old Norse:

	Pre-Old Norse		Old Norse	
	SG	PL	SG	PL
1	*blōtu	*blōtum	blót	blótum
2	*blótiR	*blōteð	blótr	blóteþ
3	*blótiR	*blōta(n)	blótr	blóta

Table 124: Present tense of blóta ‘sacrifice’ (Wurzel 1980:451-452)

The same as in older stages of West Germanic (see e.g. OHG in Table 83), the 2SG and 3SG of some verbs must have been characterized by /i/-bearing inflections in ancestral stages of Scandinavian. I-Umlaut, thus, should, by regular sound change, have resulted in a 2SG/3SG vs 1SG/PL alternation in the language. However, we find that in Old Norse, in this and other verbs (e.g. *fara* ‘travel’: 1SG *fer*, or *koma* ‘come’: 1SG *kø̄m*), the umlauted vowel has been extended to the 1SG, thus resulting in morpho-semantically neat distinctions between singular and plural number values.

Because of the distribution the change has achieved, the first impulse (this is, for example, how Wurzel analyzes it) is to attribute the change to morphosyntax, that is, to language users’ desire to mark certain semantic distinctions overtly or to have more one-to-one form-meaning exponence patterns. There is, however, an alternative interpretation of these analogical changes.

Already in the Pre-Old-Norse stage, some very high frequency verbs (preterite presents like *kunna* ‘can’: 1SG *kann*, or *vita* ‘know’: 1SG *veit*) would have contained SG vs PL vowel

apophonies in the present tense. The analogical change of the i-Umlaut-generated stem vowel apophonies in Old Norse could, therefore, constitute also a formally-driven analogical change on the basis of these verbs. Because, in this and in many other cases semantic and formal templates converge, there is no way of unmistakably identifying the force that is (most) responsible for the change.

4.4 Conclusion

Chapter 4 has explored the ways in which morphemes arise, change, and disappear from a language and the forces and reasons behind it. Sound changes (in various ways), semantic drift, analogical change (both morphosyntactically and formally motivated), pattern interactions, grammaticalization, and maybe even language contact have been identified as forces that may be involved in morpheme emergence. Some of these (e.g. morphosyntax-driven analogy, grammaticalization, and [maybe] language contact) might be considered quite surprising given the nature of the patterns (N, L, PYTA) and families (Romance) that have dominated morphomic research to date. The only possible conclusion is that basically any process that can produce a change in the paradigmatic distribution of some form(s) can be involved in processes of morpheme emergence and loss.

The forces involved in morpheme emergence and loss seem at first sight not to be exceedingly different from the ones at play in morpheme diachrony. However, although more quantitative research into this matter would be welcome, the particularities of morphemes seem to make certain diachronic origins common (e.g. sound change) and others uncommon (e.g. borrowing). Of those morphemes in the synchronic database (see Chapter 5) whose diachronic history I have been able to track, as many as 50 involve sound change, another eight involve morphosyntactically-driven analogy, six form-driven analogy, five grammaticalization, four pattern interactions, and one semantic drift. Very often, more than one of these is involved in the history of a given morpheme.

The relative predominance of sound-change-generated morphemes seems, in any case, clear. The present diachronic section has also contributed to our knowledge of morphemes by

calling attention to and typologizing the various ways in which sound changes create morphemes. On the basis of their domain of application, sound changes can happen in both the morpheme and in its complement cells. On the basis of their result, sound changes can create morphemes by disrupting previous formal invariance (i.e. $A \sim A > A \sim B$), or by erasing a formal difference (i.e. $A \sim B > A \sim A$) between word forms that do not bear any particular morphosyntactic affinity. In addition to these types, as discussed in Section 4.1.1.1, it has been found to be quite common (a total of 12 such cases have been found here) for morphemes to emerge from zero-vs-marked morphological configurations. This fact (consider Zipf's law) might well lead to some cross-linguistic tendencies in the paradigmatic distribution of these morphemes (see Section 5.4.11).

When robust enough, the accidental paradigmatic results of any of these processes (i.e. sound change, grammaticalization, etc.) will be acquired. However, language is by definition a productive system. On the basis of a Zipfian input, language-users need to infer/construct a watertight system without holes. This means that these paradigmatic patterns, even when morphosyntactically unnatural, will not be learned simply as a long list of word forms and lexemes. Language users will need to actively employ the formal and predictive regularities they observe in their input to infer and produce unencountered forms. This is the mechanism that allows morphemes, whatever their origin, to sometimes become productive/active morphological categories that may, on a par with morphosyntactic values, participate in exponence rules and steer morphological change.

5 Morphemes in Synchrony

5.1 Criteria for inclusion into the synchronic morpheme database

The common practice in morphomic literature has been to identify morphemes on a case-by-case basis, taking into account a wide range of unstructured and relatively subjective criteria. The most important of these are i) the failure to identify a semantic or morphosyntactic property exclusive to those cells, and ii) some diachronic evidence that that particular set of cells has behaved in a unified way in analogical changes. Other criteria are seldom stated openly or discussed, but I suspect that theoretical morphological notions like blocking or defaults, the generality of a pattern across the lexicon, the degree of allomorphy involved and others are, sometimes at least, lurking in the back of the mind of the morphologist when they try to assess whether or not a given pattern is a morpheme or not.

It is evident that in the context of a typological investigation, such an approach is unsuitable. To quantify and classify morphemes, clear criteria are needed in order to overcome any personal biases of the analyst or of the different grammar writing traditions, and to allow for the replicability of the research and the falsifiability of typological claims. Since morphemes and morphemes are probably not natural kinds, their definition and borders are subjective to a great degree and open to debate. In order to make this research useful to the greatest possible audience, my goal in this respect will be to restrict my attention to the higher morphomicity end of the morpheme-morpheme scale. I will therefore set purposefully high requirements for inclusion of a particular morphological structure into the present cross-linguistic morpheme database.

5.1.1 Unmistakably unnatural paradigmatic distribution

In earlier sections it has already been established that, of the various loosely-connected meanings of the term morpheme, this dissertation is only going to be concerned with what Round (2015) called metamorphemes, that is, with cells that, within the inflectional paradigm of a given lexeme, share particular exponents.

When assessing if a set of cells constitutes a natural class or not, the assumed feature structure plays a crucial role. For someone who is maximally reticent to grant the status of natural class, the syncretism of any two or more values but not all (e.g. dual and plural; dative, genitive and ablative) will count as morphologically stipulated.

Many (maybe most) morphologists will be more permissive in this respect and argue for the existence of feature structures of some sort which allow for certain values (maybe those which are perceived to be closer semantically or those which are more frequently syncretic cross-linguistically) to be able to feature together in rules of exponence as a sort of macro-value. Empirical evidence tells us, for example, that first and second person tend to be syncretic far more frequently than first and third person (Baerman et al. 2005). With that reasoning in mind, we could classify the former as natural and the latter as unnatural.

Because I want the threshold for ‘naturalness’ to be high, I will go still a step further and **I will allow any two or more values of a feature to form a natural class**. Consider the following paradigm:

	Present		Past	
	SG	PL	SG	PL
1	lugub	lu'ggəm	lugiz	lugizmə
2	lugud	lu'ggəd	lugist	lugist(ə)
3	lugub	lu'ggəbəd	lugiz	lugist(ə)

Table 126: Livonian (Uralic) ‘read’ (Corbett & Baerman 2006:240)

As shown in Table 126, the 1SG and 3SG cells are syncretic in Livonian, in present and past, with different formatives, which suggests a degree of systematicity. Those values are also not generally considered to be particularly close semantically or prone to syncretization cross-linguistically. However, because of the criterion espoused above, this syncretism will not count as morphomic for the purposes of the inclusion into the present synchronic study.

A consequence of imposing these restrictions is that patterns of morphological identity will need to be at least two-dimensional (i.e. will have to involve at least two features) for them to be considered unnatural here. Furthermore, to be absolutely sure that a given syncretism, whether partial or total, is morphomic, and to be able to measure the degree to which it is morphomic, the **features and values involved will need to be perfectly orthogonal**. It is clear that many cells in a paradigm do not meet these requirements. Consider the following:

	Indicative				Subjunctive			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	á	eigum	átti	áttum	eigi	eigum	ætti	ættum
2	átt	eigið	áttir	áttuð	eigir	eigið	ættir	ættuð
3	á	eiga	átti	áttu	eigi	eigi	ætti	ættu
Infinitive	eiga							
Participle	átt							

Table 127: Paradigm of *eiga* ‘own’ in Icelandic

In Icelandic, as in the paradigm above, every single verb except for the verb ‘be’ has the same stem in the infinitive, in the plural of the present indicative, and in the present subjunctive. Furthermore, the infinitive and the 3PL present indicative are whole-word syncretic, again except in the verb ‘be’.

There is distributional-semantic (Bonami 2017) and syntactic evidence that finite and nonfinite forms are more different from each other than any two finite forms. Thus, any morphological syncretism of a finite with a nonfinite form which does not extend to the totality of the paradigm should probably be regarded as morphomic. However, these morphological affinities will not be included in the present synchronic survey. The lack of orthogonality between the features and values involved makes it impossible to measure the degree of morphosyntactic coherence (see Section 4.2.2) of a metamorphome consisting e.g. of 3PL.PRES+INF. I will therefore limit my attention in this dissertation to those parts of the

paradigm where orthogonality does hold, excluding therefore those paradigm cells (e.g. non-finite forms, imperatives, 1PL inclusives etc.) where the orthogonality to other features is jeopardized. It has to be noted, in relation to this approach, that if a given syncretism is unmotivated within an orthogonal subset of the paradigm, then it will necessarily remain unmotivated in any larger superset.

Another challenge in identifying *bona fide* morphemes, thus, is presented by cases of TAM morphemes (Smith 2013). Whereas some features, like person and number, or case and number, are more or less well-behaved regarding their orthogonality, others like TAM are more complicated. Thus, it is often difficult to find perfect orthogonality of tense and aspect, aspect and mood, or tense and mood. The common difficulty or impossibility to organize these into orthogonal features has the consequence that establishing what counts as a natural class is difficult at this level.

Consider for example a morphological stem alternation pattern which is present in Daai Chin (Sino-Tibetan) in around 20% of the verbs. One stem (arbitrarily labeled Stem A by So-Hartmann 2009) is used in i) indicative transitive verbs (unless negative or in the presence of a focus shift), ii) subjunctive, iii) applicatives, iv) most non-final adverbial clauses, and v) most nominalizations. The other stem, stem B, is present in i) indicative intransitive verbs, ii) interrogative (unless in the presence of narrow focus), iii) imperative and in iv) non-final clause chains. Each of the stems seems to be, thus, involved in the expression of a 'hodgepodge' of values with no obvious relation to one another. This suggests that these are unnatural classes. However, because of the unstructured nature of the values involved, there is no way to assess this, let alone quantify this as I intend to do in this dissertation. Because of this, I will exclude this kind of morpheme from the present cross-linguistic study, even if it includes some of the most famous morphemes in the literature like PYTA (as present in Spanish or Portuguese) or the Latin third stem.

The last type of paradigms that will be excluded from here are those that, even in the presence of perfect feature orthogonality, involve features that are very closely related by virtue of having similar or identical values. Consider the present paradigm from Komnzo:

		Patient number		
		SG	DU	PL
Agent number	SG	-wr	-n	-wr
	DU	-n	-n	-n
	PL	-wr	-n	-wr

Table 128: Form of a number marking formative in Komnzo verbs (Döhler 2018:218)

Agent number and patient number are, of course, different things. A suffix that appears in the patient dual and/or agent dual is thus, from this point of view, as unnatural as any of the most well-known morphemes in the literature like the N-morpheme (SG and/or 3) or the L-morpheme (1SG.PRES and/or PRES.SBJV). It is true, however, that the form *-n* in Komnzo is clearly marking duality, which is more morphemic than morphomic. Cross-linguistic evidence shows that, when the same values appear in two orthogonal axes of the paradigm, distributions of this type are not infrequent and may arguably be morphosyntactically derivable depending of what we allow rules of exponence to do. Apart from agent number and patient number (see also Erzya in Table 67), other combinations where this may be found are agent person and patient person, possessor number and possessee number etc. These paradigms will also be excluded here preemptively from the ranks of morphemes.

As mentioned before, the exclusion of the structures that have been presented throughout this section responds to a desire to focus on the higher morphomicity end of the morpheme-morphome scale. The result of this is that, most often, the metamorphemes in this synchrony-oriented part of this dissertation will be found in person-number and case-number inflection. It is hoped that the greater morphomhood and measurability achieved with these standards outweighs the loss of variability and datapoints in general.

5.1.2 Unmistakably systematic formal identity

The previous requirement involved setting a high bar for considering a particular paradigmatic distribution unnatural. This section is devoted to setting a high bar for regarding a formal identity as systematic.

The impulse to classify morphological identities as systematic (in other words, those which are allegedly meaningful and part of the fabric of grammar) or accidental (those that should be understood as mere homophonies and largely irrelevant for the deeper grammatical system) is a generalized one among morphologists. As far as I understand it, the reasoning behind this distinction is that speakers, in their inner cognitive grammatical representation of their language, may code two identical forms into separate entries (e.g. [/mʌsl/₁: body tissue] vs [/mʌsl/₂: mollusk]) or instead code them as different meanings of a single entry (e.g. /mʌsl/₁: body tissue, strength). This distinction is obviously problematic for our present purposes because of its empirical inaccessibility (see however Section 2.1.2).

Many linguists, thus, have faced the challenge of finding some test or property to tell apart these two kinds of formal identities or to at least discard most unsystematic cases. One of these (mentioned e.g. in Zwicky 1991, and Haspelmath & Sims 2010) is the ability of a form to resolve syntactic feature conflicts (see Section 2.1.1). This test is unsuitable in a large typological endeavour such as the present research because i) it can only possibly be used in cases of whole-word syncretism (and morphomic structures may involve stem or affixal material separately) and because ii) the typologist hardly ever has access to the wealth of descriptive data that would be required in every language to have enough information on these morphosyntactic-conflict resolution-triggering constructions. Other tests and criteria, as already discussed in Section 2.1, are also unsuitable.

Undoubtedly for reasons similar to these, some of the linguists that have faced this challenge before (e.g. Johnston 1996 and Stump 2014) have opted for a different, less sophisticated but more easily implementable solution to discard accidental homophonies.

I propose to rely primarily on the criterion of co-extension of the homonymy under allomorphy (...) in assessing systematicity. The reasoning is this. If we find that a suffix x in a certain context realizes properties a and b, it is entirely possible that the homonymy is accidental and of no more account than the two senses of bank in English. But if we find that in another context a suffix y also realizes properties a and b, then it becomes more likely that the homonymy is systematic. (...) Naturally one's confidence in systematicity rises as the number of co-extensive homonymies does. (Johnston 1996:15).

This solution to regard a pattern as systematic if it is found to be **instantiated with more than one formal exponent** is in line with current morphomist practice⁴⁶ and will be adopted here too for inclusion of a morphological identity into my synchronic morpheme database. There are, however, two more caveats to be presented regarding the nature of those forms.

The first one is that, although suprasegmental features like tone or stress can obviously be phonemic and can perform grammatical functions, I will not include here any morphemes which are based on these formal exponents. The only reason for this is that, because the number of tones or stress possibilities in a word tends to be small within a particular language (i.e. smaller than the language’s segmental inventory), the chance of accidental formal identity is very high regarding those phonological traits.

The second is that, as mentioned in Section 2.3, formal identity is not enough. The identity has to be **exclusive to the paradigm cells constitutive of the putative morpheme**. In other words, there must be minimally one segment which appears in every single one of the cells constitutive of the metamorpheme and in no other paradigm cell outside of it. Consider again the following whole-word syncretism:

	First conjugation, <i>m̄n̄n̄j̄</i> 'go'				Second conjugation, <i>daśanj̄</i> 'prepare'			
	Present		Future		Present		Future	
	SG	PL	SG	PL	SG	PL	SG	PL
1	m̄n̄i-śko	m̄n̄i-śko-m	m̄n̄-o	m̄n̄-o-m	daśa-śko	daśa-śko-m	daśa-lo	daśa-lo-m
2	m̄n̄i-śko-d	m̄n̄i-śko-d̄j̄	m̄n̄-o-d	m̄n̄-o-d̄j̄	daśa-śko-d	daśa-śko-d̄j̄	daśa-lo-d	daśa-lo-d̄j̄
3	m̄n̄-e	m̄n̄-o	m̄n̄-o-z	m̄n̄-o-z̄j̄	daśa	daśa-lo	daśa-lo-z	daśa-lo-z̄j̄

Table 129: Verb agreement in Udmurt (Uralic) (Csúcs 1988:142)

The 3PL present and the 1SG future (and only these two cells) are always whole-word syncretic in Udmurt. The formatives involved in this syncretism, however, are not exclusive to these two cells. Both *-o* in the first conjugation and *-lo* in the second appear all through the

⁴⁶ Maiden (2018b:20) goes as far as arguing that the replication of a pattern with a different form is what “guarantees that such data are morphomic”.

future tense cells. Thus, a description of the inflectional exponence of Udmurt need not make any reference to the class 3PL.PRES+1SG.FUT. It is the cells 3PL.PRES+FUT that fulfill the requirements for morphomehood. The absence of a formative (or in other words, a zero-morpheme), will not count as a formal affinity for the purposes of inclusion into my morphome sample, where only overt formatives will be considered.

Also in this same vein of trying to avoid reference to dubious objects and/or theoretical analyses in the identification of morphomes here, subtractive affixes will not be allowed to feature in synchronic morphology. Consider the following French paradigm:

	PRES.IND		PRES.SUB		IMPERF		FUT		COND	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	li	lizõ	liz	lizjõ	lizɛ	lizjõ	liʁɛ	liʁõ	liʁɛ	liʁjõ
2	li	lize	liz	lizje	lizɛ	lizje	liʁa	liʁɛ	liʁɛ	liʁje
3	li	liz	liz	liz	lizɛ	lizɛ	liʁa	liʁõ	liʁɛ	liʁɛ

Table 130: Paradigm of French *lire* ‘read’

In the above verb’s inflection, the segment /z/ appears at the end of the stem only in the plural forms of the present indicative and in the forms of the present subjunctive and the imperfect. In other verbs, this additional consonant can be /n/ (e.g. in *prendre* ‘take’ or *venir* ‘come’), /s/ (e.g. in *connaître* ‘know’ or in regular second conjugation verbs like *finir* ‘finish’), /ʁ/ (in *atteindre* ‘attain’), /j/ (in e.g. *broyer* ‘crush’), /v/ (in *écrire* ‘write’ or *boire* ‘drink’), and the shared form can also be longer, such as /ɔlv/ in mildly suppletive alternations like the one found in *résoudre* (ʁe.zɔlv-/ʁe.zu-) ‘solve’.

An analytical option could involve assigning these segments to the stem and positing an invariable underlying stem (e.g. /liz/ or /ekʁiv/) everywhere in the paradigm. In those paradigm cells where the stem surfaces without the final consonant, this would be due to the presence of a subtractive suffix rather than due to an inherently different stem. This synchronic analysis would, indeed, recapitulate to some extent the diachronic origin of these patterns, which are sometimes the result of sound changes from Latin to French which in

some contexts obliterated the last consonant(s) of the stem. The Latin ancestor of *écrire* ‘write’, for example, showed a stem-final consonant /b/ everywhere through the paradigm (Lat. *scrīb-ō scrīb-is scrīb-it scrīb-imus scrīb-itis scrīb-unt*). Its French offshoot, by contrast, shows this consonant, which has become /v/, in certain paradigm cells only (Fr. *ekvi ekvi ekvi ekviv-ō ekviv-e ekviv*). I will not pronounce myself as for the virtues of these and similar analyses but will simply reiterate here my commitment to stick to the presence or absence of overt surface forms throughout this dissertation.

5.1.3 Other requirements

Theoretical notions like ‘basic’ vs ‘derived’, or ‘default’ vs ‘non-default’ have sometimes played a role in the identification of which structures should be regarded as morphomic. However, as one can observe from the following two excerpts, opinions vary in this respect:

The contexts are not reducible to a single dimension of the paradigm, i.e. they cannot be handled through underspecification. In addition, they are not simply the result of the application of defaults. As such, these are morphomic since they cannot be reduced to syntax, semantics or phonology. (Carroll 2016:332-333)

The third stem is no less ‘morphomic’ for being (potentially) definable as a default and the notion of ‘default’ should not blind us to the heterogeneous reality of the forms allegedly bound together by it. (Maiden 2013:495)

In line with earlier sections, I will align with Maiden here in allowing largely no role to theoretical notions like defaults in the definition of what will count as a morpheme in this dissertation. This will be so, first of all, because I want to remain close to the empirical data but also, secondly, because of the lack of consensus in the literature on how to identify defaults in the first place.

That said, the literature on metamorphemes has, indeed, focused overwhelmingly to this date on stem alternants that share some characteristics beyond the ones that have been presented

here so far. It is quite revealing, for example, that the literature has talked about the N-morphome, the L-morphome or about PYTA, but not about the complements of these cells. Consider the following Italian paradigm:

	PRES.IND	PRES.SUB	IPF	PAST	IPF.SUB	FUT	COND
1SG	cuocio	cuocia	cuocevo	cossi	cuocessi	cuocerò	cuocerei
2SG	cuoci	cuocia	cuocevi	cuocesti	cuocessi	cuocerai	cuoceresti
3SG	cuoce	cuocia	cuoceva	cosse	cuocesse	cuocerà	cuocerebbe
1PL	cuociamo	cuociamo	cuocevamo	cuocemmo	cuocessimo	cuoceremo	cuoceremmo
2PL	cuocete	cuociate	cuocevate	cuoceste	cuoceste	cuocerete	cuocereste
3PL	cuociono	cuociano	cuocevano	cossero	cuocessero	cuoceranno	cuocerebbero

Table 131: Non-PYTA root in the Italian verb *cuocere* ‘cook’ (Maiden & Robustelli 2013:226)

The complement cells of many of the most well-known morphemes often qualify as morphemes in their own right according to the criteria that are usually employed for morpheme identification. In this concrete case, for example, the shaded cells contain a stem *cuoc-* (vs *coss-*) whose paradigmatic distribution is also unnatural. Those cells have segments of their own (/w/ and /tʃ/) that are not present outside of them. In addition, the same pattern is repeated in other lexemes with different formal exponents (e.g. *romp-* [vs *rupp-*] in *rompere* ‘break’, *fa-* [vs *fec-*] in *fare* ‘do’, *esprim-* [vs *espress-*] in *esprimere* ‘express’ etc.).

If this were not enough, the shaded set of cells also shows other properties entirely comparable to more traditional morphemes. For example, in the verb *cuocere* above, the stem uniformity of /kwɔtʃ/ within the shaded cells has been achieved by analogical changes that have levelled other formal alternations (wɔ/o, tʃ/k) that were formerly present as the regular product of sound change.⁴⁷ Thus, the reason why complement sets like this are not usually discussed as morphemes of their own is not entirely clear to me but is, I suspect, not unrelated to notions like basic/default.

⁴⁷ The existence of analogical processes that are aimed at preserving or extending a particular pattern could also be thought of as a possible definitional requirement in the identification of morphemes. The evidence most often available to the typologist, however, does not include access to detailed knowledge about the history of most languages, which makes this criterion impractical for a cross-linguistic investigation.

Languages need lexemes, and lexemes need at least one phonological form to exist in a language. Thus, the form *cuoc-*, because it occurs in the vast majority of the cells, would usually be conceived of as merely the form of the lexeme. Thus, only the ‘odd man out’ (i.e. the stem *cross-*) would need to be really “explained” somehow. These concerns may be partially understandable. Because of this, and also in part to allow for some continuity with earlier morphomic literature, a concession will be made on this particular point to those morphologists worried by defaults by not including in this synchronic database of Section 5.2 any converse-type morphemes when this set of cells constitutes a clear majority within the paradigm (over 70% of the cells). Only when two complementary patterns are relatively balanced as for the number of cells that they span (and only if they fulfill the earlier two requirements, of course), will both morphological patterns be included here. This requirement also implies that converse-type morphemes of a single cell will not be included either in the present typological database.

5.1.4 Some excluded morphemes

What these high standards for morphomhood are doing, obviously, is attempting to increase the ‘cleanliness’ of the data at the cost of reducing the number of datapoints in my sample. To have a better idea of what the actual effects of these requirements are, it might be interesting to present in a bit of detail some of those structures that come painfully close to making it into my morphome database but had to be excluded. Consider, for example, the following morphological syncretisms in Binandere (Trans-New-Guinea):

	Future		Far past	
	SG	PL	SG	PL
1EXCL	adu ana	adu ara	adu ema	adu ewa
1INCL	-	adu ana	-	adu ema
2	adu ata	adu awa	adu ata	adu awa
3	adu aina	adu ara	adu ena	adu ewa

Table 132: Partial paradigm of *adu ari* ‘fear’ in Binandere (King 1927:23)

As Table 132 shows, 1SG and 1PL.INCL are always syncretic in the language. The same thing happens with 1PL.EXCL and 3PL. These syncretisms are also implemented with two different formatives in different tenses. Notice how the key shared segments are /m/ and /w/ respectively in the far past but /n/ and /r/ in the future. The cells that syncretize would, in addition, not count as a natural class for most morphologists and typologists. Cross-linguistically, when a 1SG form is syncretic with a plural cell, this is usually either the 1PL as a whole (i.e. both inclusive and exclusive) or only the 1PL.EXCL (see Cysouw 2003:161 and Sauerland & Bobaljik 2013).⁴⁸ This makes sense also semantically, since the 1SG is necessarily exclusive we can hardly be surprised if it syncretizes preferably with the 1PL.EXCL. The Binandere type of conflation seems to be, in fact, typologically unique (Cysouw 2003:95). In addition, this formal identity cannot be obviously handled by defaults either because of the intersecting (and also cross-linguistically very infrequent) syncretism of 1PL.EXCL and 3PL.

Because of the way in which unnaturalness has been defined here, however, neither of the two morphological identities can be included in the cross-linguistic database. In the case of 1PL.EXCL+3PL, the conflation happens between different person values of a single number value 'plural'. This configuration did not qualify here as unmistakably unnatural (see the Livonian example in Table 126). In the case of 1SG+1PL.INCL, the problem concerns feature orthogonality. Because clusivity cannot logically apply to the 1SG, we are missing here the neat feature-value orthogonality that we need to measure morphosyntactic coherence. This lack of logical orthogonality also gives rise to different analytical choices (i.e. treating the inclusive as a distinct value of person or not to do so) which would affect our assessment of a given pattern as natural or unnatural.

Despite not qualifying for morphomehood here, structures like these are still typological rara that are, of course, very interesting for our understanding of morphology. The Binandere system, for example, seems to be a sort of compromise, caught between two different systems in the area and in the phylogenetic tree of the Binanderean family:

⁴⁸ From a sample of 241 languages, 31 show an undifferentiated first person (i.e. 1SG=1PL) and 22 show an inclusive vs exclusive difference with no number distinctions (i.e. 1SG=1PL.EXCL).

	Orokaiva		Suena ⁴⁹	
	SG	PL	SG	PL
1EXCL	-n	-r	-n	-n
1INCL	-	-r	-	-n
2	-e	-v	-s	-w
3	-i	-r	-i	-r

Table 133: Person suffixes of two Binanderean languages (Larsen 1977:23, Wilson 1974:59)

In Binandere's closest relatives Orokaiva and Korafe (but also in unrelated languages of the area like Ömie, see Austing & Upia 1975), a syncretism with /r/-containing exponents aligns the 3PL with the 1PL as a whole. In these languages, thus, the /n/ suffix is confined to the 1SG (see Larsen 1977:23 and Farr 1999:37). In slightly more distantly related languages like Suena, however, the /n/ suffix appears in the whole of first person and /r/ is thus confined to the 3PL (Wilson 1974:59).

Evidence from other Trans-New-Guinea languages seems to point toward Suena presenting the older system and Orokaiva/Korafe being the innovative ones. A system like Binandere's could well represent a transitional state between these two systems. It would be interesting to find out the reasons for the change and its progression in Binandere. Because it does not qualify for morphomhood here, however, I shall leave that to future research.

Other quite clearly morphomic structures that, due to the exclusion of inclusive forms here, do not qualify for inclusion in the database are the following ones:

⁴⁹ Suena also has a dual, and suffixes to distinguish dual from plural. These suffixes are added on top of the ones shown in Table 133. They are not shown here because they are irrelevant to the present discussion.

	Thulung 'drink' (Lahaussois 2002:162)			Ngití 'mother' (Kutsch Lojenga 1994)	
	SG	DU	PL	SG	PL
1EXCL	ɖu-u	ɖu-tsuku	ɖu-ku	íyà-du	íyà-kà
1INCL		ɖu-tsi	ɖuŋ-i		àle- tsá-nà
2	ɖu-na	ɖu-tsi	ɖu-ni	íyà-nɛ	íyà-kɛ
3	ɖuŋ-y	ɖu-tsi	ɖu-mi	kà- tsá-nà	abádhí- tsá-nà

Table 134: Two morphemes that involve the 1PL inclusive

In the case of Thulung (Tibeto-Burman), a longer /ŋ/-final stem is used in 3SG and 1PL.INCL. In other verbs (e.g. *lwa-mu* 'see') the added segment is /s/ instead of /ŋ/. In the case of Ngití (Sudanic), stem suppletion (stem in bold) and suffixation both follow the same unnatural pattern whereby 3 shares its form with 1PL.INCL.

As Table 134 above shows, these morphological affinities in Thulung and Ngití rely on the 1PL.INCL cell for morphomehood. The exclusion of that cell would leave the patterns as morphosyntactically natural and this is the reason why they have been excluded from the present morpheme database. Note, however, that morphemes will not be excluded if they include a/the first inclusive cell but remain morphomic after the exclusion of this cell (see e.g. the morphemes of Bantawa [Section 5.2.6], and Kele [Section 5.2.31]). In these cases, the 1PL.INCL cell(s) will only be excluded in the assessment of the pattern's morphosyntactic coherence (see Section 5.4.9).

It must be clarified that the orthogonality of features and values that concerns us here is predicated on logical grounds over semantic values. Thus, for example, because speech act role of an individual or group and their quantity are logically independent, I will regard person and number as orthogonal features here. In the vast majority of cases, a particular linguistic description's view on this respect will agree with the one that is adopted here. However, I reserve myself the right to contradict the analysis in a source when this has a motivation clearly at odds with the goal of this dissertation.

In Kariña, for example, and in various other Cariban languages, the morphological affinities holding between the different pronouns and between their associated agreement morphology in the verb are unusual. The system is frequently described along these lines:

	SG	PL
1	aau	-
2	amooro	amoññaaro
3	mojko	mojkaaro
1+2	kümüooro	kümüoññaaro
1+3	na'na	-

Table 135: Kariña pronominal system as described by Mosonyi & Mosonyi (2000:407)

The system sketched above seems to fall short of the orthogonality that in principle characterizes person and number from a logical point of view, as some of the posited person categories only have a singular. Other oddities are also evident. For example, some of the forms that have been classified as singular (1+2 and 1+3) evidently refer to more than one individual. Although in their description they go as far as saying, for example, that “the first person lacks a plural” (Mosonyi & Mosonyi [2000:407], translation mine), this paradigmatic representation is evidently an attempt to reflect the morphological affinities in the language and not the semantic values involved. This is obviously not a convenient *modus operandi* if what we are researching is the relation between morphological and extramorphological structure. In line with the rest of this dissertation, 1+3 will be considered the plural of 1 (the same as 2+3 and 3+3 are considered the plurals of 2 and 3 respectively). Rearranged in the way which is most appropriate here, then, this is the paradigmatic distribution of verbal inflectional formatives in Kariña:

	Present			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	voonaae	konoonaano		voonai	noonai	
1INCL	-	kotoonaano	kotoonaatu	-	kotoonai	kotoonatu
2	moonaae	moonaaatu		moonai	moonatu	
3	konoonaano	konoonaatu		noonai	noonatu	

Table 136: Partial paradigm of ‘cultivate’ (Mosonyi & Mosonyi 2000:425)

As Table 136 shows, the form of the verb used with the 1PL.EXCL is identical to the 3SG. In explaining this puzzling behaviour it must be mentioned that the 1PL.EXCL pronoun *na’na* behaves, syntactically, quite differently from the other pronouns.⁵⁰ This may be a synchronic reflection of a nominal origin, which would explain its morphological affiliation with the 3SG.

Be that as it may, it must be observed that, despite the whole-word syncretism of 3SG and 1PL.EXCL, and despite the abundance of formatives in the paradigm above, no set of cells qualifies for morphohood here. Even if syncretic, 3SG and 1PL.EXCL never share any formative (let alone two as required here) to the exclusion of the rest of the paradigm. Every form in Table 136 (e.g. *ko-*, *n-*, *-no*, *-i*) has a paradigmatic distribution which is unnatural but unparalleled by other forms. Thus, no morphemes can be identified in Kariña with the criteria that have been adopted here.

The last class of structures that will be excluded from the present database involved complements and default forms (i.e. formatives that appear in a majority of the cells). As mentioned before, those morphological identities that represent the complement cells of a more paradigmatically restricted morpheme or of a single cell will not be included in the present morpheme database. Consider the following paradigms:

⁵⁰ For example, for the 1PL.EXCL interpretation to emerge, *na’na* must be overtly present, which is not the case with the rest of the pronouns. Similarly, whereas prepositions usually inflect for person in a single word (e.g. *amaaro* ‘with.2SG’, *miaaro* ‘with.3SG’), nouns and *na’na* simply precede the uninflected preposition (i.e. *Juan maaro* ‘Juan with’, *na’na maaro* ‘1PL.EXCL with’).

	'eat'		'drink'		'save/keep'		'roast'	
	FINAL	NFINAL	FINAL	NFINAL	FINAL	NFINAL	FINAL	NFINAL
PAST	kor	kor	kom	kom	tjər	tjər	tjər	tjər
NPAST	ko	kor	ko	kom	tjə	tjər	ka	tjər

Table 137: Formal alternations of some Gavião (Macro-Je) verbs (Amado 2004)

In the Gavião language, as well as in many other Jê languages (see Amado 2004:100-108), verbal inflectional morphology is structured along the opposition of two stems. Unlike it could be expected, however, the choice of form does not depend on one but various factors/features. Most notable among these is tense (past vs non-past) and position in the sentence (final vs non-final position).⁵¹ One form (the one shaded above, usually labelled 'long form' in the literature) occurs in non-final positions in the sentence regardless of tense and, also in final position when past.

The mapping of this formal distinction in Gavião to morphosyntactic/semantic properties is, therefore, unnatural as defined in this dissertation. In addition, as Table 137 shows, the formal alternations involved are very varied.⁵² However, because these stems could be understood to be the default, i.e. a complement of a single stored frequent cell,⁵³ these patterns have also been excluded from the database, in a concession, as mentioned before, to those morphologists for which blocking might be a concern.

⁵¹ Note that some adverbs appear to be invisible for these purposes.

⁵² One often finds the addition of segments: /r/ (most frequent), /n/, or /m/ in the long form (or alternatively their absence from the short form). Vocalic changes can also occur (e.g. *kwir/kwa* 'hit', *tjəm/tja* 'bite'), as well as consonant changes at various locations within the word (e.g. *pus/puj* 'arrive', *jəmjör/jəmⁿgõr* 'pay', *pemter/amte* 'dream'), all the way to full suppletion (e.g. *tjər/ka* 'roast').

⁵³ Patterns similar to this one, where the most common paradigm cell lacks segments which are present in the rest of the paradigm, are not infrequent. Consider, for instance, the alternations between *mat'* and *mater-*, and between *imja* and *imen-* in Russian. Similar structures are also present in the nominal paradigms of genetically unrelated languages like Pite Saami (*båtsoj* vs *buhtsu-* 'reindeer', *bena* vs *bednag-* 'dog', Wilbur 2014) and Ingush (*jexk* vs *axkar-* 'comb', *juu* vs *aur-* 'awl', *jost* vs *aastar-* 'dust', Nichols 2001:148-149) and most likely descend via sound change from an unremarkable zero-vs-suffixed configuration (see Section 4.1.1.1).

5.2 A cross-linguistic database of morphemes

Morphemes are a very challenging object of analysis for a typologist. On the one hand, the phenomenon is only found, as defined in this dissertation, in a relatively small proportion of natural languages (my rough estimate would put the percentage at around 15%). On the other hand, the very term ‘morpheme’ is relatively recent (1994) and even nowadays not widely known and used by field linguists in their grammatical descriptions. These two factors complicate a quantitative typological approach to the phenomenon because they make it a most arduous task to assemble a sufficient number of morphemes.

The fact that the term is not part of most field linguists’ terminological toolkit prevents us from simply looking for it in grammatical descriptions to find examples. Thus, one must read most often through all the morphology and inflection-related sections of a grammar to find out whether the language in question has or lacks morphemes. The relative rarity of the phenomenon, obviously, will mean that we will have to go through quite a few grammars before we find an example which deserves to be included in this database according to the criteria we set up beforehand and which were presented in the previous section.

Because the main problem with morphemes is, thus, the scarcity of data, language sampling is particularly tricky. A “probability sample” (Bakker 2011), thus, seems inadequate for our present purposes. Because of this, the percentage of around 15% that I mentioned before is everything I will have to offer in that sense. It goes without saying that highly isolating or highly agglutinative languages will lack morphemes more frequently than the cross-linguistic average whereas highly fusional, morphologically complex languages will constitute the best breeding ground for morphemes. For this reason, languages and language families with these characteristics will be overrepresented here. The present language sample should be considered thus a “variety sample” (Bakker 2011). Every morpheme has been included in this synchronic database as long as it fulfilled the criteria in Section 5.1. Only cognate morphemes have been excluded when these agreed on their paradigmatic configuration.⁵⁴

⁵⁴ For example, because the Spanish, Portuguese and Italian N-morphemes all have the same paradigmatic extension, only one of them (the Spanish one in this case) has been included in this database.



Figure 4: Geographical location of the languages in the morpheme database

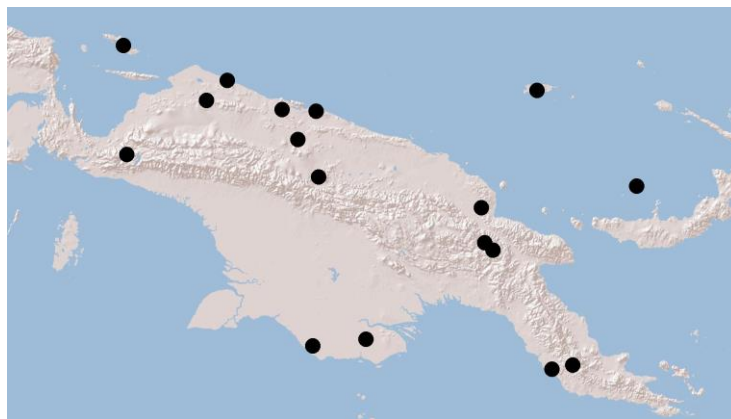


Figure 5: New Guinea zoom-in

Figures 4 and 5 above show the geographical distribution of the languages in this morpheme database. It can be seen that, despite an understandable slight European bias, the sample is by-and-large balanced geographically speaking. All the languages in which morphemes have been found and included in the database are presented below in alphabetical order along with their genetic affiliation and the source(s) employed for each of them:

Language name	Language family	Reference(s)
Achumawi	Palaihnihan	de Angulo & Freeland (1930)
Aguaruna	Chicham	Overall (2007)
Aragonese	Indo-European	Haensch (1958)
Athpariya	Sino-Tibetan	Ebert (1997)
Ayoreo	Zamucoan	Ciucci (2016)
Bantawa	Sino-Tibetan	Doornenbal (2009)
Barai	Koiarian	Olson (1973)
Basque	Isolate	Personal knowledge
Benabena	Trans-New-Guinea	Young (1964)
Biak	Austronesian	van den Heuvel (2006)
Burmeso	West Papuan	Donohue (2001)
Burushaski	Isolate	Yoshioka (2012)
Chinantec L.	Oto-Manguean	Rupp (1996)
Chinantec, P.	Oto-Manguean	Merrifield (1965)
Daasanach	Afro-Asiatic	Tosco (2001)
Daju, M.	Dajuic	Avilés (2008)
Darma	Sino-Tibetan	Willis (2007)
Ekari	Trans-New-Guinea	Drabbe (1952)
English	Indo-European	Personal knowledge
French	Indo-European	Esher (2015)
Fur	Furan	Waag (2010)
Girawa	Trans-New Guinea	Gasaway & Sims (1977)
Greek	Indo-European	Holton et al. (2012)
Icelandic	Indo-European	Jörg (1989)
Iraqw	Afro-Asiatic	Mous (1992)
Irish	Indo-European	Doyle (2001)
Italian	Indo-European	Maiden & Robustelli (2014)
Jabuti	Macro-Je	Pires (1992)
Jerung	Sino-Tibetan	Opgenort (2005)
Karamojong	Nilotic	Novelli (1985)
Kele	Austronesian	Ross (2002)

Ket	Yeniseian	Georg (2007)
Khaling	Sino-Tibetan	Jacques et al (2011), Jacques (2017)
Khinalug	Nakh-Daghestanian	Kibrik (1994)
Koasati	Muskogean	Kimball (1985)
Koiari	Koiarian	Dutton (1996)
Kosena	Trans-New-Guinea	Marks (1974)
Luxembourgish	Indo-European	Schanen (2004)
Majjiki	Tucanoan	Velie & Velie (1981)
Malinaltepec M.	Oto-Manguean	Suárez (1983)
Mazatec	Oto-Manguean	Jamieson (1988)
Mehri	Afro-Asiatic	Rubin (2010)
Menggwa Dla	Senagi	de Sousa (2006)
Mian	Trans-New-Guinea	Fedden (2011)
Murrinh-Patha	Southern Daly	Walsch (1976)
Ngkolmpu	Yam	Carroll (2016)
Nen	Yam	Evans (2015)
Nimboran	Nimboranic	Anceaux (1965), Inkelas (1993)
Nivkh	Isolate	Gruzdeva (1998), Nedjalkov & Otaina (2013)
North Saami	Uralic	Hansson (2007)
Northern Akhvakh	Nakh-Daghestanian	Creissels (2008)
Old English	Indo-European	Cassidy et al. (1971)
Pite Saami	Uralic	Wilbur (2014)
Páez	Páez	Jung (1989)
Skolt Saami	Uralic	Feist (2015)
Sobei	Austronesian	Sterner (1987)
Spanish	Indo-European	Personal knowledge
Sunwar	Sino-Tibetan	Borchers (2008)
Svan	Kartvelian	Tuite (1995)
Thulung	Sino-Tibetan	Lahaussois (2002)
Tol	Jicaquean	Dennis (1992), Holt (1999)
Turkana	Nilotic	Dimmendaal (1991)
Twi	Niger-Congo	Paster (2010), Stump (2015)

Udmurt	Uralic	Csúcs (1988)
Vitu	Austronesian	van den Berg & Bachet (2006)
Vurës	Austronesian	Malau (2016)
Wambisa	Chicham	Peña (2016)
Wutung	Sko	Marmion (2010)
Yagaria	Trans-New-Guinea	Haiman (1980), Stump (2015)
Yorno-So	Dogon	Heath (2014)
Zapotec Y.	Oto-Manguean	Butler (1980)

There is a total of 74 languages included in this database. In addition, as shown above, the genetic variety of the sample is also considerable, with a total of 30 highest-level stocks represented. In many of these languages (23, 31%), more than one structure qualified as morphomic and has been included here. This percentage is substantially higher than the overall cross-linguistic prevalence of morphemes (estimated at 15%), which means that these structures are, unsurprisingly, unevenly distributed across the world's languages. That is, having one morpheme makes you more likely to have a second one. Be that as it may, the multiple occurrences of the phenomenon in some of the languages in this database brings the total number of morphemes in this study to 110. The remaining of this (long) section will present a description of the morphemes that make up this database organized by language and in alphabetical order. The findings are found in Sections 5.4 and 5.5 and in the Appendix.

5.2.1 Achumawi (De Angulo & Freeland 1930)

The Achumawi language (Palaihnihan, California) is characterized by complex stem alternation patterns. De Angulo & Freeland (1930) explain that most verbs distinguish three different stems, which they refer to as the “normal”, “amplified” and “collapsed” stems. As one could guess by their names, the amplified and the collapsed stems usually involve an addition and a reduction respectively of phonological material with respect to the normal stem. The different stems are not aligned to TAM or person-number distinctions. The paradigmatic extension of the normal and collapsed stems varies from one verb to another and different verbs can also have different so-called ‘inflectional end-vowels’. Stem alternation and inflectional end-vowel behaviour combine to distinguish a total of six verb classes/conjugations according to De Angulo & Freeland.

Within any given verb and across conjugations stem alternation follows the same pattern in the indicative, subordinate and optative moods. In the volitional mood (present and future), the distribution of the stem alternants varies from one verb class to another. The so-called amplified stem, which appears in the indicative, subordinate and optative moods is the most relevant for the purposes of the morpheme because it is the only one that remains distributionally stable across moods and conjugations. Consider the following paradigm:⁵⁵

	Indicative			Volitional		
	SG	DU	PL			
1	s- <i>ǎ</i> :n-á	h- <i>ǎ</i> :n-á	h- <i>únn</i> -î:-má	l- <i>ú</i> :n-à	lh- <i>ú</i> :n-à	lh- <i>ú</i> :n-í:-dzà
2	k- <i>ǎ</i> :n-á	gèdz- <i>ǎ</i> :n-á	gèdz- <i>únn</i> -î:-má	t- <i>únn</i> -ô	dz- <i>únn</i> -í	dz- <i>únn</i> -ô
3	y- <i>ǎ</i> :n-á	éiy- <i>ǎ</i> :n-á	y- <i>ǎ</i> :n-íú	tsil- <i>ú</i> :n-à	tsind- <i>ú</i> :n-à	tsind- <i>ú</i> :n-í:-dzà

Table 138: Achumawi verb ‘come’ (De Angulo & Freeland 1930:110)

As Table 138 shows, the amplified stem *ǎ:n* appears in the SG, DU and 3PL forms of the indicative (and in the same cells in the subordinate and the optative). The normal stem *únn*

⁵⁵ Cumulative forms (1>2, 3>2 etc.) have been left out in an executive decision to focus on the orthogonal portions of the paradigm. The 2>1 form sometimes uses the amplified stem.

and the collapsed stem *ú:n* appear elsewhere in the paradigm. Although this is not so for every verb (e.g. in ‘come’ above), the formal alternations between the normal and the amplified stem involve usually, as was mentioned before, additional phonological material in the amplified stem. Consider, for example, the paradigm below:

	SG	DU	PL
1	s-ánwàkà:d-í	h-ánwàkà:d-í	h-á:ká:d-ù-má
2	k-ánwàkà:d-í	gêdz-ánwàkà:d-í	gêdz-á:ká:d-ù-má
3	y-ánwàkà:d-í	yeí-ánwàkà:d-í	y-ánwàkà:d-íú

Table 139: Achumawi verb ‘cut’ Indicative (De Angulo & Freeland 1930:98)

The forms that may be present in the amplified stem but absent elsewhere are very diverse. As shown in Tables 138 and 139, they may involve changes in pitch, in vowel and consonant length, in vowel quality, the infixation of segments or whole syllables etc. A look at the verbs provided by De Angulo & Freeland (1930) reveals the following possible segmental exponents for the amplified stem: *iwa, wa, o: a, a:, ʔ, owʔ, ow, uw, na, awa, eCa, nwa, n, e:, e*. The robustness of the morpheme is, therefore, as the forms and paradigms above illustrate, considerable.

5.2.2 Aguaruna (Overall 2007)

In the possessive inflection of Aguaruna nouns (also in related Achuar, see Fast & Fast 1981:60), the third person and the first person plural behave as a single morphological class and are always syncretic. Consider the paradigm of the following nouns:

	numpa ‘blood’		susu ‘beard’	
	SG	PL	SG	PL
1	numpa-hu	numpĩ	susu-hu	susu-hĩ
2	numpi-mi	numpi-mi	susu-humi	susu-humi
3	numpĩ	numpĩ	susu-hĩ	susu-hĩ

Table 140: Possessive inflection in Aguaruna (Overall 2007:200-202)

The above Table 140 shows the two main classes of nouns in Aguaruna according to the morphological expression of the possessor. The same classes are also found in related Chicham languages (see Table 142). Small irregularities occur in some nouns (see Table 141), due to sound changes or haplogogies. When this happens, however, the whole-word syncretism of 3+1PL is always preserved:

	<i>yatsu</i> 'brother (of a female)'		<i>yawaã</i> 'dog'		<i>uwiha</i> 'hand'	
	SG	PL	SG	PL	SG	PL
1	yatsu-hu	yatʃĩ	yawaã-hu	yawayĩ	uwi-hu	uwihĩ
2	yatsu-mi	yatsu-mi	yawai-mi	yawai-mi	uwi-humi	uwi-humi
3	yatʃĩ	yatʃĩ	yawayĩ	yawayĩ	uwihĩ	uwihĩ

Table 141: Possessive inflection of three irregular Aguaruna nouns (Overall 2007:200-202)

It might be interesting, in contextualizing the present morphomic pattern, to mention that in other Chicham languages (e.g. in Wambisa [Peña 2016:467] and in Shuar [Saad 2014:49]) this syncretism includes also the 2PL. Consider the following paradigms:

	<i>muuka</i> 'head'		<i>nauantu</i> 'daughter'	
	SG	PL	SG	PL
1	muuka-ru	muukĩ	nauantu-ru	nauantu-rĩ
2	muuki-mi	muukĩ	nauantu-rumi	nauantu-rĩ
3	muukĩ	muukĩ	nauantu-rĩ	nauantu-rĩ

Table 142: Possessive inflection of two Wambisa nouns (Peña 2016:467)

Notice how, besides this difference in the 2PL, the inflectional forms in Wambisa (Table 142) are completely parallel to the ones in Aguaruna (see Table 140), which results, of course, from their cognancy. It might be interesting to speculate here about which of the two patterns may represent the original paradigmatic distribution. Wambisa, Shuar and Achuar are sometimes thought to constitute a genetic unit (Shuaric) that branched only after Aguaruna had already

split from the rest of the family. Thus, Shuaric languages have some traits in common (e.g. in these paradigms, possessive suffixes start with /r/ (e.g. *-ru*, *-rumi* etc.) where Aguaruna has them in /h/ (e.g. *-hu*, *-humí* etc.) which may be derived from a closer genetic relationship.⁵⁶

Given our current knowledge of Chicham, both an extension of a 2SG form to the 2PL and a leveling of the plural forms might seem plausible diachronic developments. However, as I have argued before (Table 94), I consider it somewhat more likely that the Aguaruna/Achuar syncretism (i.e. 1PL+3) represents the original one. This is supported by the presence of this pattern in both of the deepest-level branches of Chicham and by the fact that the 2SG and 2PL pronouns both have the formative *-mi* all throughout Chicham.

5.2.3 Aragonese (Barcos 2007, Saura Rami 2003, Haensch 1954)

Local varieties of Aragonese differ as for the synchronic distribution in the verbal paradigm of the reflexes of the N-morpheme (e.g. diphthongization). The most conservative of them have those stem alternants in the cells where the alternation emerged in the first place:

	Indicative		Subjunctive	
	SG	PL	SG	PL
1	'bjengo	be'nimos	'bjengaj	ben'gamos
2	'bjen(e)s	be'niθ	'bjengas	ben'gaθ
3	'bjene	'bjenen	'bjenga	'bjengan

Table 143: Ansotano Aragonese 'have', present (Barcos 2007)

⁵⁶ This closer phylogenetic affinity of Wambisa, Shuar and Achuar has come to be questioned recently by Peña (2016:7). He sees only shared retentions between these languages (like, for example, this particular trait, because, apparently, */r/ change to /h/ in Aguaruna and simply stayed as /r/ in the other languages) and no shared innovations whatsoever.

In these varieties, the N-morpheme appears, as expected, in those cells that were rhizotonic in Latin, that is, in the SG and the 3PL of the present tense in both indicative and subjunctive. These cells continue to have stress on the root in varieties like the one of Ansotano. This paradigmatic configuration of diphthongization (i.e. /je/ vs /e/ like in Table 143, or /we/ vs /o/ in other verbs) stays the same across verbs even in the presence of another present-tense morpheme, the L-morpheme, whose exponent in this verb is the /g/ that appears in the subjunctive and 1SG indicative cells. In other varieties, however, the paradigmatic domain of diphthongization depends on the presence of this other morpheme:

	'sleep'				'twist'			
	Indicative		Subjunctive		Indicative		Subjunctive	
	SG	PL	SG	PL	SG	PL	SG	PL
1	dwérmo	dormím	dwérma	dormám	twérsko	torsém	twérska	twerskám
2	dwérmes	dormíts	dwérmas	dormáts	twérses	torséts	twérskas	twerskáts
3	dwérme	dwérmen	dwérma	dwérman	twérse	twérsen	twérska	twérskan

Table 144: Two Alta Ribagorza Aragonese verbs (Haensch 1958)

In Alta Ribagorza Aragonese, diphthongization has preserved its inherited distribution in those verbs where only the N-morpheme is found (e.g. in 'sleep' in Table 144 above) but has a different distribution in those verbs where the L-morpheme also occurs (e.g. in 'twist' above). Observe how, in the latter verb, diphthongization extends its domain to the 1PL and 2PL cells of the present subjunctive as well. These analogical changes in the paradigmatic configuration of the N-morpheme in this variety must be, therefore, the result of interaction/interference with the L-morpheme. The change could be motivated by a tendency to reduce the total number of stem alternants within a verb by making one of the two morphemes a subset of the other (see Herce 2019).

To complete the picture of the N-morpheme-related variation in the language, it must be mentioned that, in the most innovative varieties of Aragonese, the domain of the N-morpheme has changed in all verbs, even in those without an overt L-morpheme:

	Indicative		Subjunctive	
	SG	PL	SG	PL
1	dwérmo	dormém	dwérme	dwermám
2	dwérmes	dorméts	dwérmás	dwermáts
3	dwérme	dwérmen	dwérma	dwérman

Table 145: Benasque Aragonese ‘sleep’, present (Saura Rami 2003)

It is worth to point out that, due to the analogical innovations presented here, the forms associated with the N-morphemes do no longer correlate to rhizotony in these varieties.

Be that as it may, the values provided in the Appendix for Aragonese correspond to those of the morphomic diphthongization in Alta Ribagorza Aragonese ‘twist’. The diphthongization patterns that have the same paradigmatic distribution as the L- or N-morphemes in Spanish have not been included in the present database due to their cognancy to these.

5.2.4 Athpariya (Ebert 1997)

In the verbal inflection of Athpariya (Kiranti, Tibeto-Burman), 2SG, 3SG and 3PL are characterized by the same suffixal exponence. In the past and the perfect, this affinity is a mere consequence of the fact that these values lack the overt exponents of other cells. In the non-past, however, there are overt suffixes, which are shared by these cells to the exclusion of others. The suffix used varies from intransitive to transitive verbs:

	SG	DU	PL
1EXCL	khat-naʔa	khat-ciciŋa	khad-itiŋa
1INCL		khat-cici	khad-iti
2	a-khat-yuk	a-khat-cici	a-khad-iti
3	khat-yuk	khat-cici	u-khat-yuk

Table 146: Athpariya ‘go’, intransitive positive non-past (Ebert 1997:163)

	SG	DU	PL
1EXCL	lems-uŋtuŋ ⁵⁷	lem-cucuŋa	lems-umtumma
1INCL		lem-cucu	lems-umtum
2	a-lems-utu	a-lem-cucu	a-lems-umtum
3	lems-utu	lem-cucu	o-lems-utu

Table 147: Athpariya ‘beat’, transitive positive non-past, 3SG object (Ebert 1997:180)

Interestingly, this suffixal syncretism of 2SG, 3SG and 3PL is also found, albeit with completely different exponents (*-no* and *-oko*, see Tables 120 and 121), in the closely-related language Chintang, which suggests that we are not dealing with an accidental but a systematic morphological identity.

As Schackow (2016:230-231) explains (see also Section 4.1.5), some of these suffixes go back ultimately to verbs which grammaticalized into the so-called ‘tense markers’ we find synchronically. Athpariya *-yuk*, for example, is believed to be derived from the verb *yuŋ*, which meant ‘be’ or ‘stay’. That this verb grammaticalized into an inflectional formative in the 2/3SG and in the 3PL only must be related to the fact that those cells may have lacked suffixes originally (zeroes can still be found there in other East Kiranti languages like Puma, Limbu, and Bantawa, see Section 5.2.6). Be that as it may, the set of values where these formatives appear synchronically does not constitute a natural class and this counts, therefore, as a morpheme for our present purposes.

⁵⁷ Suffixes like *-uŋtuŋ*, *-utu* and *-umtum* are described by Ebert (1997:180) as involving a non-past formative *-t-* and a copy of the preceding person suffix or string.

5.2.5 Ayoreo (Ciucci 2016, Ciucci & Bertinetto 2017)

In the inflectional exponents of Ayoreo (Zamucoan, Bolivia), some verbs are characterized by a morphological affinity of SG and 3PL. In these contexts, many verbs have a longer stem; one which contains additional phonological material which is not present in the 1PL and 2PL forms. Consider the following illustrative paradigms:

	'want'		'fill up'		'deserve'	
	SG	PL	SG	PL	SG	PL
1	ji-pota	ji-pota-go	ɲĩ-rate	ɲĩ-ra-ko	ji-tiogara	ji-tio-ho
2	ba-pota	waka-pota-jo	mã-rate	wakã-ra-tɔo	ba-tiogara	waka-tio-tɔo
3	pota	pota	tɕĩ-rate	tɕĩ-rate	tiogara	tiogara

Table 148: Paradigm of 3 Ayoreo verbs (Ciucci 2016:105-107)

Most often it is a syllable that is seemingly elided in the 1PL and 2PL (i.e. before the suffixed forms). By virtue of this behaviour, these are referred to as 'mobile syllables' in the literature. These may be of various shapes: *-k(e)*, *-da*, *-go*, *-gu*, *-ni*, *-s(e)*, *-t(e)* elide always; *-di*, *-ga*, *-gi*, *-ŋa*, *-ŋo*, *-ŋu*, *-na*, *-no*, *-ra*, *-re*, *-ri*, *-ro*, *-ru*, *-sa*, *-si*, *-su*, *-so* may elide or not depending on the verb (Ciucci & Bertinetto 2017:34,35).

As Table 148 above shows, the allomorph selection in the 1PL and 2PL suffixes correlates to whether a syllable has been elided or not. As explained by Ciucci & Bertinetto (2017), this allomorphy is a by-product of the diachronic origin of the system. The suffixes must have been originally invariant (i.e. 1PL **-ko* and 2PL **-jo*) as must have been the stems. At some stage, word-internal elisions must have taken place in the suffixed forms whereas the unsuffixed forms remained unchanged. Later sound changes would have made the final segment(s) of the stem and the first consonant of the suffixes coalesce into a single segment that would have been analyzed as part of the suffix.

The changes that gave rise to the system would thus look something like this: 1PL **ɲĩ-rate-ko* > **ɲĩ-rat-ko* > *ɲĩ-ra-ko*, 2PL **wakã-rate-jo* > **wakã-rat-jo* > *wakã-ra-tɔo*. The resulting

allomorphy in the suffixes must have been reanalyzed by language users as a cue for the stem-final syllable deletion and thus it was spread to other verbs to become almost coextensive to it.⁵⁸

5.2.6 Bantawa (Doornenbal 2009)

A trademark feature of Kiranti languages (see also Athpariya in Section 5.2.4) is that they display stem alternation in the verb. In East Kiranti, stem alternation is correlated with the presence of consonant- or vowel-initial suffixes after the stem (Hence forthcoming-a). Consider the following paradigm:

	SG	DU	PL
1EXCL	kon-ŋa	kon-ca	kol-inka
1INCL		kon-ci	kol-in
2	ti-kon	ti-kon-ci	ti-kol-in
3	kon	kon-ci	mi-kon

Table 149: Paradigm of Bantawa ‘walk’ non-past (Doornenbal 2009:391).

A stem alternant (*kon-*) appears in the SG, DU and 3PL (i.e. those word forms where the stem occurs before a consonant or at the end of the word) and another one (*kol-*) appears in the 1PL and 2PL (i.e. when the stem appears before a vowel). The forms involved in these stem alternations are varied. Along with *l/n* we have *r/n*, *y/n*, *ʔ/n*, *r/t*, *ʔ/k*, *w/p* and *ʔ/p*. Some other times, the prevocalic stem is characterized by a segment which is absent from the preconsonantal stem. This can be *s*, *t*, *w*, *y*, and *ʔ*.

⁵⁸ Some mobile-syllable-related allomorphy remains in the suffixes. For example, if a velar is elided, the 1PL is *-ho* rather than *-ko*, if a syllable with */s/* is elided, the 2PL is *-so* rather than *-tso*. Isolated cases also exist where two syllables are elided (see ‘deserve’ in Table), and of the use of suffixes *-ko* and *-tso* in the absence of stem elisions (e.g. 1SG *ji-garu*, 1PL *ji-garu-ko* ‘to tie, to fasten’, Ciucci & Bertinetto 2017:34,35).

The state of affairs described so far holds in the non-past tense. In the past, all the suffixes are vowel-initial and therefore only the pre-vocalic stem alternant (e.g. *kol-*) appears in this tense. However, there is in this domain, interestingly, another form (the suffix *-a*) which has the exact same paradigmatic configuration as stem alternation does in the present:

	SG	DU	PL
1EXCL	kol-a-ŋ	kol-a-cia	kol-inka
1INCL		kol-a-ci	kol-in
2	tɨ-kol-a	tɨ-kol-a-ci	tɨ-kol-in
3	kol-a	kol-a-ci	mɨ-kol-a

Table 150: Paradigm of Bantawa ‘walk’, past (Doornenbal 2009:391).

Although such reasoning would have problems of and by itself (see Section 3.3), because of its coextensivity with coherent phonological environments (i.e. $_V$ vs $_C$), one could argue that the stem alternation in Table 149 is phonologically conditioned and thus not morphomic. However, because in this case the same distribution is replicated with a different formative in the past, phonological determination cannot be maintained and this morphological structure classifies as morphomic here.

This situation, (i.e. the system illustrated in Tables 149 and 150) is what is found in the inflection of intransitive verbs. However, the exact same formal contrasts are found, albeit with a different paradigmatic configuration, in transitive verbs:

	SG	DU	PL
1EXCL	k ^h att-u-ŋ	k ^h at-cuʔa	k ^h att-u-mka
1INCL		k ^h at-cu	k ^h att-u-m
2	tɨ-k ^h att-u	tɨ-k ^h at-cu	tɨ-k ^h att-u-m
3	k ^h att-u	ɨ-k ^h at-cu	ɨ-k ^h at

Table 151: Paradigm of Bantawa ‘take’ non-past (Doornenbal 2009:397)

	SG	DU	PL
1EXCL	k ^h att-u-ŋ	k ^h att-a-cuʔa	k ^h att-u-mka
1INCL		k ^h att-a-cu	k ^h att-u-m
2	ti-k ^h att-u	ti-k ^h att-a-cu	ti-k ^h att-u-m
3	k ^h att-u	ɨ-k ^h att-a-cu	ɨ-k ^h att-a

Table 152: Paradigm of Bantawa ‘take’ past (Doornenbal 2009:398)

This suggests that the identical paradigmatic distribution of the pre-consonantal stem in the present and the *-a* suffix in the past is not coincidental. One can also observe, in Tables 149 through 152, that an alternation between zero and *-a* indicates tense in those (darker shaded) paradigm cells where those forms appear while the rest (marked with *-i* in intransitives and with *-u* in transitives) does not make tense distinctions.⁵⁹

Given the criteria that are being used in the present dissertation, three different morphemes can be identified in Bantawa: SG/DU/3PL in intransitive verbs, SG/1PL/2PL in transitive verbs, and DU/3PL in transitive verbs. All of these cells constitute unmistakably unnatural classes and can be characterized by forms not present in the other cells of the paradigm (*-n/-t/-k/-p* and *-a* in the first and third, *-s/-t/-w/-y/-ʔ/-l/-r* and *-u* in the second).

According to the numbers provided by Doornenball (2009:134), stem alternation is present in around 92% of the lexemes in Bantawa, although only 16.6% will have (like *kon-* in Table) forms exclusive to the preconsonantal stem. This is because most stem alternations are based on ‘augments’ that are present in the prevocalic stem but absent elsewhere (e.g. *k^hatt-* vs *k^hat-*). This refers exclusively, of course, to stem alternation, since the past tense suffix *-a* and the suffix *-u* appear in every single lexical item.

Bantawa1: SG/DU/3PL

Batawa2: SG/1PL/2PL

Bantawa3: DU/3PL

⁵⁹ Notice how the realization of the morphosemantic feature of tense appears to be dependent on (or ‘nested into’, following the formulation used by Corbett 2016) a morphomic set of cells. The same happens in other languages and morphemes (see e.g. the distinction between present and progressive in Daju in Table 48).

2.5.7 Barai (Olson 1973)

Although agreement in the Barai (Koiarian, New Guinea) verb takes place most robustly with the object, some verbal formatives in the language have different allomorphs depending on the person-number of the verb's subject. The morphosyntactic distribution of this allomorphy, however, is morphosyntactically unmotivated, with 1SG and PL being characterized by an allomorph different from the one in 2SG/3SG. Consider the following:

	Past sequence 1		Past sequence 2 ⁶⁰		Future sequence		Delayed past seq.	
	SG	PL	SG	PL	SG	PL	SG	PL
1	-jo	-jo	-vo	-vo	-kuva	-kuva	-eva	-eva
2	-no	-jo	-mo	-vo	-kuma	-kuva	-ema	-eva
3	-no	-jo	-mo	-vo	-kuma	-kuva	-ema	-eva

Table 153: Allomorphy of some Barai suffixes (Olson 1973:48, 53,56)

The pattern of syncretism 1SG+PL is found in various different suffixes, although the actual alternating forms are always just two: /j/ (vs /n/), and /β/ (vs /m/). Although some analogical convergence of 1SG with plural may also have played a role (cf. closely-related Managalasi, where 1SG has sometimes an allomorph different from PL, see Parlier 1964:3), these forms seem to go back ultimately to zero. That is, at some stage, 2SG and 3SG would have been characterized by an /m/ exponent opposed to its absence from the rest of the paradigm.⁶¹ Glides would have been subsequently introduced to break vowel-vowel sequences (e.g. *-kua > *-kuwa > -kuβa). The nature of the glide (i.e. /w/ or /j/) would have depended on stress and the quality of the previous vowel (see Footnote 61).

⁶⁰ The form /jo/ is found in verbs that end in a stressed front vowel and /βo/ is found elsewhere. Note that orthographic 'v' represents, thus /β/.

⁶¹ It is interesting to note that, in related Koiari (see Section 5.2.36), /m/ appears in 1SG and 3SG and is absent from the rest of the paradigm. In related Koita (Dutton 1975), this seemingly cognate /m/ appears in all singular cells. The history of this formative seems, thus, interesting but is unclear to me at the moment.

5.2.8 Basque (personal knowledge)

The verbal inflection of Basque is mainly agglutinative and relies for the vast majority of verbs in the use of auxiliaries that bear the agreement markers. In a few high-frequency synthetic verbs, however, there are some forms which appear, in the standard language, in the PL and the 2SG forms. Consider the following paradigms:

	<i>etorri</i> 'come', present		<i>ibili</i> 'walk', past		<i>egon</i> 'be', present	
	SG	PL	SG	PL	SG	PL
1	na-tor	ga-to-z	nen-bil-en	gen-bil-tza-n	na-go	ga-u-de
2	za-to-z	za-to-z-te	zen-bil-tza-n	zen-bil-tza-te-n	za-u-de	za-u-de-te
3	da-tor	da-to-z	ze-bil-en	ze-bil-tza-n	da-go	da-u-de

Table 154: Partial paradigms of three Basque verbs.

Forms like *-z*, *-tza* and *-u-de* occur in all synthetic tenses of the verb (cf. present *za-u-de* vs past *ze-u-n-de-n*). These formatives appear in this unnatural set of cells PL+2SG in the modern language but are believed to have been straightforward markers of plurality at an earlier stage in the language. Thus, these formatives also appear in the language as markers of plurality (or rather, of 2SG+PL object) in transitive verbs (cf. *daramat* 'take.1SG>3SG' vs *darama-tza-t* 'take.1SG>3PL').

This pattern can be straightforwardly explained diachronically. In the same way as in the languages that surround it (i.e. Spanish and French, but also English or Russian), the 2PL form in Basque came to be used for polite reference to a 2SG addressee. The earlier 2SG forms (e.g. *ha-tor* 'come.2SG') were, thus, reserved for familiar address. In due time, unlike in English or French, a new 2PL pronoun and a new 2PL verbal form were innovated by adding a pluralizer (*-ek* in the pronoun, *-te* in the verb) to forms which would have ceased to be perceived as plural. Thus, in contemporary standard Basque, forms like *za-toz* can only be referentially singular but still behave morphologically the same as the plural forms. Subsequently, some sound changes (e.g. **za-tor-z* > *zatoz*, **za-go-de* > *zaude*) have blurred the distinction between stem and affix in a way that it is difficult to say whether we are synchronically dealing with stem alternants or affixes.

5.2.9 Benabena (Young 1964)

For the purposes of stem alternation, SG+1 subject agreement forms constitute a single morphological class in Benabena (Trans-New Guinea) and in related Gorokan languages (see e.g. Yagaria in Table 79). Consider the following paradigm:

	'hit'			'go'		
	SG	DU	PL	SG	DU	PL
1	ho-ʔohube	ho-ʔohuʔibe	ho-ʔohune	bu-ʔohube	bu-ʔohuʔibe	bu-ʔohune
2	ho-ʔahane	he-ʔehaʔibe	he-ʔehabe	bu-ʔahane	bi-ʔehaʔibe	bi-ʔehabe
3	ho-ʔehibe	he-ʔehaʔibe	he-ʔehabe	bu-ʔehibe	bi-ʔehaʔibe	bi-ʔehabe

Table 155: Two verbs in Benabena, past tense (Young 1964:50)

These same formal alternations can also be found in some inflectional affixes like the progressive *no-/ne-* (Young 1964:68). Note that verb compounding is common in the language and that this and other formatives are most likely derived from verbs. Secondary forms of the verb in Benabena are often based on the ones in Table 155. This is not the case, however, of the future tense, which does not show the morphomic affinities described here. Instead it shows, for the verbs above, the stems *ha-* and *bi-* respectively.

5.2.10 Biak (van den Heuvel 2006)

In the inflectional morphology of Biak (Austronesian), both in subject agreement in the verb and in possessor inflection in the noun, there is a set or cells that is characterized by a common form and by common morphophonological properties but which does not constitute a natural class from a semantic perspective. Consider the following paradigm:

	SG	DU	PC/TR	PL
1EXCL	ya-mar	nu-mar	nko-mar	nko-mar
1INCL	-	ku-mar	ko-mar	ko-mar
2	wa-mar	mu-mar	mko-mar	mko-mar
3	i-mar	su-mar	sko-mar	si-mar/na-mar

Table 156: Biak verb 'die' (van den Heuvel 2006:157)

Apart from their shared segments /ko/ (sometimes only /k/), those forms are also peculiar in that, unlike all other suffixes, they lengthen the vowel of vowel-initial stems and, at the end of an intonational unit, they require an epenthetic vowel:

	SG	DU	PC/TR ⁶²	PL
1EXCL	y-an	nuy-an	nk-áne	nk-áne
1INCL	-	kuy-an	k-áne	k-áne
2	w-an	muy-an	mk-áne	mk-áne
3	d-an	suy-an	sk-áne	s-an/n-an

Table 157: Biak verb 'eat' (van den Heuvel 2006:159)

As discussed by van den Heuvel (2006:66), all those forms in *ko-* can be traced back all the way to Proto-Austronesian **telu* 'three' (**t>k* is regular in Biak). This etymology, along with the comparison to closely-related languages (like e.g. Ambai, see Silzer 1983), suggests that the original value of the forms must have been 'trial'. It seems that, in Biak, in the first and second person (but not in the third), the use of these forms spread to denote larger numbers too. The result is that the morphological affinity of the shaded cells is no longer semantically justified.

⁶² Van den Heuvel is not consistent in the glossing of the forms. Sometimes he labels them 'trial' and some other times 'paucal'. It is thus not clear to me what is the precise value of the forms in *sk(o)-*. This, however, does not greatly affect the present analysis.

5.2.11 Burmeso (Donohue 2001)

Verbs in Burmeso (Isolate, New Guinea) agree with a single argument. This will be the direct object in the case of transitives and the subject in the case of (some) intransitives. Even though a given verb can take only one of three different prefixes, their distribution over noun classes and numbers is notoriously complicated:

Class	Conjugation 1		Conjugation 2	
	SG	PL	SG	PL
I Male	j-	s-	b-	t-
II Female, animate	g-	s-	n-	t-
III Miscellaneous	g-	j-	n-	b-
IV Mass nouns	j-	j-	b-	b-
V Banana, sago tree	j-	g-	b-	n-
VI Arrows, coconuts	g-	g-	n-	n-

Table 158: Conjugations in Burmeso (Donohue 2001:100-102)

Excluding the prefixes *s-* and *t-*, for which a coherent meaning (animate plural) can indeed be identified, the distribution of the other prefixes does not make much sense morpho-syntactically. Depending on the noun, all the prefixes may co-occur with the plural but not the singular, with the singular but not the plural, and with both singular and plural. It may be also relevant to point out that, whereas plural pronouns do occur, as expected, with the prefixes *s-* and *t-*, the singular pronouns do not agree with the gender of their referent but have fixed agreement instead. The 1SG pronoun, thus, co-occurs with *g-/n-* (i.e. behaves like female singular nouns) while the 2SG pronoun agrees with the prefixes *j-/b-* (i.e. it behaves like male singular nouns).

The assignment of particular items to each of the two agreement classes appears to be, therefore, completely arbitrary. However, because of the existence of two conjugations, we can see that these morphomic classes are systematic, since the seemingly erratic distribution of *j-* is replicated by *b-*, and that of *g-* is repeated by *n-*.

The absence of genetic relatives of Burmeso makes it difficult to make any judicious proposals as to how the system may have emerged. The pattern, however, does remind a lot of the ones found in Khinalug (see Section 5.2.34), Mian (see Section 5.2.44) and other languages. The allomorphic variation between the prefixes of different conjugations (e.g. *s-* vs *t-*) might plausibly originate from originally invariable prefixes which would have split into two different allomorphs by way of some sound change conditioned by the phonology of the following verb. As for the puzzling distribution of the prefixes, this system seems like could originate from a more unremarkable 2 or 3-gender system that somehow “went wrong”.

Burmeso1: II.SG/III.SG/V.PL/VI

Burmeso2: I.SG/III.PL/IV/V.SG

5.2.12 Burushaski (Yoshioka 2012)

Burushaski distinguishes four genders in nouns, which are indexed in the verb by means of prefixes (undergoer) and suffixes (subject). Syncretisms are common in both paradigms:

	Type I undergoer pref.		Type III undergoer pref.		'come' simple past	
	SG	PL	SG	PL	SG	PL
M	i-	u-	ée-	óo-	díimi	dúuman
F	mu-	u-	móo-	óo-	dumóomo	dúuman
X	i-	u-	ée-	óo-	díimi	dúumio
Y	i-	i-	ée-	ée-	díimi	díimi

Table 159: Some Burushaski partial paradigms (Yoshioda 2012)

While other (partial) syncretisms do not show up in every context, there is a set of cells (M.SG, X.SG, Y.SG, Y.PL) which is particularly salient for the systematicity of its identity. For the sake of full disclosure, it has to be stated that class Y nouns are less compatible with pluralizability because they are often abstract or mass nouns (Yoshioda 2012:33). This was also the case of some of the classes (like IV) in Burmeso.

5.2.13 Chinantec, Lealao (Rupp 1996, Feist & Palancar 2015)

Chinantecan languages, and the Oto-Manguean phylum more generally, are famous for their prominent use of vowel and consonantal changes in the stem in verbal inflection. Lealao Chinantec, for example, is representative of the kinds of alternations that one may find.

Inflectional affixes distinguish a total of 7 person-number values (all 3 persons and 2 number combinations plus a 1PL.INCL). The segmental and suprasegmental alternations in the stems, however, show less formal diversity in that they only distinguish four person-number “stem spaces” (1SG, 1PL, 2 and 3). This consolidation of values suggests that stem alternation in Chinantec is not completely oblivious to feature and value relations. However, despite the morphosyntactic coherence of these confluences, actual formal/segmental alternations within a single verb’s paradigm make usually little sense in morphosyntactic terms and have to be regarded as morphomic. Consider the following paradigms:

	‘grab’						‘listen’					
	Incomplet.		Irrealis		Completive		Incomplet.		Irrealis		Completive	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	sanh	sanh	sanh	xanh	sanh	xanh	nuu	nuu	nuu	niuu	nuu	niuu
2	sanh	sanh	sanh	sanh	xanh	xanh	nuu	nuu	nuu	nuu	niuu	niuu
3	sanh	sanh	sanh	sanh	sanh	sanh	nuu	nuu	nuu	nuu	nuu	nuu

Table 160: Stem alternants in two Lealao Chinantec verbs (Feist & Palancar 2015)

As the paradigms above illustrate, a particular stem alternant, characterized by various palatalizations and vowel raisings, occurs in the 1PL of the irrealis and in the 1PL and 2 of the completive. In some other verbs, this stem alternant (or rather one characterized by the same forms) appears in a superset of these cells. In addition to the contexts shaded in Table 160, the same stem appears in the third person across all aspects, as well as in the 1PL incomplete:

	'pay'						'open'					
	Incomplet.		Irrealis		Completive		Incomplet.		Irrealis		Completive	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	cø	chi	cø	chi	cø	chi	na	nia	na	nia	na	nia
2	cø	cø	cø	cø	chi	chi	na	na	na	na	nia	nia
3	chi	chi	chi	chi	chi	chi	nia	nia	nia	nia	nia	nia

Table 161: Stem alternation in two Lealao Chinantec verbs (Feist & Palancar 2015)

These alternations are also present, with a similar paradigmatic configuration, in a number of other Chinantecan languages (e.g. in Palantla Chinantec, described in the following section) and should be reconstructed for the Proto-language (see Rensch 1989:21-22). They most likely go back to a single segment /j/ which was infix, as in the verb 'open' above, between stem onset and the vowel nucleus. Similar formatives (i.e. inflectional infixes) are not unheard of in Mesoamerica (e.g. in Tol [see Holt 1999], and in distantly-related Northern Pame [see Berthiaume 2004]). The diversity of formal alternations (including, analogically, cases of suppletion) would have emerged in Chinantec from this single formative /j/ by means of later sound changes (e.g. consonant palatalizations and/or vowel fusions and raisings) which involved that yod.

Chinantec, L1: 1PL.IR/1PL.C/2.C

Chinantec, L2: 1PL/2.C/3

5.2.14 Chinantec, Palantla (Merrifield 1965, Feist & Palancar 2015)

The overall morphological system described for Lealao Chinantec in the previous section is by and large valid for Palantla as well. Also the paradigmatic distribution of the inherited stem alternations is very similar in the two varieties. The first of the morphemes that I presented for Lealao Chinantec, thus, differs from the pattern found in Palantla only in a single cell in the paradigm:

	'buy'						'smoke'					
	Progressive		Intentive		Completive		Progressive		Intentive		Completive	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	la	lye	la	lye	la	lye	hĩ	hĩ	hĩ	hĩ	hĩ	hĩ
2	la	la	la	la	lye	lye	hĩ	hĩ	hĩ	hĩ	hĩ	hĩ
3	la	la	la	la	la	la	hĩ	hĩ	hĩ	hĩ	hĩ	hĩ

Table 162: Stem alternation in two Palantla Chinantec verbs (Merrifield 1968:41)

As shown above, the morpheme extends to the 1PL progressive whereas it did not do so in Lealao. Although Rensch (1989:21-22) presented the one in Palantla as the original distribution of the alternation, comparison with other Chinantecan varieties suggests that it might be Lealao which presents the original paradigmatic distribution. The paradigmatic distribution of the alternation in Comaltepec Chinantec (Anderson 1989:7), for example, agrees with the one in Lealao.

If we considered this to be the original paradigmatic extension of the alternation, thus, the small change in Palantla would seem to be aimed at bringing the pattern of stem alternation closer to the language's other morpheme.⁶³ This one has in the language an identical paradigmatic configuration to the one in Lealao and, thus, although it is presented below for the sake of completeness, will not be included in the present morpheme database:

	'fence'						'split'					
	Progressive		Intentive		Completive		Progressive		Intentive		Completive	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	hniw	hne	hniw	hne	hniw	hne	zow	zyow	zow	zyow	zow	zyow
2	hniw	hniw	hniw	hniw	hne	hne	zow	zow	zow	zow	zyow	zyow
3	hne	hne	hne	hne	hne	hne	zyow	zyow	zyow	zyow	zyow	zyow

Table 163: Stem alternation in two Palantla Chinantec verbs (Merrifield 1968:41-42)

⁶³ In Palantla, the paradigmatic distribution of the "larger" morpheme can be stated as: 'smaller morpheme'+3, which could be taken to be a simpler description than the relationship between the two morphemes in Lealao: 'larger morpheme'='smaller morpheme'+3+1PL.PROG.

It is interesting to note how, in both Lealao and Palantla Chinantec, one morpheme constitutes a subset of the other. Something similar is found in the morphemes of Khaling, Saami and Wutung (see Sections 5.2.33, 5.2.54 and 5.2.68). This seems to be a functional trait present in the architecture of various morpheme systems. As discussed by Maiden (2018b:13-14),⁶⁴ subset-superset arrangements of allomorphy like the ones in Chinantec and Saami allow reliable (though asymmetrical) predictions of forms by language users. Thus, for example, the use of stem alternant *hne* in the 3SG (see Table 163) allows speakers to infer the use of the same form in the 1PL. Note that the predictability does not hold in the opposite direction. The use of a particular stem in the 1PL does not inform us of whether or not the same form will be used in the 3SG (notice the difference between the paradigms of ‘fence’ and ‘buy’ in this respect).

5.2.15 Daasanach (Tosco 2001)

As explained before (see Table 46) in the South Cushitic language Daasanach, verbal person-number agreement is structured morphologically in a two-way opposition between a so-called (Tosco 2001) Form A and a Form B. As the labels suggest, the paradigmatic distribution of the two forms is arbitrary from a morphosyntactic perspective. The actual formal alternations involved are also quite diverse:

	‘open’ Perfect		‘die’ Imperfect		‘drink’ Perfect	
	SG	PL	SG	PL	SG	PL
1EX	furi	fuddi	kufuma	kufunanna	fii	?iyyi
1INC		furi		kufuma		fii
2	fuddi	fuddi	kufunanna	kufunanna	?iyyi	?iyyi
3FEM	fuddi	furi	kufunanna	kufuma	?iyyi	fii
3MASC	furi	furi	kufuma	kufuma	fii	fii

Table 165: Partial paradigms of three verbs in Daasanach (Tosco 2001:112,140,172)

⁶⁴ Maiden (2018b:14) also writes that these configurations appear to be “very rare” in Romance (he even has to give an invented example to illustrate them). Judging by the data gathered here, however, there are reasons to believe that this rarity cannot be extrapolated to morphemes as a whole.

As the above paradigms show, 3SG feminine, 2 and 1PL.EXCL are whole-word syncretic and opposed to the form used in 1SG, 1PL.INCL, 3SG masculine and 3PL. This bizarre system originated from one with more-or-less unremarkable person-number agreement affixes (see Table 47) and more conservative Cushitic languages still show the original system.:⁶⁵

Although analogy must also have had a prominent role (see Sasse 1976), many of the formal alternations involved, both at the right and at the left edges of the stem, are straightforward products of run-of-the-mill sound changes that have occurred in the history of the language:

*yeet-e > *yeete > yeeḏe ('say/become' Form A Imperfect, Tosco 2001:201)

*t-yeet-t-e > *tyeette > ceete ('say/become' Form B Imperfect, Tosco 2001:201)

Daasanach1: 1SG/3SG.M/3PL

Daasanach2: 1PL/2/3SG.F

2.5.16 Daju, Mongo (Avilés 2008)

In Daju (Dajuic, Chad), verbal person-number inflection is characterized by a whole-word syncretism of SG and 3PL. This syncretism sometimes obtains merely by the absence of forms present in the rest of the paradigm but often it is also instantiated by an overt formative, which can have different phonological forms depending on tense or verb type:

⁶⁵ See also Iraqw (Section 5.2.25) for an intermediate system, that is, for a system where a form A vs form B stem alternation has emerged but where affixes still disambiguate most of the values that are collapsed in Daasanach.

	'drink' present		'drink' progressive		'hide oneself' present	
	SG	PL	SG	PL	SG	PL
1EXCL	ur-o	ur-ciga	ur-ca	ur-ciga	nol-wa	nol-din-ciga
1INCL		ur-cina		ur-cina		nol-din-cina
2	ur-o	ur-cini	ur-ca	ur-cini	nol-wa	nol-din-cini
3	ur-o	ur-o	ur-ca	ur-ca	nol-wa	nol-wa

Table 166: Some partial paradigms in Mongo Daju (Avilés 2008)⁶⁶

This system (reminiscent of Ayoreo, see Section 5.2.5) appears to have originated from a situation of zero marking in the singular and 3PL opposed to overt markers in 1PL and 2PL cells (see Section 4.1.1.1 on the cross-linguistic tendencies in zero-marking). It is not cross-linguistically uncommon for the third person not to show number distinctions even when the first and second persons do (Cysouw 2003). The idiosyncrasy of this system lies then in the fact that person-number marking is absent both from the singular forms and from the third person cell, thus resulting in an unnatural pattern of syncretism.

Sound changes would have been responsible for the later emergence of overt markers of the class SG+3PL (e.g. consider *wede* SG/3PL vs *wetcina*<**wed(e)-cina* 1PL.INCL 'walk', or *alase* SG/3PL vs *alaffina*<**alas(e)-cina* 1PL.INCL 'throw away once', Avilés 2008:71-72), although analogical processes may have also played a role (e.g. in the case of the reflexive). It is, in any case, remarkable that, within a given tense, SG+3PL are always whole-word syncretic.

5.2.17 Darma (Willis 2007)

In Darma (Sino-Tibetan), verbal agreement is characterized by a syncretism of 1PL and 2. This syncretism holds across tenses, as Table illustrates, and also, with slightly different suffixes (-*de* instead of -*he*), in transitive verbs:

⁶⁶ The 1DU forms *ur-cik* and *nol-din-cik* have not been represented in the paradigm for reasons of space. Note that they pattern like the 1PL/2PL and are thus irrelevant for the purposes of the morphomic pattern at stake.

	Non-past		Past	
	SG	PL	SG	PL
1	ra-hi	ra-he-n	ra-ju	ra-n-su
2	ra-he-n	ra-he-n(i)	ra-n-su	ra-n-su
3	ra-ni	ra-ni	ra-ju	ra-ju

Table 167: Paradigm of Darma *ra* ‘come’ (Willis 2007:350-356)

The formal affinity shaded in Table 167 is, therefore, morphomic. The situation in closely related languages looks quite confusing as to which person-number contrasts are made and how. In closest-related Byangsi (Sharma 2001a), for example, some verbs/tenses show syncretism of 1PL and 2PL, and others of 2SG and 2PL. In related Chaudangsi (Krishan 2001), the present tense has -ni in 2PL and 3SG, and -nɛ in 1PL, 2SG and 3PL, although /n/ is absent from the 3SG in the past. In Rongpo (Sharma 2001b), similarly, various /n/-containing syncretisms exist as well in different tenses:

	Present		Progressive		Past	
	SG	PL	SG	PL	SG	PL
1	rhaŋ	rha-ni	rhacɛ̃ki	rhacɛ-ni	rhaki	rha-n
2	rha-n	rha-ni	rhacɛ̃-ni	rhacɛ̃-ni	rha-n	rha-n
3	rha-n	rha-ni	rhacɛ	rhacɛ-ni	rhɛ	rhɛ̃

Table 168: Rongpo verb *rha-paŋ* ‘come’ in various tenses (Sharma 2001b:226)

As the above Table 168 shows, these may involve i) all plural cells, ii) PL+2SG, iii) 1PL+2 (as in Darma), and 2SG+3SG. The diachronic progression of these forms in West Himalayish is not clear at all to me (although see Saxena 1992). The “mess” observed in the distribution of formatives in related languages is probably derived from the loss of an earlier biargumental agreement system. It might be, although this is largely conjectural, that Darma has managed to put some order in this mess by generalizing a single paradigmatic distribution of /n/ and by organizing the allomorphy of tense markers along the same lines as well.

5.2.18 Ekari (Drabbe 1952, Doble 1987)

In the Ekari language, future tense suffixes display an allomorphic variation whose paradigmatic distribution is morphosyntactically unnatural. Formal variation minimally satisfies the criteria set for morpheme identification here. Consider the paradigms below:

	Hodiernal future			Post-hodiernal future		
	SG	DU	PL	SG	DU	PL
1	uwii-pig-a	awai-pag-e	uwii-pag-e	uwii-t-a	awai-tag-e	uwii-tag-e
2	uwii-pag-e	awai-pig-aa	uwii-pig-aa	uwii-tag-e	awai-t-aa	uwii-t-aa
3M	uwii-pag-i	awai-pig-ai	uwii-pig-ai	uwii-tag-i	awai-t-ai	uwii-t-ai
3F	uwii-pig-a	awai-pig-ai	uwii-pig-ai	uwii-t-a	awai-t-ai	uwii-t-ai

Table 169: Partial paradigm of 'go (Drabbe 1952:49-50, Doble 1987:89)

Because the languages closest related to Ekari are not sufficiently described, it is difficult to make educated guesses about the diachronic origin of these alternations. As Table 169 shows, however, the paradigmatic distribution of the allomorphs coincides with the front (e/i) vs non-front (a) quality of the following person-number agreement suffixes, which may point towards a sound-change-related origin.

Ekari1: 2SG/3SG.M/1DU/1PL

Ekari2: 1SG/2DU/2PL/3DU/3PL/3SG.F

5.2.19 English (personal knowledge)

The English language is notoriously poor in inflectional morphology compared to most other Indo-European languages. However, there are in the language two structures which minimally qualify for a morphomic status according to the criteria set here. The first one is to be found

in the paradigm of the English verb ‘be’, which shows an unnatural pattern of syncretism not found elsewhere in the language but systematic in that verb because it is repeated with more than one exponent:

	Present		Past	
	SG	PL	SG	PL
1	am	are	was	were
2	are	are	were	were
3	is	are	was	were

Table 170: Paradigm of the English verb ‘be’

As is well known, the presence of the form ‘are’ in the 2SG is due to the use of an earlier 2PL form (‘you’) for the 2SG too in the modern language.⁶⁷ Such a change was driven by the common strategy (see also Basque in Section 5.2.8) of signalling politeness by referring to singular addressees with a plural pronoun. The second morphological affinity that classifies as morphomic in English appears in three verbs which have a longer stem in the 1SG, 2SG and PL in the present:

	‘have’				‘do’				‘say’			
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	hæv	hæv	hæd	hæd	du:	du:	dɪd	dɪd	seɪ	seɪ	sed	sed
2	hæv	hæv	hæd	hæd	du:	du:	dɪd	dɪd	seɪ	seɪ	sed	sed
3	hæz	hæv	hæd	hæd	dʌz	du:	dɪd	dɪd	sez	seɪ	sed	sed

Table 171: Finite forms of three English verbs

The emergence of this particular pattern is related to the fact that those cells are the ones where the stem of a verb is not followed by a suffix. This different phonological context has made it possible for sound changes to affect those forms differently from the others. Of

⁶⁷ The presence of ‘were’ in the 2SG.PAST is a somewhat different story in that the form of the stem used with the old 2SG *thou* was already the same as the plural form in Old English. This constitutes a West Germanic trait of uncertain origin, see the section on Old English.

course, these verbs (also ‘be’ before) are among the most frequent in the language, which has made it possible for them to preserve these structures even in an ocean of invariance.

English1: 2SG/PL

English2: 1SG/2SG/PL

5.2.20 French (Esher 2015, Meul 2010)

As mentioned before (see Table 130 and the ensuing discussion) in French inflection verbs vary in the extent to which they show what could be considered their full stem throughout the paradigm. Consider the following paradigm once again:

	PRES.IND		PRES.SUB		IMPERF		FUT		COND	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	li	lizɔ̃	liz	lizjɔ̃	lizɛ	lizjɔ̃	liʁe	liʁɔ̃	liʁɛ	liʁjɔ̃
2	li	lize	liz	lizje	lizɛ	lizje	liʁa	liʁe	liʁɛ	liʁje
3	li	liz	liz	liz	lizɛ	lizɛ	liʁa	liʁɔ̃	liʁɛ	liʁɛ

Table 172: Paradigm of French *lire* ‘read’

The segment /z/ appears at the end of the stem only in the plural forms of the present indicative and in the forms of the present subjunctive and the imperfect. As explained before, the same situation obtains with other segments (/z/, /n/, /s/, /ʁ/, /j/, /v/, and /ɔlv/) in other verbs.

These morphological patterns are the result of sound changes from Latin to French which, in some contexts (but not everywhere) have eliminated the last consonant(s) of the stem. Note, however, that analogical processes have also played a big role in the emergence of this paradigmatic configuration, most clearly when the earlier inchoative infix -esc- adopted this paradigmatic configuration in what is now the second conjugation (see Meul 2010:20).

Fur (Waag 2010)

In the Fur (Furan, Sudan) verbal inflection, there is a morphological affinity of the 3SG and the 3PL non-human, which are opposed to the rest of the paradigm. Consider the following:

	'tie' imperfective		'hang' imperfective		'grind' imperfective	
	SG	PL	SG	PL	SG	PL
1	ʔirg-ɛl ⁶⁸	kirg-ɛl	ʔalg-ɛl	kalg-ɛl	ʔawan	kawan
2	jirg-ɛl	birg-ɛl	jalg-ɛl	balg-ɛl	jawan	bawan
3.HUM	rig-ɛl	kirg-ɛl-ɪ	ɪrg-ɛl	kalg-ɛl-ɪ	kɔɔn	kawne
3.NHUM	rig-ɛl	rig-ɛl-ɪ	ɪrg-ɛl	ɪrg-ɛl-ɪ	kɔɔn	kɔɔne

Table 173: Partial paradigms of three Fur verbs (Waag 2010)

Almost all verbs in Fur show stem alternation according to the above pattern, although some exceptions do exist where the two are identical (e.g. 3SG *rig-ɛl* vs 1SG *ʔa-rig-ɛl* 'lie in waiting', Waag 2010:125). As the above paradigms illustrate, the stem alternations between the two sets of cells (i.e. 3SG+3PL.NHUM vs 1+2+3PL.HUM) are extremely diverse from a formal perspective. In 'tie', for example, we find consonant/vowel metathesis, in 'hang', vowel deletion/epenthesis, and in 'grind' suppletion involving both an initial consonant and vowel apophony. The two sets of cells indicated above also differ frequently in their tone.

The language does not allow for complex onsets and so forms like 1PL **k-rig-ɛl* would not be allowed. Similarly, vowel initial onsets are also disallowed. Forms like 3SG **irg-ɛl*, therefore, would be ill-formed too (see Footnote 69). The patterns lend themselves to different analyses in terms of which (if any) is the basic form of the stem and which is the derived one. If the form of the 3SG were regarded as basic (e.g. Waag 2010:118), then the /k/ at the beginning of 'grind' will be said to be deleted in the prefixed forms. If the other stem is considered basic instead (e.g. Beaton 1968) then the formation of the 3SG in 'grind' will involve the insertion of /k/ as an onset prefix.

⁶⁸ The glottal stop occurs automatically as a subphonemic onset before a vowel (Jakobi 1990:42, Waag 2010:115) so the the 1SG should probably be thought of as unprefix.

Because different verbs will have different initial consonants in this stem, the analysis of Waag would seem preferable in that it does not lead to a proliferation of inflectional classes. However, this analysis faces challenges in other respects. Subtractive affixes are less restrictive than additive ones. In addition, the form of the 1/2/3PL.HUM stem is not always predictable from the form of the alleged basic stem. More revealingly still maybe, some verbs (e.g. ‘teach’ and ‘disagree’, see Waag 2010:120) can be homophonous in the 3SG/3PL.NHUM stem (3SG *paareɪ*) but have a different stem (1SG *ʔaareɪ* vs *ʔawreɪ* respectively) elsewhere.

Because of the great number of processes and forms involved and because of these aforementioned complications, I consider that both stems need to be stored in most cases and that the paradigmatic distribution of the stems must simply be considered morphomic. The absence of a sufficient description of Amdang, the only other close relative of Fur, makes it difficult at present to speculate about the diachronic emergence of this morpheme, although it seems related to the presence of prefixes in some person-number forms and the absence of prefixation from others.

Fur1: HUM.SG/NHUM

Fur2: 1SG/2SG/PL

5.2.22 Girawa (Gasaway & Sims 1977)

In Girawa (Trans-New.Guinea), there is a close morphosyntactic affinity of first person and second person singular which is manifested both in some verb stems and in (subject and object) agreement suffixes. Consider the following paradigms:

	'eat' present			'hit' present, 2SG subject ⁶⁹		
	SG	DU	PL	SG	DU	PL
1	je-m	je-m-ur	je-m	iw-ir-om	iw-it-om	iw-ik-om
2	je-m	jeir	jei	iw-is-om	ak-wat-om	ak-war-om
3	jeu	jeir	jei	wem	ak-wat-om	ak-war-om

Table 174: Partial paradigms of two Girawa verbs (Gasaway & Sims 1977)

As the Table 174 above illustrates, the 1 and 2SG cells constitute an internally homogeneous and externally heterogeneous class concerning certain agreement formatives (see 'eat'). In a class of verbs that also indexes the object (see 'hit'), this also constitutes the domain for stem allomorphy. These stem alternations usually involve segmental changes in the right periphery of the stem (e.g. *apa/ap/apar* 'see', *urwo/ur/urw* 'call out', *taine/tain/tainor* 'follow'). Only in *iw/ak/w(e)* 'hit' do they reach (near-)suppletion (Gasaway & Sims 1977:30).

Object suffixes, when they occur, immediately follow the verb stem. Their overall form (*-i* vs *-wa* vs \emptyset) agrees with the morphomic patterns of stem alternation, which makes it plausible to argue that the different phonological profiles of these suffixes were responsible for the emergence of stem alternations in the first place. This receives support from comparative evidence from other Madang languages (see e.g. Amele below) which seem to lack stem alternations but do have object agreement suffixes with the same pattern:

	'come' remote past			'cut' 3SG subject, progressive		
	SG	DU	PL	SG	DU	PL
1	ho-om	ho-h	ho-m	qet-it-ina	qet-il-ina	qet-ig-ina
2	ho-om	ho-sin	ho-in	qet-ih-ina	qet-ale-na	qet-ade-na
3	ho-n	ho-sin	ho-in	qet-ud-ina	qet-ale-na	qet-ade-na

Table 175: Partial paradigms of two Amele verbs (Roberts 1987:279)

⁶⁹ The form where the object itself is second singular (i.e. *iwisom*) has to be understood as having a 1SG or 1PL subject instead. There is some allomorphy of some of the object (e.g. *ir/or/ur*) and the subject (e.g. *om/em/im*) suffixes that seems to be dependent on the phonological context but which is not described in sufficient detail in Gasaway & Sims (1977) to be sure the forms I provide above are the correct allomorphs in this case. This is irrelevant, however, for my general analysis of this morphomic pattern. It is worth noting, nevertheless, that the form of this vowel may occasionally differ between 1/2SG and 1DU/PL.

Observe how in Amele, the distribution of *i-*, *a-*, or *u-*initial object suffixes mirrors the paradigmatic organization of stem alternation in Girawa. Observe also that a degree of suffixal similarity (involving also the segment /m/) also exists in Amele between 1/2SG and 1PL subject suffixes as well.⁷⁰

Although the diachronic details are uncertain, it seems to be the case that a more-or-less inconsequential phonological resemblance of person object suffixes created in Girawa a pattern of stem alternation whereby the same stem was shared by 1 and 2SG. This pattern, in turn, would have been learned as a morphomic grammatical entity by language users. This would have contributed to the promotion of 1/2SG identity in other domains, for example facilitating the spread of the suffix /m/ to the 1DU in Girawa (compare Tables 174 and 175).

5.2.23 Greek (Holton et al. 2012)

In modern Greek, a prefix, known in the literature as the augment, appears in the past tense of some verbs in the SG and the 3PL forms. Consider the following paradigms:

	'tie'		'know'	
	SG	PL	SG	PL
1	'e- <i>desa</i>	' <i>desame</i>	'i- <i>xera</i>	' <i>xerame</i>
2	'e- <i>deses</i>	' <i>desate</i>	'i- <i>xeres</i>	' <i>xerate</i>
3	'e- <i>dese</i>	'e- <i>desan</i>	'i- <i>xere</i>	'i- <i>xeran</i>

Table 176: Aorist past tense⁷¹ paradigm of two Greek verbs (Holton et al. 2012)

This affix appears usually as /e/, which must have been originally its only form. Only in a few verbs, it has the form /i/ instead. In Ancient Greek (and also in other older Indo-European

⁷⁰ This pattern is also found in Kosena and a very similar one is found in Yagaría. Both are TNG languages (distantly related to Girawa and Amele) and instantiate these syncretisms with a suffix /n/. Consider also the similarity of stem alternation in some of these languages (see Sections 5.2.9 and 5.2.69).

⁷¹ The other past tense, the imperfect, shows the same pattern.

languages), this augment *e-* was used in all past tense forms. Before consonant-initial verbs, the prefix was simply *e-* and, because it formed a syllable of its own, it is known as the ‘syllabic augment’. Before certain vowel-initial verbs (e.g. in the case of ‘know’ above, which was *exeur-* in Ancient Greek), the /e/ of the prefix and the stem-initial vowel were fused into a long vowel. This was often /e:/, which has become /i/ in the modern language by regular sound change.

Along with the addition of this prefix, past tense forms were also characterized in Greek by being stressed on the antepenultimate syllable. This meant that, in some verbs, depending on the shape of the person-number suffixes, the stressed vowel could be in the stem or in the augment. With the longer, syllabic person suffixes (1PL and 2PL) the stress fell on the root while with the shorter, non-syllabic person suffixes (SG+3PL), the stress fell on the augment. When unstressed initial vowels were elided in the medieval language, an alternation was introduced between the former, which lost the prefix, and the latter, which kept it.

The parallels between the diachronic emergence of this alternation and that of the renowned N-morpheme of Romance are remarkable. We have a stress assignment rule that, in conjunction with person-number suffixes of different phonological profiles, leads to the stressed syllable being different in different forms. Then a run-of-the-mill sound change created differences between stressed and unstressed vowels. The pattern arrived at (SG+3PL vs 1PL+2PL) is also the same in Greek and Romance.

5.2.24 Icelandic (Jörg 1989)

In the verbal inflectional system of Icelandic and other conservative Germanic languages, there are complex patterns of stem alternation involving mostly, but not only, vowel apophony. Alternations were at the earliest stages more-or-less correlated with semantic distinctions but later sound changes and analogical reshapings have meant that parts of the paradigm have come to share form despite a lack of semantic or morphosyntactic common thread. Consider the following paradigms:

	Indicative				Subjunctive			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	sé	sjáum	sá	sáum	sjái	sjáum	sæi	sæjum
2	sérð	sjáið	sást	sáuð	sjáir	sjáið	sæir	sæjuð
3	sér	sjá	sá	sáu	sjái	sjái	sæi	sæju

Table 177: Paradigm of *sjá* ‘see’ in Icelandic

	Indicative				Subjunctive			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	á	eigum	átti	áttum	eigi	eigum	ætti	ættum
2	átt	eigið	áttir	áttuð	eigir	eigið	ættir	ættuð
3	á	eiga	átti	áttu	eigi	eigi	ætti	ættu

Table 178: Paradigm of *eiga* ‘own’ in Icelandic

In Icelandic, every single verb except for the verb ‘be’ has the same stem in the infinitive, in the plural present indicative, and in the present subjunctive. The actual concrete forms being shared by these cells can vary. In verbs without stem alternation, of course, no particular morphological affinity will be apparent between the shaded cells. In other cases (see *sjá* ‘see’ in Table 177), one segment /j/ is shared by the shaded cells. In yet other cases (see *eiga* ‘own’ in Table 178), the whole of the stem /ei:ɣ/ is exclusive to the mentioned paradigm cells.

5.2.25 Iraqw (Mous 1992)

Verbs in the South Cushitic language Iraqw show a morphological affinity of 2 and 3SG feminine, which are opposed to the rest of the paradigm (i.e. 1+3SG.M+3PL) in a number of ways. Consider the following paradigms:

	'leave' present		'follow' present		'eat' present	
	SG	PL	SG	PL	SG	PL
1	máw	mawáan	eehár	eeharáan	ᶑáay	ᶑaayáan
2	méer	méeraʔ	eehát	eehatáʔ	ᶑág	ᶑagáʔ
3.M	máy	mayáʔ	eehar	eeharír ⁷²	ᶑaay	ᶑaayír
3.F	méer	mayáʔ	eehát	eeharír	ᶑág	ᶑaayír

Table 179: Present paradigms of three Iraqw verbs (Mous 1992:156-157)

As illustrated in Table 179, the two sets of cells show formal differences which can be very diverse (*a/eer*, *r/t*, *ay/g* above, but also *w/b*, *h/t*, *r/n*, *V:/V* elsewhere). There is evidence, in addition, (see Kießling 1994:132) that these cells have behaved as a unit in processes of analogical change. These are facts that identify this pattern as robustly morphomic.

Most of the alternations we see today, however, can be traced back to regular sound changes. Following the common Afroasiatic pattern (still readily observable, for example in modern Arabic, or in more closely related Afar, see Kamil 2004:81), the 2nd and the 3SG.F would have been characterized by a /t/ (or t-containing) affix in older stages of the language. In this part of Cushitic, these formatives were suffixed to the stem. In the course of time, certain sound changes (most importantly the lenition of stops [*g>y*, *b>w*, *d>r*] in certain positions, the shortening of vowels before a consonant cluster, and the loss of certain word-final segments, see Mous 1992:160) introduced stem alternations in the language and obliterated the original conditioning environment. Consider, for example:

eat.3SG.F *ᶑaag-t > *ᶑaag-t > *ᶑag-t > ᶑag

eat.3SG.M *ᶑaag-i > *ᶑaay-i > *ᶑaay-i > ᶑaay

This development is entirely parallel to (but also completely independent from) the emergence of the morphomic agreement system described earlier here for the East Cushitic

⁷² There are two alternative forms for the 3PL in these verb and others. The two alternatives however (e.g. *eehariyáʔ* and *eeharír* in this verb) always share the exponence (/r/ in this case) which is at stake here.

language Daasanach.⁷³ This could indicate that particular affixal configurations like the (accidental?, see Harbour 2008) Afroasiatic homophony of 2 and 3SG.FEM -t are particularly good breeding grounds for the emergence of morphemes.

Iraqw1: 2/3SG.F

Iraqw2: 1/3SG.M/3PL

5.2.26 Irish (Doyle 2001)

In Irish nominal declension, one can often find a whole-word syncretism of genitive singular and nominative plural, which often share some segment(s) to the exclusion of the rest of the paradigm. Consider the following:

	'boat'		'habit'		'son'		'woman'	
	SG	PL	SG	PL	SG	PL	SG	PL
NOM	bád	báid	béas	béasa	mac	mic	bean	mná
GEN	báid	bád	béasa	béas	mic	mac	mná	ban

Table 180: Declension of some Irish nouns⁷⁴ (Doyle 2001)

As the nouns above illustrate, the forms involved may be diverse: palatalization of the last consonant of the stem ($b^{\vee}\alpha:d^{\vee}/b^{\vee}\alpha:d^i$ 'boat'), sometimes along with a different stem vowel ($m^{\vee}\alpha k/m^i\alpha c$ 'son'), suffixation ($b^j e:s^{\vee}/b^j e:s^{\vee}\alpha$ 'habit'), and even suppletion occasionally ($b^j \alpha n^{\vee}/m^{\vee} n^{\vee}\alpha$: 'woman'). This morphological affinity is a very old Indo-European trait that goes

⁷³ Conservative languages in both East Cushitic (e.g. Oromo, see Ali & Zaborski 1990) and South Cushitic (e.g. Burunge, see Kießling 1994) still show the well-known Afroasiatic dental suffixes -t/d in 2 and 3SG.F. This rules out genetic inheritance of these stem alternations from a common ancestor. The two languages are also separated by almost 1000 kms, thus making areal influences similarly unlikely.

⁷⁴ The nominative and accusative cases are not distinguished in Modern Irish and the dative is most usually syncretic with them too. The vocative has not been included above either.

back to some affixal identities that are still visible in more conservative languages (e.g. Lithuanian: *ašv-os* is the GEN.SG and NOM.PL of *ašva* ‘mare’, and Russian: *knig-i* is both the GEN.SG and NOM.PL of *kniga* ‘book’).

5.2.27 Italian (Maiden & Robustelli 2014)

Italian verbal inflection, as that of other conservative Romance varieties, is characterized by morphomic stem alternation patterns. Two Italian morphemes have been included in the present database, the Italian versions of the morphemes referred to as the U-morphome and PYTA in the literature. Consider the former:

	<i>cogliere</i> ‘pick’		<i>dire</i> ‘say’		<i>apparire</i> ‘appear’	
	IND	SUBJ	IND	SUBJ	IND	SUBJ
1SG	colgo	colga	di[k]o	di[k]a	appaio	appaia
2SG	cogli	colga	di[tʃ]i	di[k]a	appari	appaia
3SG	coglie	colga	di[tʃ]e	di[k]a	appare	appaia
1PL	cogliamo	cogliamo	di[tʃ]iamo	di[tʃ]iamo	appariamo	appariamo
2PL	cogliete	cogliate	dite	di[tʃ]iate	apparite	appariate
3PL	colgono	colgano	di[k]ono	di[k]ano	appaiono	appaiano

Table 181: Present tense paradigms of three Italian verbs (Maiden & Robustelli 2014)

As is well-known, these stem alternations originated as a result of the palatalization of some consonants before front vowels (e.g. in *dire*, an older 2SG.IND $dī[k]is > dī[tʃ]is$). Note that, even if the alternations are completely morphological in the modern language, the original conditioning environment and stem alternation continue to be coextensive in the paradigm and not merely due to a continuation from the inherited configuration (forms like Latin 1PL.SUBJ *dīcāmus* and 2PL.SUBJ *dīcātis* would have yielded *di[k]amo* and *di[k]ate* respectively). This same fact can be illustrated by the cognate morphome in other varieties:

	<i>pote</i> 'can'		<i>di</i> 'say'		<i>ae</i> 'have'	
	IND	SUBJ	IND	SUBJ	IND	SUBJ
1SG	pɔttso	pɔttso	diko	diko	aggjo	aggjo
2SG	poi	poi	ditʃi	ditʃi	ai	ai
3SG	pɔ	pɔttsa	ditʃe	dika	a	aggja
1PL	putimo	putimo	ditʃimo	ditʃimo	aimo	aimo
2PL	potete	potete	ditʃete	ditʃete	aete	aete
3PL	pɔ	pɔttsa	ditʃe	dika	a	aggja

Table 182: Present tense paradigms of three Servigliano Italian verbs (Camilli 1929)

As the variety of Italian spoken around Servigliano illustrates, the presence of the exceptional stem alternant (e.g. *dik-*) is correlated with the presence of a back vowel in the corresponding suffix even though the phonological changes that created these alternations have long been dead⁷⁵ and even though the paradigms have been extensively restructured (consider the loss of mood distinctions outside of the third person or the loss of number distinctions in the third person). Because of the different extension of the L/U-morpheme in this variety, this has been included as a separate morpheme in this database.

Another morphological quirk of many Italian verbs is the presence, in three cells of the past tense, of a special stem not present in the rest of the paradigm. Coming back to standard Italian, consider again (see also Section 4.2.4.2) this other pattern of stem alternation:

	<i>fare</i> 'do'		<i>cuocere</i> 'cook'		<i>rompere</i> 'break'	
	SG	PL	SG	PL	SG	PL
1	feci	facemmo	cossi	cuocemmo	ruppi	rompemmo
2	facesti	faceste	cuocesti	cuoceste	rompesti	rompeste
3	fece	fecero	cosse	cossero	ruppe	ruppero

Table 183: Past tense paradigms of three Italian verbs (Maiden & Robustelli 2014)

⁷⁵ See Section 3.3 for a discussion of these morphological patterns that correlate to a coherent phonological environment and why they are considered morphomic here.

This alternation emerged from an originally semantically motivated one: a perfective stem opposed in Latin to an imperfective one. Those roots would have thus been associated originally to whole tenses and still are in some contemporary Romance varieties like Spanish. Italian, however, lost these special roots in those paradigm cells which were arhizotonic (also less frequent, see Section 4.1.1.1). The result is a person-number morpheme (Smith 2013) that is, the same as the previous ones, formally diverse (e.g. *fec-i fac-esti* above, but also *conobb-i conosc-esti* ‘know’, *apparv-i appar-isti* ‘appear’, *nacqu-i nasc-esti* ‘be born’ etc.).

Italian1: 1SG.IND/3PL.IND/SG.SUBJ/3PL.SUBJ

Italian2 (Servigliano): 1SG/3.SUBJ

Italian3: 1SG/3

5.2.28 Jabuti (Pires 1992)

Some Jabuti (Macro-Je) verbs (also nouns, which have similar morphology) whose stem begins with /h/ are subject to a stem alternation pattern that opposes 2+1PL (which are prefixed in the language) to 1SG+3 (which are unprefixes). Consider the following:

	‘get tired’		‘fall’	
	SG	PL	SG	PL
1	habä	hi-rabä	hökü	hi-nökü
2	a-rabä	a-rabä	a-nökü	a-nökü
3	habä	habä	hökü	hökü

Table 184: Paradigms of two Jabuti verbs (Pires 1992:45-46)

The alternations displayed above go back to an originally non-alternating paradigm. Van der Voort (2007:150) shows that words like these probably had \sqrt{tj} as their original stem-initial

consonant in all the forms. This sound is preserved in the closely-related language Arikapu. In Jabuti, however, in some intervocalic phonological contexts, this phoneme changed to /r/ (maybe through some intermediate [allophonic?] stage ʒ-ʝ). Later on, some of those /r/ (those before nasal vowels) changed in turn to /n/, thus creating the diversity of alternations found in Jabuti synchronically. It is important to note that the consonants /h/, /r/ and /n/ (and also /tʃ/ for that matter) are not allophones synchronically but independent phonemes that are phonotactically unrestricted in the present language (Pires 1992:24-28).

5.2.29 Jerung (Opgenort 2005)

Jerung (Sino-Tibetan) has a morphologically-determined pattern of stem alternation which involves the same (longer) stem in the SG and 3.NSG. Consider the following paradigm:

	SG	DU	PL
1EXCL	gɔk-ma	go-cum	go-kum
1INCL	-	go-cim	go-kim
2	gɔk-nim	go-cim	go-nimme
3	gɔkt-im ⁷⁶	gɔk-cim	gɔk-me

Table 185: Paradigm of Jerung 'give', 3SG patient (Opgenort 2005:330)

As Table 185 illustrates, this pattern of stem alternation can involve both final consonant(s) and stem vowel. These alternations are confined to transitive verbs and involve most often a stem augment /t/, with or without further segments. This formative descends ultimately from a valency-increasing suffix in Proto-Tibeto-Burman (see Michailovsky 1985).

Similar stem alternations in East Kiranti languages are predictable from the vowel-initial vs consonant-initial forms of the suffix, with the longer stem appearing before a vowel and the shorter one before a consonant (see Hecce forthcoming-a). This might be the origin of the

⁷⁶ The alternation between *gɔkt-* and *gɔk-* is phonological.

stem alternation in Jerung too. Synchronically, however, it has become unmistakably morphological in this language, since the same suffix (e.g. DU *-cim*) can co-occur with both stems (contrast 2DU to 3DU). The picture is similar in closely-related Wambule:

	SG	DU	PL
1EXCL	twan-me	to-cukme	to-kume
1INCL	-	to-cime	twan-ime
2	twan-nume	to-cime	to-nime
3	twan-s-ume	twan-s-ucime	twan-mime

Table 186: Paradigm of Wambule ‘drop’ (*twan- to-*) (Opgenort 2004:874)

It is still interesting to notice, however, that where the two languages differ (i.e. in the 1PL.INCL), suffix and stem have both changed together in a way that, even if phonological determination no longer holds, seems to suggest that some degree of preference may apply.

5.2.30 Karamojong (Novelli 1985)

Verbal inflection in Karamojong (Nilotic) involves prefixes that mark, cumulatively, person-number agreement, tense, mood, and voice. In the active paradigm, 1SG, 1PL, 2, and 3 are usually distinguished, although some syncretism can also be found occasionally. In the passive, by contrast, 2 is always syncretic with 1PL. Consider the following prefixes:

	Indicative		Subjunctive		Narrative	
	SG	PL	SG	PL	SG	PL
1	aka-	iki-	k’aka-	k’iki-	ɔkɔ-	itɔ-
2	iki-	iki-	k’iki-	k’iki-	itɔ-	itɔ-
3	a-	a-	k’ɛ-	k’ɛ-	tɔ-	tɔ-

Table 187: Karamojong Conjugation 1, Past, Passive prefixes (Novelli 1985:202)

The passive prefixes seem to be derived from the active ones. Whereas the two are the same in the third person, first and second person add segments to the left of active forms. The actual forms being added differ from one mood to another and from 1PL to 2. For, example, in the subjunctive, the second person adds *-ik-* while the 1PL does not add anything. In the narrative mood, by contrast, the second person adds *i-* while the 1PL adds *it-*. It looks as if the goal of these morphological operations were to achieve a syncretism of 2 and 1PL in the passive to the exclusion of the rest of the paradigm.

5.2.31 Kele (Ross 2002)

In some Oceanic languages like Kele (see also Vurës in later Section 5.2.66) nouns in their possessive inflection are subject to stem alternations. Consider the following paradigms:

	<i>dop</i> 'basket'			mah 'taro'		
	SG	DU	PL	SG	DU	PL
1EXCL	dópu	dábo-yoru	dábo-yotu	mohí	mohé-yoru	mohé-yotu
1INCL	-	dábo-teru	dábo-titu	-	mohé-teru	mohé-titu
2	dópu-m	dábo-eru	dábo-etu	mahí-m	mohé-eru	mohé-etu
3	dábo-n	dábo-heru	dábo-su	mohé-n	mohé-heru	mohé-su

Table 188: Possessor paradigm of two Kele nouns (Ross 2002:133)

As the paradigms above illustrate, 3SG and all the nonsingular cells always share the same stem. In the cases with the maximum number of alternants (see the paradigm of 'taro') there are four stems: one used in non-possessed contexts, another one in the 1SG, another in the 2SG and the one of 3SG+NSG. Some (or all) of these stem spaces may be formally identical in particular lexemes (see e.g. 1SG and 2SG in 'basket') but the stem in 3SG and NSG is unexceptionally the same, which constitutes a morphomic formal alignment.

As explained by François (2005), these stem alternations must have originated by way of stem-vowel assimilation to the following possessive suffixes. It must be noted that the singular cells did contain overt syllabic suffixes as well in earlier stages of the language (these have been reconstructed as *-gu (1SG), *-mu (2SG) and *-ña (3SG) in Proto-Oceanic, see Lynch et al. 2002:76).

5.2.32 Ket (Georg 2007)

In Ket (Yeniseian) inflectional morphology, the neuter plural behaves as the singular. Consider the following inflectional affixes:

	Genitive suffixes		Actant suffixes		Possessive prefixes	
	SG	PL	SG	PL	SG	PL
3M	-da	-na	-o	-oŋ	da-	na-
3F	-di	-na	-u	-oŋ	d-	na-
3N	-di	-di	-u	-u	d-	d-

Table 189: Some inflectional formatives in Ket (Georg 2007:104, 119, 268)

Sometimes, this absence of distinction between neuter singular and plural, as Table 189 shows, leads to a syncretism with the feminine SG that is morphosyntactically unnatural but is repeated with several exponents. It is worth mentioning that neuter nouns do distinguish number morphologically (e.g. *dón-di* 'knife-GEN' vs *dónaŋ-di* 'knives-GEN, Georg 2007:104) so it is just their targets that fail to do so.

5.2.33 Khaling (Jacques et al. 2011, Jacques 2017)

The verbal inflectional morphology of Khaling (Sino-Tibetan) is complex when it comes to stem alternation. Although clear correlations can be found between stem and suffix forms (e.g. a nasal-initial suffix and a nasal-final stem usually appearing together) most of the formal

alternations have become morphologized. Contributing to this complexity is the fact that almost every stem coda behaves on an idiosyncratic manner (i.e. in a way that cannot be generalized to other forms) regarding these morphological alternations. Because of this, most alternations in the language cannot be labeled morphomic with the criteria I have set here. Consider, however, the following paradigms:

	‘have enough’			‘look nice’		
	SG	DU	PL	SG	DU	PL
1EXCL	soðm-ŋʌ	səp-u	soɔp-kʌ	bū:-ŋʌ	biŋ-u	bʌŋ-kʌ
1INCL		səp-i	soɔp-ki		biŋ-i	bʌŋ-ki
2	ʔi-soɔp ⁷⁷	ʔi-səp-i	ʔi-soðm-ni	ʔi-bʌŋ	ʔi-biŋ-i	ʔi-bū:-ni
3	soɔp	səp-i	soðm-nu	bʌŋ	biŋ-i	bū:-nu

Table 190: Two Khaling verbs, non-past, intransitive (Jacques et al. 2011:1102,1148)

Observe how, in the non-past tense, paradigm cells may differ in the stem they use. Despite what might seem to be the case in Table 190, none of these two alternations is a regular phonological rule of the language. Both are purely morphological, which can be seen in the existence of forms like *lê:p-nu* ‘catch-3PL>3SG’ (Jacques et al. 2011:1102) or *sîŋ-nu* ‘ask-3PL>3SG’ (Jacques et al. 2011:1150), where a suffix *-nu* (phonologically identical to the 3PL suffix in the paradigms above) does not trigger nasalization or loss of stem-final /ŋ/.

The nasal /m/ at the end of the stem in ‘have enough’ and the vowel /u:/ in ‘look nice’ are thus used in these verbs’ stems in 1SG, in 2PL and in 3PL. Other verbs show an alternation in the same paradigm cells between *-Vk* and *-V:* (e.g. *tsek* ‘be hard’, Jacques et al. 2011:1139) and between *-Vŋ* and *-V:* (e.g. *ghaŋ* ‘agree’ Jacques et al. 2011:1131) more generally. This constitutes more than enough formal variation to classify this morphological affinity as morphomic according to the criteria that have been set here.

⁷⁷ The rest of the paradigm of these verbs has /p/ and /ŋ/ respectively as the stem-final consonant. This is the reason why it might be considered the default form and has not been included in the morpheme database.

Another morphological affinity in Khaling (one which is a superset of the cells above and which is instantiated by similar forms) also deserves its inclusion into the present database of metamorphemes. Consider the following paradigm:

	Present			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	ʔλm-si-ŋλ	ʔip-si-ju	ʔλp-si-kλ	ʔλm-tλsu	ʔip-si-jtu	ʔλp-si-ktλkλ
1INCL		ʔip-si-ji	ʔλp-si-ki		ʔip-si-jti	ʔλp-si-ktiki
2	ʔi-ʔλm-si	ʔi-ʔip-si-ji	ʔi-ʔλm-si-ni	ʔi-ʔλm-tε-si	ʔi-ʔip-si-jti	ʔi-ʔλm-tε-nnu
3	ʔλm-si	ʔip-si-ji	ʔλm-si-nu	ʔλm-tε-si	ʔip-si-jti	ʔλm-tε-nnu

Table 191: Paradigm of Khaling ‘sleep’ (ʔip-) past, reflexive (Jacques 2017:6).

In Khaling, reflexive verbs require a nasalized stem in SG+2PL+3PL cells. This stem may be characterized by a stem-final /m/ (vs /p/), /ŋ/ (vs /k/) or /n/ (vs /ŋ/, /t/ or zero) and by use of the same pitch. In the past, these cells are also different from the rest of the paradigm in that the reflexive suffix *-si* does not appear immediately after the stem. Instead, the past suffix *-t(ε)* appears first.

Although their diachronic emergence and evolution are not clear, stem nasalizations with a similar formal and paradigmatic profile are found in other West Kiranti languages like Bahing (Michailovsky 1975:189) or Wayu (Michailovsky 1988:81). These alternations must have emerged from sound change; as a phonological assimilation process of stops to a following nasal suffix. The alternations would have been subsequently morphologized and left at the mercy of analogical processes and later sound changes.

Khaling1: 1SG/2PL/3PL

Khaling2: SG/2PL/3PL

5.2.34 Khinalug (Kibrik 1994)

According to their agreement morphology in the verb, Khinalug nouns fall into four different genders. These have been labelled below ‘masculine’, ‘feminine’, ‘animate’ and ‘inanimate’ on the basis of their semantic core (although membership into III or IV is less systematic as in the other two genders). The agreement markers that reveal this gender division, however, are syncretic in morphomic ways. Consider the affixes involved:

	Set 1		Set 2		Set 3	
	SG	PL	SG	PL	SG	PL
I Male	∅	b	j	v	h	f
II Female	z	b	z	v	s	f
III Animate	b	∅	v	j	f	h
IV Inanimate	∅	∅	j	j	h	h

Table 192: Gender agreement morphology in Khinalug (Kibrik 1994:387)

As the above Table 192 illustrates, for the purposes of morphology, the singular of gender I, the plural of gender III, and gender IV constitute a single class. Similarly, the plural of genders I and II and the singular of gender III are always syncretic too. These morphological affinities are systematic because they are implemented with different formatives. The different sets correspond to different slots in the verbal complex (Sets 1 and 2) and to a small number of irregular verbs in the case of Set 3.

The multiplicity of forms with which the various morphological classes are instantiated must have emerged from sound changes taking place on an originally invariable affix. The phonological affinity (e.g. the labial character of all /b/, /v/ and /f/), points in this direction. As for the history of these syncretisms, the evidence suggests that some of them are very old and quite stable diachronically:

	Proto-Lezgian		Khinalug	
	SG	PL	SG	PL
I Male	*w	*b	∅	b
II Female	*r/j	*b	z	b
III Animate	*b	*d	b	∅
IV Inanimate	*d	*d	∅	∅

Table 193: Gender agreement morphology in some Daghestanian languages

The one of I/II.PL+III.SG in Khinalug, for example, is also found in relatively distantly-related Tsakhur (Schulze 1997), Hunzib (van den Berg 1995), and Archi (Chumakina & Corbett 2015) (it also has cognates in Nakh) and, as the above Table 193 shows, is reconstructed for the ancestral language (Schulze 1997:26-27). The antiquity of this particular pattern in Khinalug does not preclude, of course, the occasional reconfiguration of these morphological gender-number morphomic classes in particular languages (see Lak in Table).

The other morphomic class of Khinalug, for example, appears to have involved the fusion of two different exponents, since I.SG does have a non-syncretic exponence in related languages like the ones mentioned above (i.e. Tsakhur, Hunzib and Archi). The merger of these two morphological classes into one in Khinalug may have resulted from their exponents falling together in some of their allomorphs (maybe as zero in Set 1) and this identity being later construed as systematic and extended to the other allomorphs.

Khinalug1: IPL/IIPL/IIISG

Khinalug2: ISG/IIPL/IVSG/IVPL

5.2.35 Koasati (Kimball 1985)

In the verbal person-number inflection in Koasati (also in the closest-related Muskogean languages like Alabama, see Lupardus 1982:140), one can identify a clear morphological affinity between the 2SG, the 1PL, and the 2PL in most conjugations. These values are marked

in the same syntagmatic position within the word (e.g. compare *hófna-l* ‘smell.1SG’ to *ho<lí>fn* ‘smell.1PL’, Kimball 1985:70) and sometimes share formal exponents as well. Consider the following paradigms:

	<i>míkkon</i> ‘be a chief’		<i>cákkín</i> ‘catch up with’	
	SG	PL	SG	PL
1	mikko-lí	mikko-t-il-ká	cákki-l	cák-h-íl-k
2	mikko-t-is-ká	mikko-t-as-ká	cák-h-ís-k	cák-h-ás-k
3	mikkó	mikkó	cák	cák

Table 194: Person-number inflection in two Koasati verbs (Kimball 1985:76, 80-81)

5.2.36 Koiari (Dutton 1996, Dutton 2003)

Koiari (Koiarian, New Guinea) tense-aspect suffixes sometimes have a different form depending on the person-number of the subject. Frequently, only two forms are distinguished, whose paradigmatic distribution does not correlate with a meaning dimension. Consider the following paradigms and suffixes:

	‘see’ Perfect aspect		Imperfect aspect		Obligatory mood	
	SG	PL	SG	PL	SG	PL
1	ereva-nu	ereva-nua	-ma	-a	-ahi-na	-iha-va
2	ereva-nua	ereva-nua	-a	-a	-iha-ma	-iha-va
3	ereva-nu	ereva-nua	-ma	-a	-ahi-ma	-iha-va

Table 195: Some Koiari TAM morphology (Dutton 1996:23, Dutton 2003:346,351)

Some morphological elements appear, as illustrated in Table 195, in 2SG+PL, which is morphosyntactically unnatural. In Koita, the closest relative to Koiari, some of these forms (e.g. Imperfect *-ima* vs *-a*, see Dutton 1975:338) correspond to a SG vs PL distinction. It is unclear to me how the Koiari system may have come about. In other languages (e.g. Basque

and English) where we find a PL+2SG morphomic pattern, this emerged when a plural pronoun started to be used to refer politely to a SG addressee. In Koiari, however, there do not seem to have been any processes of this kind, since Koita and Koiari have cognate pronouns (2SG *a* vs 2PL *ya*) with the same values. The history of this pattern is therefore unclear (but see Section 5.2.7 and Footnote 62).

5.2.37 Kosenā (Marks 1974)

In the grammar of Kosenā (Trans-New Guinea, also known as Awiyaana) and in related Usarufa there appear to be various and complex morphophonological rules operating across morpheme boundaries. Paradigms often show unmotivated morphological allegiances. For example, the 1SG and the 1PL are usually formally identical to the exclusion of 1DU. This syncretism often extends to the 2SG too. Consider the following suffix paradigms:

	DS Present			Indicative			Interrogative			Assertive		
	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL
1	-una	-uya	-una	-ne	-we	-ne	-no	-o	-no	-mpo	-vo	-mpo
2	-na	-ya	-wa	-ne	-we	-we	-no	-o	-o	-mpo	-vo	-vo
3	-isa	-ya	-wa	-we	-we	-we	-o	-o	-o	-vo	-vo	-vo

Table 196: Paradigm of various inflectional suffixes in Kosenā (Marks 1974)

This morphological affinity in mood suffixes is similar to the one found in related languages like Yagarā (see Section 5.2.69) and most probably also has a similar diachronic origin. Consider also the similarity of this pattern to the one in Amele (Table 175).

5.2.38 Luxembourgish (Schanen 2004)

In Luxembourgish, as in other West Germanic varieties, some sound changes have resulted in the presence of different stem alternants in the present tense inflection of verbs. One of these sound changes is Umlaut. An /i/ formerly present in the 2SG and 3SG suffixes raised the /a/ of the stem in many verbs (e.g. in German *tragen* ‘carry’: 1SG *trage*, 2SG *tr[e]gst*, 3SG *tr[e]gt*, 1PL *tragen*, 2PL *tragt*, 3PL *tragen*).

This sound change, thus, created a pattern of stem alternation where the 2SG and 3SG cells were opposed to the rest (i.e. 1SG+PL). Other, unrelated sound changes, e.g. closed-syllable shortening, other types of Umlaut etc. (see Albright 2010) gave (or would have given) rise to different patterns of stem vowel alternation. These, however, (even motivated alternations like SG *gib-* vs PL *geb-* in ‘give’) have often been made to conform to the (morphomic) pattern of stem alternation presented here. In German, for example, we find 1SG+PL *g[e:]b-* vs 2/3SG *g[i]b-*. The tendency to have characteristic forms in these cells is also witnessed, maybe even more strongly, in Luxembourgish. Consider the following paradigms:

	<i>sinn</i> ‘be’		<i>kommen</i> ‘come’		<i>maachen</i> ‘make’	
	SG	PL	SG	PL	SG	PL
1	sinn	sinn	kommen	kommen	maachen	maachen
2	bass	sidd	kënns	kommt	méchs	maacht
3	ass	sinn	kënnt	kommen	mécht	maachen

Table 197: Three Luxembourgish verbs, present tense (Schanen 2004)

As explained before (see Tables 118 and 119 and the ensuing discussion), the morphological affinities displayed in Table 197 have emerged analogically in Luxembourgish. The replacement, in the verb ‘to be’, of the inherited 1SG form (which started with /b/, cf. German *bin*) by the plural stem in *s-*, the introduction of stem alternation in etymologically nonalternating ‘make’ etc. show that, after emerging from sound change, the morphological identity of the 1SG+PL present has been productive in the language to some extent.

5.2.39 Maijiki (Velie & Velie 1981)

The verbal morphology of Maijiki (Tucanoan, also known as Orejon) shows an interesting shift between declarative and interrogative contexts. In the former, the 2SG is formally identical to the 1SG. In the latter, it is syncretic with the 3SG instead, which shows a gender distinction. Because some of the suffixes appear in both declarative and interrogative contexts, their overall paradigmatic distribution is unnatural as a result of the aforementioned shift of the 2SG. Consider the following paradigms:

	Declarative		Interrogative	
	SG	PL	SG	PL
1	sa-hi	sa-hi	sa-te	sa-te
2M	sa-hi	sa-hi	sa-ki	sa-te
2F	sa-hi	sa-hi	sa-ko	sa-te
3M	sa-ki	sa-hi	sa-ki	sa-te
3F	sa-ko	sa-hi	sa-ko	sa-te

Table 198: Preterite paradigm of the verb 'go' in Maijiki (Velie & Velie 1981:124-125)

As Table 198 illustrates, suffixes like *-ki* and *-ko* appear only with the 3SG in declaratives but with both 3SG and 2SG in interrogatives. This constitutes a morphomic paradigmatic distribution as defined in this dissertation.

Comparative evidence from related Tucanoan languages suggests that the morphological formatives which are involved in this unusual morphological phenomenon started as more run-of-the-mill gender agreement markers. In closely-related Koreguaje (Cook & Criswell 1993), for example, the forms appear simply in SG.M and SG.F contexts. In closely-related Secoya (Johnson & Levinsohn 1990) and Siona (Wheeler 1970), the forms appear in the 3SG.M and 3SG.F instead, but always consistently. Evidence from more distantly-related Tucano (West 1980) and Desano (Silva 2012) suggests that the latter distribution (i.e. 3SG gender markers) must have been the original one. The similarity of the suffixes to the 3SG.M and

3SG.F pronouns (e.g. *kũ* vs *ko* in Tucano) suggests that their incorporation and grammaticalization as gender/number markers lies at the ultimate source of the formatives.

It is at present unclear what the motivation might be for the innovation in Maijiki that caused the emergence of this morphomic structure. Areal influences, however, might constitute a promising avenue for explanation. This system (i.e. the change in the value of suffixes from declarative to interrogative) resembles conjunct/disjunct systems which are present in the area (e.g. in Barbacoan languages).

5.2.40 Malinaltepec Me'phaa (Suárez 1983)

As in other Oto-Manguenan languages, verbal inflection in Malinaltepec Me'phaa is complex. A tense prefix occurs first. As in the present tense below, tense prefixes tend to have an /a/-containing allophone in the singular and an /o/ or /u/-containing allophone in the plural. Next comes a 2SG prefix (which has many different allomorphs)⁷⁸ in a great number of verbs but not in all. After this comes the verb stem (which may or may not show alternations) and, finally, in many verbs (but, again, not in all), person-number agreement suffixes. These suffixes, even when they appear, are quite rich in syncretisms (e.g. 1PL and 2PL are always syncretic). Person clitics can be suffixed to the whole complex described so far in order to disambiguate the referent:

	Present		Past	
	SG	PL	SG	PL
1EXCL	na-ci:n	nu'-ci:n=so'	ni-ci:n	ni'-ci:n=so'
1INCL	-	nu'-ci:n=lo'	-	ni'-ci:n=lo'
2	na-ra-ci:n	nu'-ci:n=la	ni-ra-ci:n	ni'-ci:n=la
3	na'-ci:n	nu-ci:n	ni'-ci:n	ni-ci:n

Table 199: Some inflectional forms of 'play' in Malinaltepec Me'phaa (Suárez 1983: 122)

⁷⁸ The allomorphs are: *ta-*, *t-*, *tha-*, *ra-*, *tra-*, *štr-*, *šta*. As you may notice, these are all characterized by an alveolar stop (which has sometimes become /r/ as a result of sound change).

The morphological trait that is relevant in the context of the present discussion is that there are several irregular verbs in the language that show forms in 2SG+PL cells which are not present in the 1SG and 3SG. Consider the following paradigms:

	'carry' (whole form, past)		'close' (stem+suff.)		'throw' (stem)		'bathe' (stem)	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	ni-gongo:	ni-rango:=so'	rogo	ru'gwa	grwi'ya	dri'ya	wan	na
1INCL	-	ni-rango:=lo'	-	ru'gwa	-	dri'ya	-	na
2	ni-rango:	ni-rango:=la	rugwa	ru'gwa	dri'ya	dri'ya	na	na
3	ni-gongo:	ni-rango:	rogo	ru'gwa	grwi'ya	dri'ya	wan	na

Table 200: Some inflectional forms in Malinaltepec Me'phaa (Suárez 1983: 155, 158, 160)

As Table 200 shows, the forms involved can be diverse and include stems (from changes in stem-initial consonants or syllables all the way to suppletion) and sometimes suffixes (in inflection classes 5 [see 'close'] and 6). That stem alternation is not phonologically predictable is clear from the occurrence of different stems with the same affix.

The concrete changes by which these stem alternations emerged are not entirely clear to me at the moment but must involve the effects in the stem of both i) the 2SG agreement prefix present in a great number of verbs (see Table 199 and the 2SG prefix forms in Footnote 79) and ii) the back vowel allomorph of the tense prefixes that precede plural subjects (see the present in Table 199).

Alternations between velar stops in the singular and alveolar stops in the plural are found in some irregular verbs (e.g. SG *gu'ma* vs PL *tima*: 'be outside', SG *kra'mu*: vs PL *tra'ma*: 'be on top', Suárez 1983:159-160). In some other irregular verbs, there is a triple alternation between 1SG/3SG, 2SG and PL (e.g. *ganu*, *ja'nu*, *gwa'nu* 'arrive') or alternations for almost every paradigm cell. A common thread to some or most of these is that the 2SG (see Footnote 79) and the PL alternants are often characterized by alveolars opposed to velars in the rest of the paradigm. This situation, which probably came about by regular sound changes, may have been the reason for the occasional merger of some of those 2SG and PL forms into a single stem allomorph general to 2SG+PL.

5.2.41 Mazatec, Chiquihuitlan (Jamieson 1988, Feist & Palancar 2015)

It is common for Mazatec languages (Otomaguean) to display a morphological affinity of the 1SG and 3 (in both stems and agreement suffixes) and, to a smaller extent (only stems), of the converse cells 1PL and 2. Consider the following paradigms:

	'remember'		'forbid'		'scratch'	
	SG	PL	SG	PL	SG	PL
1EXCL	base	časin	tsičoʔo	ničoʔin	hentsun	čhentsin
1INCL	-	časen	-	ničoʔon	-	čhentsun
2	čase	časun	ničoʔe	ničoʔun	čhentsin	čhentsun
3	base	base	tsičoʔo	tsičoʔo	hentsun	hentsun

Table 201: Chiquihuitlán Mazatec verbs, positive, neutral aspect (Feist & Palancar 2015)

Table 201 above shows how 1SG+3 often share a stem opposed to the one in 2+1PL. These formally diverse alternations originate from a system of auxiliaries, many of which already showed these unnatural morphological affinities, that became prefixed to the main verbs (see Baerman 2013 and Pike 1948). In around 90% of the verbs, 1SG and 3 are whole-word syncretic, since they also share their person-number suffix, as in the paradigms above. Other syncretisms (e.g. between 1PL, 2SG and 2PL) are less systematic.

Mazatec1: 1SG/3

Mazatec2: 2/1PL

5.2.42 Mehri (Rubin 2010)

As in other Semitic languages, the verbal conjugation of Mehri is characterized by the heavy use of vowel apophony upon a more-or-less invariable consonantal skeleton. There is, in the perfect, a syncretism of the third singular masculine and the third plural feminine. Consider the following paradigms:

	‘put on the fire’			‘break’		
	SG	DU	PL	SG	DU	PL
1	arákb-ək	arákb-əki	arákb-ən	təbr-ək	təbr-əki	təbr-ən
2M	arákb-ək	arákb-əki	arákb-əkəm	təbr-ək	təbr-əki	təbər-kəm
2F	arákb-əš	arákb-əki	arákb-əkən	təbr-əš	təbr-əki	təbər-kən
3M	arökəb	arkəb-ē	arákb-əm	t̥ɪbər	təbr-ō	təbr-əm
3F	arkəb-ēt	arkəb-tē	arökəb	təbr-ūt	təbər-tō	t̥ɪbər

Table 202: Perfect paradigms of two Mehri verbs (Rubin 2010:91,94)

Affixally, both cells are characterized merely by the absence of an affix, which would not qualify as a formal identity here. The two forms, however, also behave alike in every verb concerning ablaut, sometimes, as Table 202 illustrates, sharing a form to the exclusion of every other paradigm cell.

5.2.43 Menggwa Dla (de Sousa)

In the Papuan language Menggwa Dla (Senagi), also known as Dera, a few verbs display a stem alternation pattern that is phonologically and morphosyntactically unmotivated. Consider the following paradigm:

	SG	DU	PL
1	numb-ahahwa	numb-ehyahwa	numb-efahwa
2M	numb-afahwa	numb-afahwa	nung-umahwa
2F	numb-afahwa	numb-efyahwa	numb-eihwa
3M	nung-uhwa	numb-afahwa	nung-umahwa
3F	nung-wahwa	numb-efyahwa	numb-eihwa

Table 203: Menggwa Dla 'stand' past (de Sousa 2006:541)

As Table 203 illustrates, the 3SG and 2/3PL.M cells show a stem alternant different from the one found in the rest of the paradigm. Notice that these cells are also characterized by suffixes which begin with a high back vowel. Although this differential phonological context (i.e. front vs back vowel) may have been the origin of this pattern, the alternation is not phonologically derived, because /g/ and /b/ are fully-fledged phonemes that can both appear in all phonological environments synchronically (cf. *yangifi* /jagiϕi/ [jaŋgiβi] ‘wake (someone) up’, *ambuha* /abuxa/ [ʔambuɣa] ‘cockatoo’). The pattern is clearly morphological in nature and also systematic, since the forms involved can be even suppletive:

	SG	DU	PL
1	s-ahaambi	s-ehihwaambi	s-efuhuambi
2M	s-afuambi	s-af(ani)naambi	ah -umuwuambi
2F	s-afuambi	s-ef(ya)naambi	s-eihiambi
3M	ah -yaambi	s-af(ani)naambi	ah -umuwuambi
3F	ah -yaambi	s-ef(ya)naambi	s-eihiambi

Table 204: Menggwa Dla 'think/call' present (de Sousa 2006:541)

As the above paradigm illustrates, in the verb ‘think/call’, the stem *ah-* appears in that same paradigmatic environment even in the absence, sometimes, of the back-vowels that appeared in those cells’ suffixes in the earlier paradigm. Other suppletive alternants with this distribution include *eh-* (vs *s-* ‘talk’) and *ap-* (vs *e-* ‘sleep’).

5.2.44 Mian (Fedden 2011)

Gender agreement in Mian (Trans-New-Guinea) is similar to that in other languages presented here before like Khinalug (Section 5.2.34) and Burmeso (5.2.11). Thus, the same agreement affixes are required by a class of nouns in the singular, by another class in the plural, and by a third class in both singular and plural:

	Subject		Direct Object		Indirect Object IPFV	
	SG	PL	SG	PL	SG	PL
M	-e	-ib	a-	ya-	-ha	-ye
F	-o	-ib	wa-	ya-	-we	-ye
N1	-e	-o	a-	wa-	-ha	-we
N2	-o	-o	wa-	wa-	-we	-we

Table 205: Gender-number agreement affixes in Mian (Fedden 2011:163)

Feminine singulars, neuter 1 plurals, and neuter 2 nouns behave all as a single class in terms of agreement. As Table 205 shows, the agreement formatives take on a different form in different grammatical roles, so this pattern of morphological identity is systematic.

Although what we know about the history of these languages is not enough, there are, as explained by Fedden (2011:168-169), plausible ways in which these systems can emerge diachronically. In a typological parallel he mentions:

*It is well-known that for some classical daughter languages of Proto Indo-European (PIE), suffixes in the feminine singular (nominative) and the neuter plural (both nominative and accusative) are identical, namely -a; e.g. Latin femin-a 'woman' (feminine singular); don-a 'presents' (neuter plural). An account for this homophony is that in early PIE and pre-IE, neither of which had a category 'gender', there was a single collective form marked with *-h which expressed low individuation later developing into the feminine singular and the neuter plural form. The marker *-h was (among others) in opposition to *-s, which had an individualizing force and a specific meaning (cf. Lehmann 1958: 189-90) and later became the masculine form. Similarly, in Mian the masculine marker =e is used to refer to individual, singular objects (whether animate or inanimate), whereas the feminine marker =o is associated with a collective meaning. (Fedden 2011:168-169)*

5.2.45 Murrinh-Patha (Mansfield n.d., Nordlinger 2015, Walsch 1976)

Murrinh-Patha (Southern Daly, Australia) verbal inflection is extraordinarily complex. For one, it is the only language to date reported to have an inflectional siblinghood category (Nordlinger 2015:501). What concerns us here is that the expression of this category interacts with number (SG, DU, PC and PL) in an idiosyncratic way. The suffixes for non-sibling (masculine or feminine) apply to the form of the verb that is otherwise used for the number form immediately lower to the value they actually express. That is, dual non-sibling suffixes attach to the otherwise singular form and paucal non-sibling suffixes attach to what is otherwise the dual form. The misadjustment of this category effectively means that all person-number forms have an unnatural distribution in the Murrinh-Patha paradigm. Consider the following:

	1EXCL ⁷⁹		2		3	
	SIBLING	N-SIBLING	SIBLING	N-SIBLING	SIBLING	N-SIBLING
SG	ɲem		t̥im		dim ⁸⁰	
DU	ɲarimka	ɲem-ɲinda ⁸¹	nirimka	t̥im-ɲinda	karimka	dim-ɲinda
PC	ɲarim	ɲarimka-ɲime	nirim	nirimka-ɲime	karim	karimka-ɲime
PL	ɲarim		nirim		karim	

Table 206: Perfect paradigm of 'sit' (Walsch 1976:327)

As the above Table 206 shows, almost every person-number exponent in the language adopts a paradigmatic configuration that is unnatural with respect to number marking. Some forms (e.g. the /di/ in bold in the paradigm above, but also forms like /t̥i/ and /ɲe/) appear, within a given person, in the singular and the dual non-sibling. Other forms (e.g. the shaded /ka/ but also /ɲa/ and /ni/) appear, within a given person, in the opposite set of cells, that is, in the

⁷⁹ The inclusive forms are not represented in this paradigm because they are not sensitive to the same number distinctions as other forms.

⁸⁰ The form *dim* indicates proximity. It is replaced by *kem* to signal a greater distance. For reasons of space, only proximate forms are displayed here.

⁸¹ For reasons of space, only feminine forms are given. Masculine forms are only used with groups made up exclusively of males and thus the feminine can be thought of as the default.

dual sibling and in the paucal and plural. Other forms (e.g. /ri/ in Table 206 above) are not limited to a particular person but appear in the ‘larger number’ region of the paradigm across all persons. Note, however, that the form /ri/ also appears in the opposite side of the paradigm in other verbs (e.g. in the past tense of ‘stand’, it is the SG+DU.NSIB that are marked with that suffix, which is then absent from the rest of the cells, see Mansfield n.d.:5).

The association of concrete exponents with particular morphemes is, thus, also not always straightforward. Because of this ‘misplaced’ number morphology in non-sibling forms, only tense forms are semantically well-behaved. Thus, the perfect marker *-m* in the Table 206 above is opposed to zero in the future and to *-ni/-ne* in the imperfect).

The morphomic categories described here can be instantiated by many different forms depending on the person and the particular tense or verb/conjugation. SG+DU.NSIB can be instantiated by person-specific forms like 1 /ŋe/, /ŋa/, 2 /t̥i/, /d/, /n/, 3 /di/, /w/, /j/ etc. or by person-indifferent forms like /ri/, /r/, /n/, /l/. DU.SIB+PC+PL, in turn, can also be instantiated by either person-specific forms like 1 /ŋa/, /ŋ/, 2 /n/, 3 /p/, /k/, /ka/ or by person-indifferent forms like /ri/, /ra/, /je/, /ɹ/, /ŋ/, /nn/, /ll/, or /dd/.

Murrinh-Patha1: SG/DU.NSIB

Murrinh-Patha2: DU.SIB/PC/PL

5.2.46 Ngkolmpu (Carroll 2016)

The Papuan language Ngkolmpu (Yam) is characterized by a very complex verbal morphology whose mapping into morphosyntactic values is often notoriously complicated. For the purposes of the present discussion, the undergoer prefixes⁸² are particularly interesting. These are illustrated below in the paradigm of the copula:

⁸² “The undergoer prefix indexes O arguments, S arguments in the intransitive construction and R arguments in the recipient-indexing ditransitive construction and the benefactive applicative” (Carroll 2016:134). There are several sets of prefixes used with different TAMs. These are referred to as ‘series’ (α , β and γ) in the literature.

	Hodiernal past		Imperative-hortative		Future-irrealis	
	SG	PL	SG	PL	SG	PL
1	u-rei	n-rei	b-ront	kn-ront	b-ront	nt-ront
2	n-rei	y-rei	kn-ront	s-ront	nt-ront	s-ront
3	y-rei	y-rei	s-ront	s-ront	s-ront	s-ront

Table 207: Three tense subparadigms of the copula in Ngkolmpu (Carroll 2016:245)

The form of the undergoer prefixes, as Table illustrates, changes according to person and number. Two of the three forms distinguished, however, are not aligned to any particular value. The 2SG and 1PL are always syncretic, and so are 3 and 2PL. The syncretisms are instantiated by different allomorphs depending on the particular TAM.

Ngkolmpu1: 2SG/1PL

Ngkolmpu2: 3SG/2PL/3PL

5.2.47 Nen (Evans 2015)

Although these syncretisms are systematic in Ngkolmpu (because they always hold and are repeated under several allomorphs), this is not so in related Yam languages:

	Nen (Evans 2015:548)						Komnzo (Döhler 2018.:238)					
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	w-	y-n-	q-	t-n-	ḡ-	d-n-	wo-	n-	kw-	nzn-	zu-	nzn-
2	n-	y-a-	kn-	t-a-	gn-	d-a-	n-	e-	gn-	th-	nzn-	th-
3	y-	y-a-	t-	t-a-	d-	d-a-	y-	e-	s-	th-	s-	th-

Table 208: Undergoer prefixes in Nen and Komnzo

As Table 208 illustrates, 2SG and 1PL are not always syncretic beyond Ngkolmpu. Similarly, 3SG and 2/3PL are usually not syncretic either. Although it is, at present, not entirely clear which diachronic developments one should assume, the syncretism of 2SG and 1PL seems to go all the way back to Proto-Yam. That of 2/3PL and 3SG is less clear. In one of the series, these two cells are reconstructed by Evans et al. (2017:760) as two different formatives which fell together in Ngkolmpu probably because of sound change ($\theta > s$).

Although it is difficult to be sure about the details, it seems that while Ngkolmpu appears to have systematized the (partially inherited) unmotivated syncretisms, other languages have evolved towards more well-behaved paradigms with less syncretism. Consider, for example, the extension of the 3/2PL morphology to the 1PL in Nen, which effectively prevents the expected syncretism of that cell with the 2SG. This seemingly newly-acquired morphological affinity of PL+3SG in Nen should also be regarded as morphomic according to my present criteria and has been included in the database.

5.2.48 Nimboran (Anceaux 1965, Inkelas 1993)

The Nimboran language (Nimboranic, New Guinea) is well known for its baroque verbal complex. The most interesting feature for our current purposes is the variety of stem alternations, which appear to correlate, although imperfectly, with the marking of number. Three stems are distinguished, whose distribution also matches that of certain suffixes:

	SG	DU	PL
1EX	ŋgedúo-d-u	ŋgedóu-ke-d-ú	ŋgedói-d-i-u
1INC	-	ŋgedúo-man-d-ám	ŋgedóu-ke-d-ám
2	ŋgedúo-d-e	ŋgedóu-ke-d-é	
3M	ŋgedúo-d-am	ŋgedóu-ke-d-ám	ŋgedói-d-i-am
3N	ŋgedúo-d-um	ŋgedóu-ke-d-úm	

Table 209: Nimboran ‘draw’, unspecified object, momentary, future (Anceaux 1965:186)

As the Table 209 above illustrates, what Anceaux labels ‘singular stem’ occurs in 1SG, 2SG and 3SG but also in 1+2 (i.e. in the 1DU inclusive). The so-called ‘dual’ stem, and also the suffix *-ke*, in turn, occur in 1DU.EXCL, 2DU and 3DU, but also in 1PL.INCL and, maybe more surprisingly, in the 2PL. The ‘plural’ stem, in turn, can occur with 1EX and 3 but, crucially, not with 1INCL or 2. These last facts are crucial to regard this system as unmistakably morphomic since, although it reminds of a minimal-augmented number system, a restructuring of the above paradigm in those terms would not solve the form-meaning mapping maladjustments in Nimboran, since the 2PL form *ηgedóukedé* (instead of the expected **ηgedóidie*) makes the so-called ‘dual’ stem morphomic.

Stem alternations are formally diverse (e.g. *suáj*(SG) *sáoj*(DU) *saóij*(PL) ‘water’, *ηgeduá*(SG) *ηgedáu*(DU) *ηgedói*(PL) ‘shave’) and found in a majority of verbs. They tend to involve stress and vowel changes on the right edge of the stem, maybe originating from anticipatory assimilations to the following number suffixes. The original number-marking function of this morphology is clear. It is revealing, in this respect, that, in the durative aspect and with plural objects, the paradigmatic distribution of these stems is “shifted to the left” as it were:

	SG	DU	PL
1EX	ηgedóu-t-emné-y	ηgedói-t-i-emné-y	
1INC	-	ηgedóu-t-emené-m	ηgedói-t-i-emné-m
2	ηgedóu-t-emné-i	ηgedói-t-i-emné-i	
3M	ηgedóu-t-emné-m	ηgedói-t-i-emné-m	
3M	ηgedóu-t-emné-m	ηgedói-t-i-emnyé-m	

Table 210: Verb ‘draw’ in Nimboran, durative, present (Anceaux 1965:236)

As Table 210 shows, the earlier dual stem occurs now in the singular, and the earlier plural stem has spread to the dual. In any case, the synchronic distribution of the so-called “dual”⁸³ and “plural” stems in Nimboran is synchronically morphomic.

⁸³ The dual stem is sometimes regarded as a default in the literature. This theoretical status may be derived from the greater formal and distributional diversity of this stem compared to the others.

Research into other languages in the family has been sparse but it seems that some of the morphomic affinities that exist in Nimboran might also be present elsewhere, even though with a somewhat different distribution in the paradigm. In Kemtuk (Wilden 1976:73-74), for example, the “dual” suffix *-ke* that we saw in Table 209, is used in the same contexts as in Nimboran except for the 1PL.INCL, which shares form (*-i*) with the 1PL.EXCL instead .

Nimboran1: DU.M/2PL.M/SG.D

Nimboran2: 1PL.M/3PL.M/NSG.D

5.2.49 Nivkh (Gruzdeva 1998, Nedjalkov & Otaina 2013)

Some verbal forms in Nivkh (Isolate, Russia) agree with their subject. These formatives (manner converbs, temporal converbs and finite forms, see Gruzdeva 1998:55), however, can take only two forms and the values with which they occur do not constitute a natural class. Consider their distribution:

	Non-future						Future					
	Narrative		Distant		Coordinating		Narrative		Distant		Coordinating	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	-t	-t	-tot	-tot	-ta	-ta	-n	-n	-non	-non	-na	-na
2	-r	-t	-ror	-tot	-ra	-ta	-r	-n	-ror	-non	-ra	-na
3	-r	-t	-ror	-tot	-ra	-ta	-r	-n	-ror	-non	-ra	-na

Table 211: Nivkh (East Sakhalin) converb inflection (Nedjalkov & Otaina 2013: 40-42)

As Table 211 above illustrates, the first person singular and the plural subjects occur with the same forms. The form of this 1SG+PL suffix varies (/t/ vs /n/) according to tense and is thus systematic. The diachronic origin of these alternations might be sound change. In a way similar to Celtic mutations, morphologized consonant alternations (between voiced stops, voiceless stops, and fricatives) occur frequently at word and morpheme boundaries in Nivkh.

The alternation between /t/ and /r/ is part of this bigger system in the language (Nedjalkov & Otaina 2013:15-16). In synchrony, however, the alternations between the forms in Table 211 do not correlate to different phonological environments, as all of them simply follow the verb stem synchronically. The pattern is, therefore, morphomic.

5.2.50 North Saami (Hansson 2007)

The variety of North Saami spoken in Eastern Finnmark has a systematic diagonal syncretism between comitative singular and locative plural. Consider the following partial paradigms:

	'house'		'who'	
	SG	PL	SG	PL
LOC	viesu-s	viesu-in	gea-s	gea-inna
COM	viesu-in	viesu-iguin	gea-inna	gea-iguin

Table 212: Two partial paradigms in East Finnmark North Saami (Hansson 2007:25,28)

The syncretism in 'house' and other polysyllabic stems (i.e. with the formative /jn/) happens in various other Saami varieties and might even be reconstructible for the Proto-language. The syncretism in monosyllabic stems like 'who', by contrast, is a local analogical innovation that has extended what was originally the COM.SG form to the LOC.PL as well on the basis of the large class of polysyllabic nouns where the two cells were initially syncretic.

It is worth noting that the /i/s in these two *-in* suffixes could potentially lend themselves to different segmentations. One may feel justified to segment one as an inseparable part of the comitative singular suffix (*-in*) and the other as a recurrent plural suffix, which would be followed in this particular cell by a LOC.PL formative (*-i-n*). Because of this, Feist (2015:137) even refers to this as a syncretism that is only "apparent". It is therefore surprising to see that despite the availability of potential cues that this is an accidental (as opposed to a systematic) homophony, the two cells lead parallel lives in North Saami and elsewhere:

	SG	PL
ILL	kuasska	kūs'ket'
LOC	kuas'kes't	kūs'ken'
COM	kūs'ken'	kūs'keguejm
ABE	kues'kxa	kūs'kexa

Table 213: Partial paradigm of Kildin Sami *kuess'k* 'aunt' (Rießler forthcoming)

As illustrated in Table 213, in varieties with morphological stem alternation, the stem in the two cells is usually the same, even when this means (as in COM.SG *kūs'k-* above) deviating from a more natural distribution. Whole-word syncretism, thus, is usually maintained even in the presence of various non-linear inflectional devices (i.e. consonant gradation or vowel apophones) that could disrupt it. This is suggestive of a systematic morphological identity.

5.2.51 Northern Akhvakh (Creissels 2008)

The perfective positive suffixes in Northern Akhvakh (Nakh-Daghestanian) are characterized by allomorphy structured along various orthogonal axes. Observe the following paradigms:

	Conjunct		Disjunct	
	SG	PL	SG	PL
M	w-uᵅ-ada	ba-ᵅ-idi	w-uᵅ-ari	ba-ᵅ-iri
F	j-iᵅ-ada	ba-ᵅ-idi	j-iᵅ-ari	ba-ᵅ-iri
N	b-iᵅ-ada	r-iᵅ-ada	b-iᵅ-ari	r-iᵅ-ari

Table 214: Perfective positive paradigm of 'grasp' (Creissels 2008)

	Conjunct		Disjunct	
	SG	PL	SG	PL
M	gw-ēda	guj-idi	gw-ēri	guj-iri
F	gw-ēda	guj-idi	gw-ēri	guj-iri
N	gw-ēda	gw-ēda	gw-ēri	gw-ēri

Table 215: Perfective positive paradigm of 'do' (Creissels 2008)

As corresponds to a conjunct/disjunct system, the allomorph with /d/ appears in the first person in statements and in the second in questions, and the allomorph in /r/ elsewhere. This is a simplification but it is inconsequential for our current discussion. This allomorphy is understood to be related to the epistemological properties of speech act participants in particular speech acts and is thus not morphomic (see Section 2.9.3).

Each of those allomorphs, however, is in turn subject to various allomorphies in Northern Akhvakh. As illustrated in Table 214 and 215, the gender/number of the absolutive argument determines the concrete form to be used. Singular and neuter plural arguments occur with the same /a/-based allomorph, whereas masculine and feminine plural use a different /i/-based form. This is not the end of the allomorphy, however, as the allomorphs *-ada* and *-ari* that occur in SG+N.PL also show allomorphic differences between lexical items. In some vowel-final stems like ‘do’, for example, those vowels have blended with the suffix-initial /a/ (i.e. i+a=e:) to yield further vocalic allomorphy.

5.2.52 Old English

Germanic languages are well known for their stem vowel changes. These correlate to some extent to whole categories like past vs present. Some other times, however, a stem’s distribution is not so straightforward. This is particularly prevalent in more conservative (e.g. Icelandic) and in older Germanic offshoots. Consider the following Old English paradigms:

	Present				Past			
	Indicative		Subjunctive		Indicative		Subjunctive	
	SG	PL	SG	PL	SG	PL	SG	PL
1	sēoþ-e	sēoþ-aþ	sēoþ-e	sēoþ-en	sēaþ	sud-on	sud-e	sud-en
2	sīeþ-st				sud-e			
3	sīeþ-þ				sēaþ			

Table 216: Paradigm of *sēoþan* ‘boil’

As the Table 216 above illustrates, the same stem usually occurs in Old English (also in the other older West Germanic languages) in the past subjunctive, past indicative plural and past indicative 2SG. The actual forms involved are diverse. They can be consonant alternations resulting from Werner’s Law (/d/ [vs /θ/], /s/ [vs /r/], /h/ [vs /g/]), as well as stem vowel alternations (i u æ) derived largely from the Proto-Indo-European zero grade opposed to other grades elsewhere in the paradigm. The original distribution of these alternants did not include the 2SG past indicative. Their extension to this cell is a West Germanic innovation that happened, probably, when a subjunctive form replaced the original 2SG.PAST indicative.

5.2.53 Páez (Jung 1989)

In the Colombian isolate Páez the 2SG feminine and the 2PL are always syncretic. This is so in every single TAM and across various exponents. Consider the following suffixes:

	Declarative		Interrogative	
	SG	PL	SG	PL
1	-thu	-thaʔw	-tka	-tkhaʔw
2M	-gu	-iʔkwe	-ga	-kwe
2F	-iʔkwe	-iʔkwe	-kwe	-kwe
3	-aʔ	-taʔ	-kha	-ta

Table 217: Two suffix sets in Páez (Jung 1989:124)

The cells in Table 217 above show the same unnatural syncretism. Despite the formal diversity, one can spot a segment sequence common to all of the formatives instantiating this morphomic category. When the same pattern is found in the imperative, for example, the corresponding suffix is *-we* (e.g. *mdex* ‘sleep.2SG.M’ vs *mdex-we* ‘sleep.2SF.F/2PL’ (Jung 1989:134). Thus, although different tenses instantiate the 2SG.F+2PL syncretism with different formatives, all involve adding segments to an invariable sequence, i.e. ([iʔ]k)we.

5.2.54 Pite Saami (Wilbur 2014)

Saami languages (Uralic) are well-known for their intricate stem alternation patterns in both verbal and nominal inflection. Several sound changes in the history of the family (most notably consonant gradation [see Gordon 2009] and various vowel assimilations) have introduced allomorphy in the stem. These alternations were initially connected to particular phonological contexts but became subsequently morphologized when the conditioning environments disappeared as a result of later sound changes. As a result of these processes, non-concatenative morphology is prominent in Saami and various patterns qualify here for morphomic status. Consider the following nominal paradigms:

	<i>luakkta</i> ‘bay’		<i>bärrgo</i> ‘meat’	
	SG	PL	SG	PL
NOM	luakkta	luokta	bärrgo	biergo
GEN	luokta	luoktaj	biergo	biergoj
ACC	luoktav	luoktajd	biergov	biergojd
ILL	luokktaj	luoktajda	bärrgoj	biergojda
INESS	luoktan	luoktajn	biergon	biergojn
ELAT	luoktast	luoktajst	biergost	biergojst
COM	luoktajn	luoktaj	biergojn	biergoj
ABESS	luoktadak	luoktadaga	biergodak	biergodakta
ESS	luokktan	luokktan	bärrgon	bärrgon

Table 218: Two nominal paradigms of Pite Saami (Wilbur 2014:96,101)

As Table 218 illustrates, the strong grade⁸⁴ of the stem, and also a different stem vowel (/wa/ [vs /o/] and /ε/ [vs /e/]) appear in nominative and illative singular, and in the essive, whose singular and plural forms are the same. This constitutes, thus, a morphosyntactically unnatural morphological affinity.

⁸⁴ Strong grade in Pite Sami most usually involves gemination with respect to the weak grade (as in the examples) but can it also involve adding a segment, more concretely /t/, /p/ or /k/ (e.g. /va:jmo/ ‘heart.NOM.PL’ vs /va:jpmo/ ‘heart.NOM.SG’).

Nominal declension can also show a different morphological pattern in addition to this one. This other alternation involves vowel apophonies different from the ones that participated in the previous morpheme. In this case, we are dealing with vowel raisings which include the following: /e/ > /i/, /o/ > /u/, /a:/ > /ε/, /a:/ > /i/, /ɔ/ > /u/, /a/ > /ε/, /a/ > /e/ and /a/ > /i/. Consider the following paradigms:

	<i>guolle</i> 'fish'		<i>vágge</i> 'valley'	
	SG	PL	SG	PL
NOM	guolle	guole	vágge	vágge
GEN	guole	gulij	vágge	vággij
ACC	guolev	gulijd	vággev	vággijd
ILL	guolláj	gulijda	vággáj	vággijda
INESS	guolen	gulijn	vággen	vággijn
ELAT	guolest	gulijst	vággest	vággijst
COM	gulijn	gulij	vággijn	vággij
ABESS	guoledak	guoledaga	vággedak	vággedaga
ESS	guollen	guollen	vággen	vággen

Table 219: Two nominal paradigms of Pite Saami (Wilbur 2014:101)

As the paradigms above illustrate, a high-vowel stem appears in various cases in the plural and also in the comitative singular. These patterns originate, as is probably apparent from the synchronic form of the suffixes, by means of anticipatory assimilation to a following high vowel /i/. It must be stressed, however, that, unlike the paradigms above might suggest, it is not possible synchronically to identify a phonological context where these forms occur or to consistently derive one vowel from the other (Wilbur 2014:79). Consequently, this pattern must be classified here as morphomic.

Turning to the verbal domain we also find the morphological vestiges of the same sound changes that produced alternations in nominal declension. Regarding the first of these processes, i.e. consonant gradation, consider the following paradigm:

	PRES			PAST		
	SG	DU	PL	SG	DU	PL
1	vies-ov	viess-on	viess-op	viess-ov	vies-ojmen	vies-ojme
2	vies-o	viess-obähten	viess-obähtet	viess-o	vies-ojden	vies-ojde
3	viess-o	viess-oba	viess-o	vies-oj	vies-ojga	viess-on

Table 220: Pite Sami *viessot* 'live' (Wilbur 2014: 172)

The same as in nouns, the strong grade also may occur along with stem vowel apophony (/wa/ [vs /o/] and /ε/ [vs /e/]). The one shown in Table 220 is the distribution of the strong grade in all Pite Sami verbs that show gradation. Vowel raising, however, shows a different picture, as there are two classes of verbs where raising differs on its paradigmatic extension:

	PRES			PAST		
	SG	DU	PL	SG	DU	PL
1	bas-av	biss-in	bass-ap	biss-iv	bas-ajmen	bas-ajmä
2	bas-a	bass-abähten	bass-abähtet	biss-e	bas-ajden	bas-ajdä
3	bass-a	bass-aba	biss-e	bas-aj	bas-ajga	biss-in

Table 221: Pite Sami *bassat* 'wash' (Wilbur 2014: 174)

	PRES			PAST		
	SG	DU	PL	SG	DU	PL
1	bas-áv	biss-in	bass-ep	biss-iv	bis-ijmen	bis-ijmä
2	bas-á	bass-ebähten	bass-ebähtet	biss-e	bis-ijden	bis-ijdä
3	bass-a	bass-eba	biss-e	bis-ij	bis-ijga	biss-in

Table 222: Pite Sami *basset* 'fry' (Wilbur 2014: 174)

In the first of these classes (Table 221), vowel raising applies to 1DU.PRES, 3PL.PRES, 1SG.PAST, 2SG.PAST and 3PL.PAST. It must be noted that this set of cells is a subset of the cells with stems in the strong grade. In this way, its intersection with it only generates three as opposed to four stem alternants. Notice how the weak+high stem *bis-* does not occur.

This might well be a desirable trait in morpheme interactions (Herce 2019) but does not extend to the other verbal class (Table 222). Here, vowel raising applies to a superset of the cells where it applied in *bassat* because it extends to the entirety of the past tense. These two different distributions of raising in the past tense are also found in other Saami varieties (e.g. North Sami, see Kahn & Valijärvi 2017) and may be conceived to be stable due to their use of two different types of morphological niches: a formal one (i.e. the strong consonant grade) in *bassat* and (partially) a semantic one (i.e. past) in *basset*.

Pite Saami1: NOM.SG/ILL.SG/ESS

Pite Saami2: COM.SG/GEN.PL/ACC.PL/ILL.PL/INESS.PL/ELAT.PL/COM.PL

Pite Saami3: 3SG.PRES/DU.PRES/PL.PRES/1SG.PAST/2SG.PAST/3PL.PAST

Pite Saami4: 1DU.PRES/3PL.PRES/1SG.PAST/2SG.PAST/3PL.PAST

Pite Saami5: 1DU.PRES/3PL.PRES/PAST

5.2.55 Skolt Saami (Feist 2015)

Skolt Sami's morphomic structures are similar to those in Pite Saami. In the verbal domain, however, there are a few differences which are relevant for the present discussion. One is the loss of the dual. Since a whole value (i.e. a column of cells) has disappeared, the paradigmatic profile of the alternations has been modified, even in the absence of changes in the surviving cells. The other one is the emergence of qualitative consonant gradations. Some alternations which were originally quantitative (e.g. p: vs p, t: vs t) have become qualitative (e.g. p: vs v, t: vs ð) in Skolt Saami. Consider the following paradigm:

	Present		Past	
	SG	PL	SG	PL
1	njoor[ɣ]-am	njorgg-ap	njurgg-em	njoor[ɣ]-im
2	njoor[ɣ]-ak	njorgg-e'ped	njurgg-ič	njoor[ɣ]-id
3	njorgg	njorgg-a	njoor[ɣ]-i	njurgg-e

Table 223: Inflectional paradigm of *njorggad* 'whistle' (Feist 2015:204,210)

In the paradigm above, the weak grade (/ɣ/) appears in 1SG and 2SG present and in 3SG, 1PL and 2PL past. The strong grade (/g:/) appears in the rest of the paradigm. The paradigmatic distribution of the two forms is, therefore unnatural. In addition to this, as Table shows, the paradigmatic distribution of vowel raising is different in Skolt and Pite Sami. In Skolt Saami, it appears exclusively in the past tense:

	Present		Past	
	SG	PL	SG	PL
1	njoor[ɣ]-am	njorgg-ap	njurgg-em	njoor[ɣ]-im
2	njoor[ɣ]-ak	njorgg-e'ped	njurgg-ič	njoor[ɣ]-id
3	njorgg	njorgg-a	njoor[ɣ]-i	njurgg-e

Table 224: Inflectional paradigm of *njorggad* 'whistle' (Feist 2015:204,210)

In some Skolt Saami verbs, as Table 224 shows, vowel raising appears in 1SG, 2SG and 3PL past. This is morphosyntactically unnatural and it contrasts with the distribution of raising in Pite Saami (see Table 221), where it also occurred in two cells in the present. In other Skolt Saami verbs, in the same way as in Pite Saami (see Table 222), raising extends to all the past cells (Feist 2015:209). Due to its confinement to the past, this alternation has become semantically motivated in this class of verbs and does not classify as morphomic here.

Skolt Saami1: 1SG.PRES/2SG.PRES/3SG.PAST/1PL.PAST/2PL.PAST

Skot Saami2: 3SG.PRES/PL.PRES/1SG.PAST/2SG.PAST/3PL.PAST

Skolt Saami3: 1SG/2SG/3PL

5.2.56 Sobei (Sterner 1987)

In the Oceanic language Sobei, some 20 verbs show a stem vowel apophony in their person-number infection in the present tense. Consider the following paradigms:

	'slide'				'come'			
	Realis		Irrealis		Realis		Irrealis	
	SG	PL	SG	PL	SG	PL	SG	PL
1EX	i-tosis	me-tosis	ye-tasis	'e'-tasis	yo-mi	mi-mi	i-ma	'a'-ma
1INC	-	te-tosis	-	te-tasis	-	ti-mi	-	ta-ma
2	u-tosis	me-tosis	a-tasis	'e'-tasis	u-mi	mi-mi	a-ma	'a'-ma
3	e-tasis	re-tosis	a-tasis	rie-tasis	e-ma	ri-mi	a-ma	ria-ma

Table 225: Partial paradigm of two Sobei verbs (Sterner 1987:41,43)

As Table 225 shows, in the 1SG, 2SG and PL of the Realis, the stem vowel is different from the one found elsewhere in the paradigm. This happens with only a few verbs and with the forms /o/ (vs /a/, e.g. 'slide' above), /i/ (vs /a/, e.g. 'come' above) and /i/ (vs /ei). The forms and paradigmatic distributions involved mean that both parts of the paradigms qualify for morphomehood according to the requirements that have been pre-established here.

Sobei1: 1SG/2SG/PL

Sobei2: 3SG.R/I

5.2.57 Spanish (personal knowledge)

Romance languages are well known for being the family where morphomic stem alternation patterns have been more thoroughly studied (see e.g. Maiden 2018b). Spanish will be taken here as a representative of two of the most often discussed ones: the N-morphome and the L-morphome. Consider the former:

	Present Indicative		Present Subjunctive	
	SG	PL	SG	PL
1	pierdo	perdemos	pierda	perdamos
2	pierdes	perdéis	pierdas	perdáis
3	pierde	pierden	pierda	pierdan

Table 226: Present tense paradigm of Spanish *perder* 'lose'

As Table 226 shows, a diphthong (i.e. /je/ vs /e/) appears, in the present, in the singular and 3PL cells. That form is absent from the rest of the paradigm. In other verbs (e.g. *poder* 'be able to'), the alternation /we/ vs /o/ has the exact same paradigmatic distribution. The presence of the diphthong coincides with the location of stress in the stem. Note, however, that stress is free in Spanish. Consider now the L-morphome:

	Present Indicative		Present Subjunctive	
	SG	PL	SG	PL
1	caigo	caemos	caiga	caigamos
2	caes	caéis	caigas	caigáis
3	cae	caen	caiga	caigan

Table 227: Present tense of Spanish *caer* 'fall'

As the Table 227 above illustrates, some Spanish verbs show a different stem in the 1SG indicative and in the present subjunctive. Most often (e.g. *caig-o* vs *ca-es* 'fall', *pare[θk]-o* vs *pare[θ]-es* 'seem') the stem has a velar extension absent from the rest of the paradigm. In one verb (*quep-o* vs *cab-es* 'fit') the alternation is quasi-suppletive. As these verbs illustrate, the formal alternations involved in the L-morphome are, thus, varied.

Other Romance varieties closely related to Spanish have similar paradigmatic alternations. An interesting one, cognate with the one in Table 226 but with a different paradigmatic configuration, is one present in a dialect of the west of Asturias:

	Present Indicative		Present Subjunctive	
	SG	PL	SG	PL
1	'mordo	mur'demos	'morda	'mordamos
2	'mwerdes	mur'deis	'mordas	'mordais
3	'mwerde	'mwerden	'morda	'mordan

Table 228: Present tense paradigm of *murder* 'bite' in western Asturias (Bybee 1985:73)

Diphthongization occurs in this variety, in some 35 verbs, in the 2SG, 3SG, and in the 3PL of the present indicative. Some of these (e.g. *ferber* 'boil' have the other diphthong (i.e. /je/) with the same paradigmatic configuration, which makes this pattern morphomic as defined here. The diachronic origin of this alternants is to be found in the interaction between the two morphemes that have been described above for Spanish. These cells constitute, precisely, those that participated in the N-morpheme allomorphy but not in the L-morpheme one (see Herce 2019 for more details).

Spanish1: SG/3PL

Spanish2: 1SG.IND/SBJV

Spanish3 (Asturian):2SG/3SG/3PL

5.2.58 Sunwar (Borchers 2008)

Like other Western Kiranti languages (Sino-Tibetan), Sunwar shows morphological stem alternation in some of its verbs. Consider the following paradigm:

	SG	DU	PL
1	ma-jog-u	ma-jo-sku	ma-jo-ka
2	ma-jog-i	ma-jo-si	ma-jo-ni
3	ma-jog-a	ma-joga-se	ma-joga-me

Table 229: Paradigm of Sunwar 'understand', neg. past (Borchers 2008:200)

In the case of the verb ‘understand’, as Table 229 shows, a stem augment *-g(a)* appears, in the negative past, in the SG and in the third person.⁸⁵ Other lexemes show this exact same paradigmatic configuration with stem extensions in /d/ or /ŋ/ instead.

5.2.59 Svan (Tuite 1995)

In the Kartvelian language Svan, the past indicative tenses (aorist and imperfect) of most verbs show an opposition between the forms used in 1SG+2SG and those in 3SG+PL. Consider the following aorist paradigms:

	‘extinguish’		‘cut’		‘wreck’	
	SG	PL	SG	PL	SG	PL
1EXC	o-dəg	o-dig-d	o-č'k'or	o-č'k'wer-d	žoxw-žwem	žoxw-žom-e-d
1INC	-	al-dig-d	-	al-č'k'wer-d	-	žolw-žom-e-d
2	a-dəg	a-dig-d	a-č'k'or	a-č'k'wer-d	žoxw-žwem	žoxw-žom-e-d
3	a-dig	a-dig-x	a-č'k'wer	a-č'k'wer-x	žoxw-žom-e	žoxw-žom-e-x

Table 230: Aorist tense paradigm of three Svan verbs (Tuite 1998:12, 1994:4)

As shown in Table 230, the formal instantiations of this morphological opposition are very diverse. Some verbs (e.g. ‘cut’ above) mark these cells by umlauting⁸⁶ the stem vowel. Some other verbs show umlauting of the 1SG and 2SG instead (see ‘wreck’) as well as suffixation on

⁸⁵ Note that in other lexemes these stem augments occur in the singular forms exclusively. These cases, of course, do not classify as morphomic.

⁸⁶ This started (it is no longer a synchronic phonological rule) as the anticipatory assimilation of /a/, /o/, /u/ (and /ə/?) to a following front high vowel, which yielded /æ/, /œ/, /y/ (and /i/?) respectively. Note that the form /we/ shown in Table is due to a later development in some Svan varieties, which unpacked front rounded vowels into a labial+front vowel sequence (i.e. /œ/ > /we/).

3SG+PL. Yet other verbs (e.g. ‘extinguish’) show more ancient vowel apophonies⁸⁷ which have the same paradigmatic distribution synchronically. In the tenses besides the past indicative, the stem vowel can match the one in 3+PL aorist or the one in 1SG/2SG aorist.

The diachronic origin of this paradigmatic alternation is not entirely understood (see Tuite 1995 for some hypotheses) and is necessarily complex. It may boil down ultimately, however, to a situation where zero-marked 1SG and 2SG were opposed to overt suffixes in the rest of the person-number combinations. Sound changes (e.g. the loss of final vowels) would have caused a (past?) suffix /i/ to be erased from the unsuffixed forms (i.e. *o-č'k'or-i > o-č'k'or) but not from other cells (i.e. *a-č'k'or-i-a > *a-č'k'or-i). Later anticipatory vowel assimilations gave rise to some of the stem alternations we see in synchrony.

Be that as it may, as Tuite (1995:29) explains, this morphological opposition in Svan “is sufficiently implanted in the grammar that all sorts of formal means, varying from region to region, have been recruited to express it”. This might be the case, for example, of some of the aforementioned vowel apophonies (those known as ‘ablaut’), whose reflexes in other Kartvelian languages have a different paradigmatic distribution from the one they show in Svan (namely 1/2 vs 3 in Old Georgian, see Tuite 1995:12 and left of Table 231). It seems, thus, that the paradigmatic distribution of a more ancient vowel alternation (Ablaut) might have been changed to fit that of a more recent and robust one (Umlaut). This may have been facilitated by the formal and distributional similarity of the two patterns:

	Old alternations				Umlaut-derived metaphonies								Re-structuring			
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	ə	ə	e	e	u	y	o	œ	ə	i	a	æ	ə	i	e	æ
2	ə	ə	e	e	u	y	o	œ	ə	i	a	æ	ə	i	e	æ
3	i	i	æ	æ	y	y	œ	œ	i	i	æ	æ	i	i	æ	æ

Table 231: Converging patterns of vowel apophony in Svan

⁸⁷ These are the alternations known as Ablaut in Kartvelian studies. These vowel apophonies (which are reminiscent of the PIE vowel grades) are very ancient and can be traced all the way back to Proto-Kartvelian (see Gamkrelidze 1966). They surface as /ə/-/i/ and /e/-/æ/ in Svan.

The morphological variation encountered between the different Svan dialects also appears to confirm the productivity and diachronic resilience of this 1SG/2SG vs 3SG/PL split. Unrelated affixes appear to have been repeatedly recruited to express the distinction:

	Becho		Laxamul		Lashx	
	SG	PL	SG	PL	SG	PL
1EXC	xwamar-a- sgw	xwamar-a-d	xwamar-∅	xwamar-a-d	xwamār- is	amār-(d)ad
1INC	-	lamar-a-d	-	lamar-a-d	-	amār-(d)ad
2	xamar-a- sgw	xamar-a-d	xamar-∅	xamar-a-d	xamār- is	amār-(d)ad
3	amar-a	amar-a-x	amar-a	amar-a-x	amār-(d)a	amār-(d)ax

Table 232: The verb ‘prepare’ in the imperfect tense in various Svan varieties (Tuite 1995:30)

The morphological means to distinguish 1SG+2SG from 3SG+PL differ from one variety of Svan to another. Looking at 1SG/2SG, we see a suffix *-sgw* in Becho, a suffix *-is* in Lashx, and the absence of a suffix in Laxamul. This indicates unmistakably that at least some of these strategies must be innovations, which suggests that the morphomic opposition that I have described in this section is still productive or has been so in recent times.

5.2.60 Thulung (Lahaussis 2002)

Stem alternations in Thulung (Sino-Tibetan) are numerous and they often involve the addition of segments in particular paradigm cells. Consider the following verb’s inflection:

	Non-past			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	ge-ŋu	ge-tsuku	ge-ku	ge-ŋroro	get-tsoko ⁸⁸	get-toko
1INCL		ge-tsi	geɟ-i		get-tsi	geɟ-dʒi
2	ge-na	ge-tsi	ge-ni	geɟ-na	get-tsi	geɟ-ni
3	ge	ge-tsi	ge-mi	geɟ-ɟa	get-tsi	ge-miri

Table 233: Paradigm of Thulung ‘come (up)’, intransitive (from Allen 1975:204)

⁸⁸ The alternation between /d/ and /t/ is automatic (i.e. phonological).

The pattern displayed in Table 233 can also be instantiated with the forms -k (vs -∅), -p (vs -m) and -ɖ (vs -n). This stronger/longer stem appears, thus, in the non-past, in the 1PL.INCL and in the past everywhere except in the 1SG and 3PL. In the case of transitive verbs, the distribution of these formal alternations is slightly different:

	Non-past			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	rep-u	rem-tsuku	rem-ku	rep-to	rep-tsoko	rep-toko
1INCL		rem-tsi	rep-i		rep-tsi	rep-ɖi
2	rem-na	rem-tsi	rem-ni	rep-na	rep-tsi	rep-ni
3	rep-y	rem-tsi	rem-mi	rep-ɖy	rep-tsi	rep-miri ⁸⁹

Table 234: Paradigm of Thulung 'look', transitive, 3SG patient (Lahaussois 2002:158).

As shown in Table 234, the long stem appears in a superset of the contexts where it showed up in transitive verbs, extending to the 1SG and 3PL present and to the whole of the past.

The stem alternations in all of these paradigms in Thulung seem to originate from the deletion/lenition of stem-final consonants in concrete phonological environments. Although the correlation is no longer perfect, the consonants tend to surface in the present before vowel-initial suffixes. In the past, the "survival" of the (stronger) stem-final consonant appears to be due to it having been protected (or reinforced) by a former past tense suffix (-ɖ-) which subsequently disappeared in many contexts. Traces of this suffix can still be found. Compare, e.g. the 3SG and the 1PL.INCL past to their present tense counterparts.

Thulung1: 2SG.PAST/3SG.PAST/DU.PAST/1PL.PAST/2PL.PAST

Thulung2: 1SG.PRES/3SG.PRES/1SG.PAST/2SG.PAST/3SG.PAST/DU.PAST/PL.PAST

⁸⁹ Lahaussois mentions the existence of variation in the 3PL, in both past and present as for the particular stem used in those two cells. This, however, does not come in the way of this pattern being unavoidably morphomic.

5.2.61 Tol (Dennis 1992, Holt 1999)

Person-number agreement inflection in Tol (Jicaquean, Honduras) is characterized by complex segmental alternations in stems.⁹⁰ Consider the following paradigms:

	<i>sipi</i> 'hit'		<i>nuku</i> 'see'		<i>pake</i> 'write'	
	SG	PL	SG	PL	SG	PL
1	syip ^h	syipik ^h	t ^h inyuk ^h	t ^h inyukuk ^h	tepyak ^h	tepyakek ^h
2	syip ^h	sipi	t ^h inyuk ^h	t ^h unuku	tepyak ^h	tapake
3	syipa	sip ^h	t ^h inyuka	t ^h unuk ^h	tepyaka	tapak ^h

Table 235: Past tense inflection of some class 1 Tol verbs (Holt 1999:23)

As Table 235 above shows, Class 1 verbs in Tol (which are mostly transitive) are characterized by a morphological affinity of SG and 1PL. In these forms (also in the present, but not in the future), a glide occurs before the stem vowel. In those verbs (e.g. 'see' and 'write') where a past tense prefix is present, the vowel in the prefix may also vary from SG+1PL to 2PL/3PL. In addition to the above, stem consonant allomorphy can also be found in some verbs of Class 1. These formal alternations, however, follow a different pattern:

	Present ⁹¹		Future	
	SG	PL	SG	PL
1	hyok ^h	hyoʔ-o-k ^h	mo-hok ^h	mo-hoʔ-o-k ^h
2	hyok ^h	hoʔ-o	mo-hoʔ-o-n	mo-hoʔ-o
3	hyoʔ-o	ha-hok ^h	mo-hoʔ-o-s	mo-hoʔ-o-k ^h

Table 236: Inflectional paradigm of verb 'cut', class 1 (Dennis 1992:21,33)

⁹⁰ Holt (1999:16) derives many of these surface forms from more concatenative underlying forms by means of highly complex morphophonological rules (e.g. *myaʔna* 'gives birth' is allegedly derived from an underlying *himanunua*). Holt (1999:18) mentions that this system of underlying forms and morphophonological rules "may also bear some relation to a supposed underlying competence on the part of present-day speakers of Tol". Some of the transformations he posits are likely to recapitulate former sound changes in the language, however, I am skeptical of the validity of this analysis in synchrony and I will only deal with surface forms here.

⁹¹ The past tense behaves as the present for the purposes of this morphological alternation except in a few verbs that show no stem alternation in the present (in which case they have the pre-zero stem alternant in all of the present cells).

Some verbs in Tol show a morphological alternation in the right periphery of the stem, with one stem (e.g. *hok^h*) appearing in some (unsuffixed) paradigm cells and another one (*hoʔ[o]*) elsewhere. The alternations are very diverse formally (parallely to *hok^h/hoʔ* [see Table 236] we have *tat^h/taʔ* ‘have’, *k^hol/k^hok^h* ‘grind’, *sok^h/sok^{ht}* ‘untie’, *la/lah* ‘eat’, *ʔinan/ʔiʔn* ‘kill’ etc., see Dennis 1992:54-55). Although the differential phonological context suffixed vs unsuffixed was probably responsible for their diachronic emergence, the formal alternations involved suggest that there is little hope for a phonological explanation synchronically.

As the present tense paradigm of ‘cut’ suggests, the two morphomic patterns discussed here so far (the first one chiefly vocalic and with a locus on the left of the word form, the second one involving the presence of a vowel and/or consonant alternations at the right edge of the stem) are fully compatible and participate actively in the system of formal distinctions in the language. Some roots can only be inflected with an accompanying particle. In this case, the form of the particle will abide by the first morpheme and the stem by the second:

	Form of particle		Form of stem		Whole paradigm	
	SG	PL	SG	PL	SG	PL
1	le	le	koy	keye	le koy	le keye-k ^h
2	le	lo	koy	keye	le koy	lo keye
3	le	la	keye	koy	le keye	la koy

Table 237: Present tense inflection of *le koy* ‘throw away’ (Dennis 1992:42)

The other big morphological class of verbs in Tol (which is mostly intransitive) shows a different system of morphological allegiances. In this class, for the purposes of the vocalic alternation at the left periphery of the stem, the singular forms pattern with 2PL instead:

	Present		Past	
	SG	PL	SG	PL
1	ʔosʔis	ʔyasʔikek ^h	t ^h aʔasʔis	t ^h eʔyasʔikek ^h
2	ʔosʔim	ʔosʔike	t ^h aʔasʔim	t ^h aʔasʔike
3	ʔosʔi	ʔyasʔiŋ	t ^h aʔasʔi	t ^h eʔyasʔiŋ

Table 238: Inflectional paradigm of *ʔasʔi* ‘bathe’, class 2 (Holt 1999:29)

As Table 238 above shows, in contradistinction to class 1, these verbs show the infix -y- and its associated vowel frontings in 1PL and 3PL, thus leaving SG+2PL as an unnatural class with shared forms. The same as these vowel apophonies, stem-right-edge alternations also show a very different pattern in class 2. Consider the following paradigm:

	Present ⁹²		Future	
	SG	PL	SG	PL
1	miʔ-i-s	myis-ikek ^h	ka miʔ-i-s	ka mis-ikek ^h
2	mis	mis-ike	ka miʔ-i-m	ka mis-ike
3	miʔ-i	myiʔ-i-n	ka miʔ-i-m	ka miʔ-i-n

Table 239: Partial paradigm of verb 'drink, class 2 (Dennis 1992:65,74)

The stem alternations illustrated in Table 239 are also formally diverse, even if less so than those of class 1. Near *miʔ/mis* we find *p^hak/p^haʔ* 'hear' and *peʔ/pec* 'defecate'. Last but not least, the irregular verb 'go' shows still another morphological pattern whereby it is the 3PL that patterns with the SG forms across all tenses:

	Present		Past		Future	
	SG	PL	SG	PL	SG	PL
1	hum	leke	t ^h um	tleke	mis	nlaka
2	hay	lowa	t ^h ay	tlowa	mim	nlawu
3	hama	hil	themey	t ^h il	mim	mal

Table 240: Paradigm of Tol 'go' (Holt 1999:30)

In a way which is reminiscent of the Romance L-morpheme (e.g. in French *aller*), the Tol verb for 'go' also shows a different stem in SG+3PL and in 1PL/2PL.

Tol1: SG/1PL

Tol2: 3SG.P/1PL.P/2PL.P/2SG.F/3SG.F/PL.F

Tol3: 1SG.P/2SG.P/3PL.P/1SG.F

⁹² The past tense behaves again as the present for the purposes of this morphological alternation.

ToI4: SG/2PL

ToI5: 1SG.P/3.P/SG.F/3PL.F

ToI6: 2SG.P/1PL.P/2PL.P/1PL.F/2PL.F

ToI7: SG/3PL

5.2.62 Turkana (Dimmendaal 1991)

In Turkana (Nilotic) inflection, partial and whole-word syncretisms are widespread. There are two inflectional classes in the language. Consider the following paradigm of a class 1 verb:

	Perfective present		Perfective past	
	SG	PL	SG	PL
1	a-los-it	ki-los-it	a-los-o	ki-los-o
2	i-los-it	i-los-it-o	i-los-o	i-los-os(i)
3	e-los-it	e-los-it-o	a-los-o	a-los-os(i)

Table 241: Partial paradigm of 'go' in Turkana (Dimmendaal 1991:283-284)

The prefixal syncretism of 1SG.PRES+1SG.PAST+3.PAST seen in class 1 is also repeated in class 2 with the prefix *e-* instead, which makes this morphological affinity systematic as defined here. As explained by Dimmendaal, these two inflectional classes in Turkana emerged due to the presence of an early causative prefix *i-* in class 2 verbs. This suffix became at some point unproductive and increasingly lexicalized. The vowels of the person-number agreement prefixes in class 2 merged with this former prefix to yield a new set of markers where the vowels are raised one degree from their height in class 1 (i.e. *a-i-STEM > e-STEM, e-i-STEM > i-STEM).

This system is widely shared across most of the languages closely related to Turkana (see Dimmendaal 1991:290) and must thus be inherited from the Proto-language. One variety,

however, shows an interesting deviation from this family pattern in that the 1PL form does not have the expected *ki-* but takes a form that patterns as 3. Observe the forms involved:

	Turkana								Toposa							
	Class 1				Class 2				Class 1				Class 2			
	PRES		PAST		PRES		PAST		PRES		PAST		PRES		PAST	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	a-	ki-	a-	ki-	e-	ki-	e-	ki-	a-	e-	a-	a-	e-	i-	e-	e-
2	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-	i-
3	e-	e-	a-	a-	i-	i-	e-	e-	e-	e-	a-	a-	i-	i-	e-	e-

Table 242: Person-number prefixes in some Turkana varieties (Dimmendaal 1991:290)

What happened in this variety is that a formerly impersonal construction (based on the third person morphologically) came to replace the inherited 1PL.⁹³ Paradoxically, this does not give rise to a systematic (as defined here) morphological identity of 3 and 1PL, but merely results in the extension of the morphomic pattern described above to the 1PL.PAST. This extended pattern in Toposa is also morphomic and has been included in the database.

Turkana1: 1SG/3.PAST

Turkana2 (Toposa): 1SG.PRES/1.PAST/3.PAST

5.2.63 Twi (Stump 2015)

In the Niger-Congo language Twi, there is a morphological polarity configuration in the expression of past vs perfect tense and positive vs negative polarity.⁹⁴ Observe the following partial paradigm of the verb *tó* ‘buy’:

⁹³ This constitutes a cross-linguistically recurrent development. Consider, for example, my earlier comments (Section 5.1.4) on Kariña and more well-known cases like the contemporary uses of the impersonal in colloquial French, where the etymological 1PL (e.g. “nous allons”) is being replaced by the impersonal (i.e. “on va”) which is morphologically a third person.

⁹⁴ A similar configuration can be found in Texmelucan Zapotec (Speck 1984), where the morphology that marks the positive potential appears in the negative of all the tenses except the potential.

	Before complement		Elsewhere	
	Affirmative	Negative	Affirmative	Negative
Past	tò-ò	à-ń-tó	tò-ò-yè	à-ń-tó
Perfect	à-tó	ń-tó-ò	à-tó	ń-tó-ò-yè

Table 242: Past and perfect forms of ‘buy’ in Twi (Stump 2015:136)

As Table 242 illustrates, leaving aside the nasal prefix which consistently occurs in the negative, the rest of the morphology is distributed in an unexpected way. The prefix à- occurs in the perfect affirmative and in the past negative. Conversely, stem-vowel lengthening (tò > tòò) and the suffix -yè both occur in the past affirmative and in the perfect negative. The latter morphological affinity qualifies for morphomehood here.

The diachronic origin of this system is uncertain, however, some observations may help to shed some light. The first is that the TAM system of Twi is characterized by less distinctions in the negative than in the positive (4 vs 9 respectively according to Osam 1994:103). The second one is the incompatibility of the past (sometimes labelled ‘completive’) and negation in related languages (e.g. in Anufo, see Smye 2004:88).

The explanation I would like to propose here is, thus, that the tense nowadays labelled ‘past’ (also sometimes ‘remote past’) in Twi must have formerly expressed completive aspect and must have been semantically incompatible with negation at this stage. One can understand the logic of this, since what has been completed cannot be expected to not have happened at all. At a later stage, the semantics of the tense must have drifted to include past tense uses which were no longer logically incompatible with negation. Because of the absence of a negative form for the tense, however, the semantically closest thing (i.e. the negative perfect) would have been used instead:

	Before complement		Elsewhere	
	Affirmative	Negative	Affirmative	Negative
Past	*tò-ò	*à-ń-tó	*tò-ò-yè	*à-ń-tó
Perfect	*à-tó	*à-ń-tó	*à-tó	*à-ń-tó

Table 243: Proposed system of morphological oppositions in Pre-Twi ‘buy’

The developments up to this point are not surprising and the system at this stage would have been similar to the one found in related Anufo (Smye 2004:88) and in comparable TAM/negation morphology in Twi itself in other tenses.⁹⁵

The later (quite striking) development that sets this pattern apart would be the innovation of a negative form for the perfect in Twi on the basis of the past.⁹⁶ It might make functional sense to try and (re)introduce in the negative some of the TAM distinctions that hold in the positive. Thus, the impulse to de-syncretize negative past and negative perfect seems understandable. The morphological form used to mark the past was available as a potential source for innovating this distinction. However, its use to mark the negative perfect, rather than the negative past, seems surprising and may demand additional explanations to the ones offered here. The development would appear to make sense, for example, only if there was some sort of pressure (e.g. a lower frequency of use) that made changing the perfect negative preferable to or more likely than changing the past negative.

5.2.64 Udmurt (Winkler 2001, Csúcs 1988)

In the Uralic language Udmurt, verbs are conjugated for past, present, future and pluperfect. The future tense and the 3PL present appear to have an unnatural morphological affinity. Consider the following paradigm:

	First conjugation, <i>m̄ñĩñĩ</i> 'go'				Second conjugation, <i>daśanjĩ</i> 'prepare'			
	Present		Future		Present		Future	
	SG	PL	SG	PL	SG	PL	SG	PL
1	m̄ñĩ-śko	m̄ñĩ-śko-m	m̄ñĩ-o	m̄ñĩ-o-m	daśa-śko	daśa-śko-m	daśa-lo	daśa-lo-m
2	m̄ñĩ-śko-d	m̄ñĩ-śko-dĩ	m̄ñĩ-o-d	m̄ñĩ-o-dĩ	daśa-śko-d	daśa-śko-dĩ	daśa-lo-d	daśa-lo-dĩ
3	m̄ñĩ-e	m̄ñĩ-o	m̄ñĩ-o-z	m̄ñĩ-o-zĩ	daśa	daśa-lo	daśa-lo-z	daśa-lo-zĩ

Table 244: Verb agreement in Udmurt (Csúcs 1988:142)

⁹⁵ For example, as explained by Osam (1994), the mark of the progressive tense in Twi is a prefix *re-* and the mark of the future is a prefix *bɛ-*. However, the negative form of the two tenses has *re-*.

⁹⁶ See the language Triqui (Otomanguean, discussed in Baerman 2007b) for a very similar reversal involving aspect and negation and for a diachronic scenario similar to the one proposed here.

The shaded cells in Table 244 share a suffix (or a stem extension) to the exclusion of the rest of the paradigm. This element takes slightly different forms in the two conjugations of the language and therefore classifies as a morpheme here.

Our knowledge of the Udmurt verb's history is incomplete but we can make educated guesses as for how the present system may have emerged. I propose here two steps. The first one involves the intrusion of the formative (-sk-) in the first and second person forms of the present tense. These forms are absent from Udmurt's closest relative Komi (Avril 2006), where present and future are only distinguished in the third person. The incorporation of this suffix into the person-number agreement complex, thus, constitutes unmistakably an innovation of Udmurt motivated, maybe, by the will to distinguish present and future consistently. It has been proposed that the suffix denoted a frequentative meaning originally (see Winkler 2001:50).⁹⁷

A second step would have involved the emergence of the second conjugation from the first. It is usually assumed (see e.g. Frodl 2013:21-22) that the /l/ that characterizes this verb class was originally part of the stem and appeared, therefore, throughout the whole paradigm. Sound change would have then deleted the final consonant in coda positions (e.g. 1SG.PRES **daśal-śko* > *daśa-śko*, 3SG.PRES *daśal* > *daśa*) while leaving intervocalic /l/ in place (e.g. in 3PL.PRES *daśal-o*).

5.2.65 Vitu (van den Berg & Bachet 2006)

TAM particles in Vitu (Oceanic) change form according to the person-number of the subject.

Consider the following particles in Table :

⁹⁷ Note the similarity of this development with the evolution of the inchoative suffix (-sk-) from Latin to some modern-day Romance languages (Meul 2010). There is, in a different part of the Udmurt verbal paradigm, another parallel to this borrowing of a derivational formative for the expression of inflectional values. The 2PL and 3PL of the second past show an infix -l'a- that is also a frequentative marker in the language (Winkler 2001:50).

	Realis		Irrealis		Perfect		Continuity	
	SG	NSG	SG	NSG	SG	NSG	SG	NSG
1	ta	ta	na	na	te	te	ka	ka
2	tu	ta	nu	na	tu	te	ku	ka
3	e	ta	ni	na	ti	te	ki	ka

Table 245: Forms of some TAM particles of Vitu (van den Berg & Bachet 2006:97)

As van den Berg & Bachet (2006:97) mention, the inflection of these particles is “somewhat unusual in that, with a few exceptions, the first person singular and all duals and plurals are grouped together, while the second and third person singular have separate forms”.

5.2.66 Vurës (Malau 2016)

In some other Oceanic languages (see also Kele in Section 5.2.31 before), nouns have stem alternations in their possessive paradigms. Consider the paradigm below:

	SG	DU	PL
1EXCL	vulu-k	vulu-mōrōk	vulu-mem
1INCL	-	vulu-dōrōk	vulu-nēn
2	vōlō-ñ	vulu-mōrōn	vulu-mi
3	vōlō-n	vulu-r	vulu-r

Table 246: Possessive paradigm for ‘hair’ (Malau 2016:275)

The same pattern is shown by various other vowel pairs, more specifically i vs ē, iē vs ia, ö vs o and ë vs a. As explained for Kele previously, these vowel apophonies originated from the anticipatory vowel assimilation of the stem vowel(s) to the vowel in the following suffix. In the contemporary languages, however, the patterns do not always agree and analogical changes have undoubtedly played a big role. This is seen clearly, for example, if we compare Vurës with its close relative Mwotlap (François 2001). In this language, 1EXCL and 2 (all numbers) share a stem different from the one found in 3 and 1INCL.

5.2.67 Wambisa (Peña 2016)

In the possessive inflection of Wambisa (Chicham) nouns (also in related Shuar, see Saad 2014:49), the third person singular and the plural cells behave as a single morphological object in that they are always syncretic. Observe the following paradigm:

	<i>muuka</i> 'head'		<i>nauantu</i> 'daughter'	
	SG	PL	SG	PL
1	muuka-ru	muukĩ	nauantu-ru	nauantu-rĩ
2	muuki-mi	muukĩ	nauantu-rumi	nauantu-rĩ
3	muukĩ	muukĩ	nauantu-rĩ	nauantu-rĩ

Table 247: Possessive inflection of two Wambisa nouns (Peña 2016:467)

This falling-together of an unnatural class of cells with different formatives constitutes a morpheme according to our definition (see the Section 5.2.2 on Aguaruna for diachronic insights on this pattern). Another area of the Wambisa grammar where a morphological affinity is observed within an unnatural class of cells is the different-subject morphology of the verb. Here, the 1SG and the third person that are characterized by shared morphology:

	Simultaneous DS		Sequential DS	
	SG	PL	SG	PL
1	puha-ku-ĩ	puha-ku-ri-nĩ	puhu-sa-mataĩ	puhu-sa-ri-nĩ
2	puha-ku-mi-nĩ	puha-ku-rumi-nĩ	puhu-sa-mi-nĩ	puhu-sa-rumi-nĩ
3	puha-ku-ĩ	puhu-ina-ku-ĩ	puhu-sa-mataĩ	puhu-sa-ara-mataĩ

Table 248: Different subject inflection in the Wambisa verb *puhu* 'live' (Peña 2016:808)

As shown above, these cells are characterized by shared exponence, which changes from simultaneous to sequential DS.⁹⁸ The same morphomic affinity holds in the related Chicham

⁹⁸ The alternation -nĩ vs -ĩ is presented by Peña as a morphophonological process in Wambisa. According to him, there is just one suffix -(n)ĩ which is realized as -nĩ after /i/ and as -ĩ elsewhere. This is, as one can probably guess from the forms involved, not a phonologically regular process. Saad (2014:127) does not favour the same analysis in closely-related Shuar and for him the two forms (-n and -ĩ in Shuar) are different in a deeper sense.

languages Achuar (Fast & Fast 1981:107) and Shuar (Saad 2014:115). Aguaruna, by contrast, shows a slightly different picture whereby that affinity extends to the 1PL as well:

	Simultaneous DS		Sequential DS	
	SG	PL	SG	PL
1	anta-ku-ĩ	antu-ina-ku-ĩ	antu-ka-mataĩ	antu-ka-aha-mataĩ
2	anta-ku-mi-nĩ	anta-ku-humi-nĩ	antu-ka-mi-nĩ	antu-ka-humi-nĩ
3	anta-ku-ĩ	antu-ina-ku-ĩ	antu-ka-mataĩ	antu-ka-aha-mataĩ

Table 249: Different subject inflection in Aguaruna *antu* ‘hear’ (Overall 2007:398-399)

There is reason to believe that Wambisa, Achuar and Shuar continue the original system and that it is Aguaruna that has innovated. This is suggested by two different facts. The first is that the appearance of the pluralizers *-ina* and *-aha* in the 1PL is not common at all in Aguaruna. Other, closely-related paradigms, like the imperfective DS one (Overall 2007:400), show *-ina* only in the 3PL. A second factor that suggests the chronological precedence of the 1SG+3 system is that there is a formative *-taĩ* which appears in Aguaruna (Overall 2007: 397-398) but also, crucially, in Wambisa (Peña 2016:812) in the first person (both SG and PL) and in the third. This formative could have, thus, provided the analogical model in earlier Aguaruna to extend the suffix *-mataĩ* to the 1PL. In addition, the absence of 1SG marking (*-ha* elsewhere) in the DS verbal inflection makes the 1SG form look like the (unaffixed) 3SG. That (directional) syncretism (1SG<3SG) could have been simply extended to the plural in Aguaruna, which would be the reason why today we find a form like *antu-ina-ku-ĩ* in the 1PL instead of the expected *antu-ku-hi-nĩ*.

Wambisa1: 3SG/PL

Wambisa2: 1SG/3

5.2.68 Wutung (Marmion 2010)

The language Wutung (Sko, New Guinea) is characterized by considerable morphological complexity in the domain of verbal person-number inflection. The language is plagued by syncretisms and exponence patterns that appear to be completely oblivious to natural morphosyntactic classes. The morphological identities often contradict one another and the initial impression is of total chaos. On closer scrutiny, however, several patterns recur in the language. Most notable among these is the formal identity of 1SG and 2PL, which in the vast majority of verbs are whole-word syncretic. Consider the following paradigms:

	'be here'		'be under'		'follow'	
	SG	PL	SG	PL	SG	PL
1	punga	nua	qang	ne	ha-qe	hna-ne
2	mua	punga	me	qang	hma-me	ha-qe
3M	mua	mua	nyi	qing	qa-nyie	hnya-eng
3F	ma	mua	ing	qing	hwa-eng	hnya-eng

Table 250: Three Wutung verbs (Marmion 2010:305-306)

As Table 250 above shows, 1SG and 2PL often share form to the exclusion of the remaining paradigm cells. The forms shared can be varied (e.g. /pũ/, /ã/, /ʔ/ above)⁹⁹ although segmentation into exponents is exceedingly complicated. Lexical verbs may consist of a single inflecting root (e.g. 'be here' and 'be under' in Table 250) but they are often also compounds of either a) two inflecting roots (e.g. 'follow' in Table 250) or b) an inflecting root and an invariable root (e.g. *qang-qwur*, *me-qwur*, *nyi-qwur*... 'lie down').

Despite the synchronic complexity of the Wutung verbal agreement system, its diachronic emergence is quite straightforward. Comparative evidence from other Skou languages (e.g. Skou [Donohue 2004] and Vanimo [Ross 1980]) as well as a look at the regularities within Wutung make it clear that the system emerged from the prefixation of relatively unremarkable person-number markers. Later sound changes would have often fused those prefixes and the initial consonants of the stems in an unsegmentable form:

⁹⁹ The digraph 'ng' indicates nasalization of a previous vowel and 'q' represents /ʔ/.

	Pronouns		Proto-prefixes		+pṼ		+lV		+qṼ	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	nie	ne(tu)	*∅-	*n-	p	n	l	d	q	n/hn
2	me	e(tu)	*m-	*∅-	m	p	b	l	qm	q
3M	qey	te(tu)	*q-	*t-	m	m	ql	t/s	q	q
3F	cey	te(tu)	*c-/w-	*t-	m	m	c/hl	t/s	qw	q

Table 251: Wutung free pronouns, proto-prefixes and their phonological outcomes with different stem initials (Marmion 2010)

The reason why 1SG and 2PL are almost always syncretic, as Table 251 suggests, is simply derived from the fact that those two forms had a zero prefix and would have left the original stem-initial consonant always unchanged. An original stem-initial /p/ would, thus, only be regularly continued as /p/ in 1SG+2PL. Other stem-initial consonants would have been preserved in other phonological contexts as well. Initial /l/, for example, is not regularly altered by the 3M.SG prefix /ʔ/ either (i.e. ʔ + l = ʔl). Stem-initial /ʔ/ would “survive” in addition, in combination with the 3PL prefix /t/ as well (i.e. t + ʔ = ʔ).

It must be stressed, however, that there is no phonological rule that would account for the forms we find synchronically. There is also evidence of widespread analogical changes that have reorganized the regularly-inherited forms. Thus, both sound change and analogy have conspired to create the 1SG/2PL morphomic syncretism displayed above, as well as the rest of the (less robust) morphological alliances. Because of the reasons explained, these constitute often supersets of the 1SG+2PL morpheme. Consider the following paradigms:

	‘do’		‘rub’		‘take’ 1/2/3M.SG OBJ	
	SG	PL	SG	PL	SG	PL
1	ley	dey	qo	do	qai	qdi
2	bey	ley	bo	qo	qbi	qai
3M	q-ley	tey	qo	to	qai	si
3F	cey	tey	co	to	qwi	si

Table 252: Three Wutung verbs (Marmion 2010:303,305,311)

Some other times, as Table 252 illustrates, it is 1SG, 2PL and 3M.SG which share segments to the exclusion of the remaining paradigm cells, sometimes (e.g. ‘rub’ and ‘take’) resulting in whole-word syncretism. The shared forms can also be diverse (i.e. /l/, /ʔ/, /ai/ above). Therefore, this morphological pattern also qualifies as morphomic here. Other patterns constitute still larger supersets:

	‘wait’		‘hide’		‘be with’	
	SG	PL	SG	PL	SG	PL
1	qangqie	qmie	qaing	qni	la	da
2	qmie	qangqie	qmi	qaing	ba	q-la
3M	qangqie	qangqie	qaing	qaing	q-la	q-la
3F	qwie	qangqie	qwing	qaing	wa	q-la

Table 253: Three Wutung verbs (Marmion 2010:303,305,311)

In the paradigms in Table 253, 3PL is added to the previous morpheme as the set of cells which displays shared formatives. It must be stressed again that many of these patterns have come about by analogy. As Marmion (2010:303,305) mentions, the forms of the 1PL, 2SG and 3F.SG are all unexpected in ‘wait’, the same as the 2PL and 3PL in ‘be with’, which would have been expected to be *la* and *sa* respectively by regular sound change. One last pattern that is relatively recurrent¹⁰⁰ in Wutung involves the same morphology in 3SG.F and 3PL:

	‘cut’		‘be on top’		‘lie down’	
	SG	PL	SG	PL	SG	PL
1	hur-lang	hur-na	qa-le	da-ne	qang-qwur	ne-qwur
2	hur-ma	hur-lang	ba-me	qa-si	me-qwur	qang-qwur
3M	hur-qlang	hur-nya	jie-lie	qi-li	nyi-qwur	ing-qwur
3F	hur-nya	hur-nya	qi-li	qi-li	ing-qwur	ing-qwur

Table 254: Three Wutung verbs (Marmion 2010:321,326)

¹⁰⁰ Many patterns exist in Wutung that are completely exceptional. Many (maybe most) one-root lexemes would classify as singleton inflection classes. This is probably possible because of the relatively small number of inflecting roots in the language (around 200), which are recycled into compounds to form more lexemes.

Observe in Table 254 how these cells can be whole-word syncretic and share various forms (i.e. /ɲ/, /i/ and /i/ above) not present elsewhere in the paradigm.

Wutung1: 1SG/2PL

Wutung2: 1SG/2PL/3SG.M

Wutung3: 1SG/2PL/3SG.M/3PL

Wutung4: 3SG.F/3PL

5.2.69 Yagaria (Stump 2015, Haiman 1980)

In Yagaria (also called Hua) and other Gorokan languages (also in the related Kainantu family of Trans-New-Guinea, e.g. in Awa, see Loving 1973) there is a morphological affinity, in mood suffixes, between the 2SG and the 1PL, which share their exponence to the exclusion of the rest of the person-number values. Consider the following paradigm:

	SG	DU	PL
1	hu-ve	hu-'-ve	hu-pe
2	ha-pe	ha-'-ve	ha-ve
3	hi-ve	ha-'-ve	ha-ve

Table 255: *hu* 'do', interrogative mood (Stump 2015:128)

As presented before in Section 3.4, the morphological contrast between a -p in the 2SG/1PL and a -v in the rest of the paradigm is repeated in other moods with different exponents, for example in the indicative (-n vs -∅), in the relative (-p vs -m), in the medial coordinate (-n vs -g) or in the counterfactual (-s vs -h). A total of 12 mood suffixes show this morphomic pattern of exponence, although the actual alternating forms are always these.

The explanation for these alternations was advanced by Foley (1986:251). The subject suffixes he reconstructed for Proto-Gorokan are the following:

	SG	DU	PL
1	-u	-us	-uN
2	-a:N	-a:s	-a:
3	-i	-a:s	-a:

Table 256: Proto-Gorokan subject suffixes (Foley 1986:74)

These subject suffixes would have been followed by invariable particles marking illocutionary force (e.g. interrogative *pe*: *-u pe* for 1SG vs *-uN pe* for 1PL etc.). Later sound changes would have treated (some of) those particles differently depending on whether they followed the nasal or not. In this case, for example, the intervocalic /p/ in the sequence *-upe* would have been lenited (to *-uve* in Yagaria and to *-ufi* in Benabena) whereas the non-intervocalic /p/ in the sequence *-uNpe* would have been preserved as /p/ because it was “protected” from lenition by the previous nasal. Similar sound changes would have given rise to the rest of the synchronically attested formal alternations (except for *-n* vs \emptyset , which would just continue the initial situation, albeit with a resegmentation of the final nasal, which would have been simply “chopped away” from the subject agreement morphology and grouped together with the mood suffixes instead):

	SG	DU	PL
1	ormu-e	ormu-'-e	ormu-ne
2	ormi-ne	ermi-'-e	ermi-e
3	ormi-e	ermi-'-e	ermi-e

Table 257: *ormi* ‘come down’ Indicative mood (Haiman 1980:121)

5.2.70 Yorno-So (Heath 2014)

The verbal agreement inflection of Yorno-So (Dogon, Mali) is characterized by a morphological affinity of 1PL and 3PL, which are opposed to the rest of the paradigm, i.e. SG+2PL. Consider the following paradigms:

	‘fall’, imperfective		‘hit’, imperfect negative		‘see’, exper. perf. neg.	
	SG	PL	SG	PL	SG	PL
1	númò-jè-m	númòy	lágà-lè-m	lágàynè	yé:té-r-úm	yé:téné
2	númò-jè-w	númò-jè-y	lágà-lè-w	lágà-lè-y	yé:té-r-úw	yé:té-r-íy
3	númò-jè	númòy	lágà-lè	lágàynè	yé:té-r	yé:téné

Table 258: partial paradigms of three Yorno-So Dogon verbs (Heath 2014:186,190,191)

In the inflection of many tenses there is a morphological opposition of SG+2PL and 1PL+3PL. Both sets of cells, as the table above illustrates, are able to take exponents of their own. For the purposes of the present dissertation, SG+2PL qualifies as a morpheme.

The story of this morphological opposition is an interesting one. Person-number agreement seems to be a relatively recent innovation in Dogon because some languages (e.g. Togo Kan, see Heath 2011) do not have it. What all Dogon languages do have is some sort of number agreement in the verb. This morphological contrast applies, most frequently, only to third person arguments, particularly to animates, thus creating an opposition between a plural-marked 3PL and the rest of the paradigm (unmarked).

As its presence across the whole family suggests, this morphological contrast must be older than the person-number suffixes and is thus more robustly hardwired into the inflectional system, which means that cumulative forms and allomorphy have had time to develop. The main innovation that separates Yorno-So from its sister languages (e.g. from closely-related Tommo-So, see McPherson 2013) is that the earlier 3PL forms have spread to the 1PL.

5.2.71 Zapotec, Yatzachi (Butler 1980)

In some varieties of Zapotec (Otomaguean), the 3PL agreement morphology stands out as dramatically different from the rest of the person-number agreement forms. In the variety spoken in Yatzachi el Bajo, this cell is characterized by (plural) morphology (in bold in Table 259) that is absent from the rest of the plural cells:

	SG	PL
1EXCL	ch-sed-a' ¹⁰¹	ch-sed-to'
1INCL	-	ch-sed-cho
2	ch-sed-o'	ch-sed-le
3	ch-sed-bo'	ch-əsə'ə-sed-bo'

Table 259: Partial paradigm of 'study', progressive (Butler 1980:147-148)

In some TAMs, this has led to stem alternants being confined to the 1+2+3SG of one aspect and opposed to the majority stem in the 3PL and other aspects:

	Progressive		Stative		Completive		Potential	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	da	da	da	da	da	da	ta	ta
1INCL	-	da	-	da	-	da	-	ta
2	da	da	da	da	da	da	ta	ta
3	da	da	da	da	da	da	ta	da

Table 260: Stem of 'walk' (Butler 1980:82)

In the potential mood, a fortition of the first consonant of the stem happens when the stem is word-initial (e.g. compare 3SG.POT *ta-bo'* to 3PL.POT *yesə'ə-da-bo'*). This process is more or less predictable and may be argued to have a partial phonological motivation. This is not the case, however, of stem alternations in the completive aspect:

	Progress.		Stative		Complet.		Potential	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	bez	bez	bez	bez	lez	lez	cuez	cuez
1INCL	-	bez	-	bez	-	lez	-	cuez
2	bez	bez	bez	bez	lez	lez	cuez	cuez
3	bez	bez	bez	bez	lez	bez	cuez	bez

Table 261: Stem of 'wait' (Butler 1980:86)

¹⁰¹ The progressive is marked with *ch-* and *-sed-* is the stem.

Here, 3SG.COMP *gw-lez-bo'* is opposed to 3PL.COMP *gosa'a-bez-bo'*. This pattern must also have started as a result of sound changes operating in different phonological environments (observe how the plural prefix comes between aspect prefix and stem), however, these stem alternations are phonologically unmotivated synchronically and unpredictable (compare *bez/lez/cuez* in Table 261 to an alternation *bež/chež/cuež* in the verb 'cry', or to *yis/dis/chis* in 'distribute'). The same phonological context that was responsible for the emergence of these consonantal stem alternations also created vocalic alternations in vowel-initial stems:

	Progress.		Stative		Comple.		Potential	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	ene'e	ene'e	ene'e	ene'e	one'e	one'e	ene'e	ene'e
1INCL	-	ene'e	-	ene'e	-	one'e	-	ene'e
2	ene'e	ene'e	ene'e	ene'e	one'e	one'e	ene'e	ene'e
3	ene'e	ene'e	ene'e	ene'e	one'e	ene'e	ene'e	ene'e

Table 262: Stem of 'want' (Butler 1980:97)

These alternations are also formally diverse. Compare *ene'e/one'e* above to *on/en* 'do' (note the reversed vowel qualities) or to *ol/il* 'sing'. Some vowel-initial verbs also add a consonant to the stem in these paradigm cells (e.g. *ao/dao* 'come').

In other Zapotec varieties (e.g. Zaniza and Texmelucan), rather than being "missing" from some cell where they might have been expected, these roots have instead spread in the paradigm beyond the completive and into the 1st person forms of all other TAMs (see Operstein 2002). Consider the following stem alternation:

	Habitual		Unreal		Completive		Potential	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	lez	lez	lez	lez	lez	lez	lez	lez
1INCL	-	lez	-	lez	-	lez	-	lez
2	yez	yez	yez	yez	lez	lez	yez	yez
3	yez	yez	yez	yez	lez	lez	yez	yez

Table 263: Stem of 'distribute' in Texmelucan Zapotec (Speck 1984:156)

Different forms are involved in other verbs (e.g. *loo* vs *boo* ‘remove’, *dub* vs *ub* ‘catch’, *ruz* vs *az* ‘be beaten’ etc.). The alternations are very similar to the ones presented for Yatzachi Zapotec, which confirms that they are cognate structures. The extension of the completive stem in Texmelucan is taken to have started in the 1PL.¹⁰² According to Operstein (2002), hortative/imperative forms (which have a close morphological affinity to the completive in Zapotec) would have come to be used in the 1PL of other (e.g. indicative) TAMs. An entirely parallel development can be found in standard Italian, where the former 1PL subjunctive has been extended to the 1PL indicative.

Zapotec1: SG/1PL/2PL

Zapotec2 (Texmelucan): 1.C/2.C/3.C/1.H/1.U/1.P

5.3 Typologizing morphemes

It is usually agreed that the object of analysis of morphology is the **form** and the **meaning** of elements within the word and the relation between them. The following are some representative expressions of that sentiment:

Morphological structure exists if there are groups of words that show identical partial resemblances in both form and meaning (Haspelmath & Sims 2010:2)

The primary goal of morphological typology and theory is to analyze the ways in which languages establish relations between forms and meanings when they build words, and to discover the principles underlying the cross-linguistic variation in this domain. (Arkadiev & Klamer 2018:2-3)

¹⁰² This state of affairs (i.e. the completive root extended to the 1PL, but not the 1SG, of other tenses) seems to be have held in a variety of Zapotec documented in the XVI century.

Any attempt to typologize morphological elements, whether morphemic or morphomic, will thus need to make reference to these two main aspects of 'form' and 'meaning'. The first one relates to the segmental and suprasegmental similarities and differences between (paradigmatically) related words. The second one refers to the morphosyntactic or semantic distribution¹⁰³ of sub-word units. These two dimensions of morphological signs are, however, complex, in that they subsume different and independent axes of variation.

In order to systematically analyze these, some of the most useful frameworks are Canonical (Corbett 2005) and Multivariate (Bickel 2010) Typology. These approaches (more extensively explained in other publications, e.g. Bickel & Nichols 2002, Brown & Chumakina 2013 etc.) basically consist on taking a broad but relatively well-defined phenomenon (e.g. clause linkage, agreement, negation, gender etc.) and unpacking which are the dimensions across which particular instances of the phenomenon may vary. One can afterwards assess whether variation is random or constrained, for example by finding out whether all logically possible combinations are attested or whether naturally occurring cases actually cluster around a restricted set of frequent values.

The challenges of applying this methodology to the study of the morpheme are, obviously, considerable. First, whereas terms like 'agreement', 'negation' or 'gender' belong to the terminological toolkit of most theorists and field linguists, the term 'morpheme' does not. Consequently, finding morphemes in grammatical descriptions is a much more painstaking process. Second, there is a broad consensus in the linguistic community that phenomena like 'agreement', 'negation' or 'gender' do exist (even if they may be defined or analyzed with some discrepancies). By contrast, the term 'morpheme' has been applied to many different phenomena and objects in ways which are not always entirely clear. Some linguists even reject the notion altogether. This makes it, therefore, a more difficult object of study than the average linguistic phenomenon. It may also explain why there have not been any typological approaches to the morpheme so far.

¹⁰³ I will avoid the term 'meaning' whenever possible in subsequent discussion because it leads one to make assumptions about the realizational role of morphological forms. Very often, especially when dealing with idiosyncratic elements, it is not easy to tell when a particular element 'means' something and when it simply occurs 'meaninglessly' in particular morphosyntactic configurations. I will try to keep discussion neutral in this respect by speaking here of the 'distribution' of forms rather than of their 'meaning'.

Taking as the starting point the definition of the morpheme I advanced in Section 5.1, this section will present the various ways in which morphemes (i.e. morphological units with an unnatural morphosyntactic distribution) may differ from one another. Following the spirit of the AUTOTYP¹⁰⁴ research programme, the various dimensions/categories/variables that will be presented throughout this section have emerged inductively from the individual examples of morphemes that were presented in Section 5.2.

5.4 Variables of cross-morphomic variation

The overall **distribution** of a form can be decomposed in different dimensions: the overall domain to which all instances of a form are confined (if any), the 'shape' of its paradigmatic distribution, the total number of contexts/cells where it can be found etc. Different aspects of a morpheme's **form**, in turn, are also important: how many different exponents it has, how long these exponents are etc. If we want to reach a high level of granularity and observe generalizations and dependencies, these various largely independent variables should not be conflated. Different aspects about the distribution and form of formatives, thus, have to be captured and operationalized in different ways. I will present in this section what the variables are that underlie these and I will propose ways to measure variation objectively. After a theoretical exposition of each variable I will present an overview of the empirical data in the morpheme database of Section 5.2 (the values of all morphemes for all variables can be consulted in the Appendix). Statistical analysis will be presented in Section 5.6.

¹⁰⁴ See the principles online at <http://www.autotyp.uzh.ch/theory.html>. Most important among these is that: "Rather than starting with a predefined list of categories, AUTOTYP databases rely on an automatic generation of category lists during data input. When entering a new language, one first checks whether the previously established notions are sufficient for this language. If not, new notions are postulated (...) This procedure is time-consuming in the beginning because each new type requires review (and possibly revision) of all previous entries, but after a few dozen languages, new types become less likely to emerge and the typology stabilizes. In our experience this happens after about 40 languages are entered."

5.4.1 External morphosyntactic constraints

Not only morphemes, but also morphemes, can be circumscribed to particular inflectional subdomains. Even some of the most famous morphemes in the literature are somewhat unmorphomelike, as it were, in that they, the same as “meaningful” formatives, are limited in their distribution to particular morphosyntactic or semantic contexts/values.

Consider, for instance, the paradigmatic distribution of the Spanish L-morpheme, which occurs in the 1SG of the present indicative and through the present subjunctive. All of its cells, thus, share a tense value ‘present’. This is what I have called a strong morphosyntactic constraint: all the cells within a morpheme have a certain value in common.

Weak constraints, by contrast, are those by which a morpheme’s cells never adopt some value(s) of the ones that are possible for a given feature. One could say, for example, that the cells of Romance PYTA never have a value present. This morpheme, thus, would be subject to a weak morphosyntactic constraint.

The morphosyntactic constrainedness of a morpheme, therefore, has been measured here into two different variables simply by counting the number of distributional constraints of each kind that a morpheme’s exponents are subject to.

In the present database, morphemes have been found to range between being subject to two strong and two weak constraints simultaneously, and complete morphosyntactic unconstrainedness. Consider, first, one of the most restricted morphemes:

	Present Indicative		Present Subjunctive	
	SG	PL	SG	PL
1	‘mordo	mur’demos	‘morda	‘mordamos
2	‘mwerdes	mur’deis	‘mordas	‘mordais
3	‘mwerde	‘mwerden	‘morda	‘mordan

Table 264: Partial paradigm of *mur’der* ‘bite’ in western Asturias (Bybee 1985:73)

The diphthongizations that sometimes (e.g. in Spanish) constitute the exponents of the well-known N-morpheme have sometimes narrowed down their paradigmatic domain in this variety of Asturian in a way that its three constitutive cells are now all ‘present tense’ (strong constraint 1), ‘indicative mood’ (strong constraint 2), and ‘non-1’ (weak constraint). Despite all these morphosyntactic restrictions, the forms have stayed morphomic according to the definition used here.

In the opposite pole of this variable, many morphomic formatives have been found to be completely unrestricted in their paradigmatic distribution. Consider the following one:

	PRES		PAST	
	SG	PL	SG	PL
1	kuul-am	kuull-âp	ku'll-em	kuul-im
2	kuul-ak	kuull-ve'ted	ku'll-iĵ	kuul-id
3	kooll	ko'll-e	kuul-i	ku'll-e

Table 265: Skolt Sami *kuullâd* 'hear' (Feist 2011: 115)

In Skolt Sami *kuullâd*, the distribution of the weak grade stem *kuul-* is paradigmatically unrestricted: it can appear in both present and past, in both singular and plural, and in both first, second and third person. Its distribution is, thus, morphosyntactically unconstrained.

As for the findings in the present database, the average constrainedness of stem-based morphemes has been found to be noticeably less than that of affix-based morphemes (0.43 vs 0.70),¹⁰⁵ which fits with the role traditionally attributed to each of those elements, whereby affixes serve by default as exponents of morphosyntactic properties while stems do not (e.g. Spencer 2012). However, it must be pointed out that the full range of variation can be found in both morpheme types.

Figure 6 below gives an overview of how the morphemes in this database fare overall as for their morphosyntactic restrictedness:

¹⁰⁵ In this section, only averages and other basic descriptive statistics will be presented. The analysis of correlations between variables and statistical significance matters will be presented in Section 5.6.

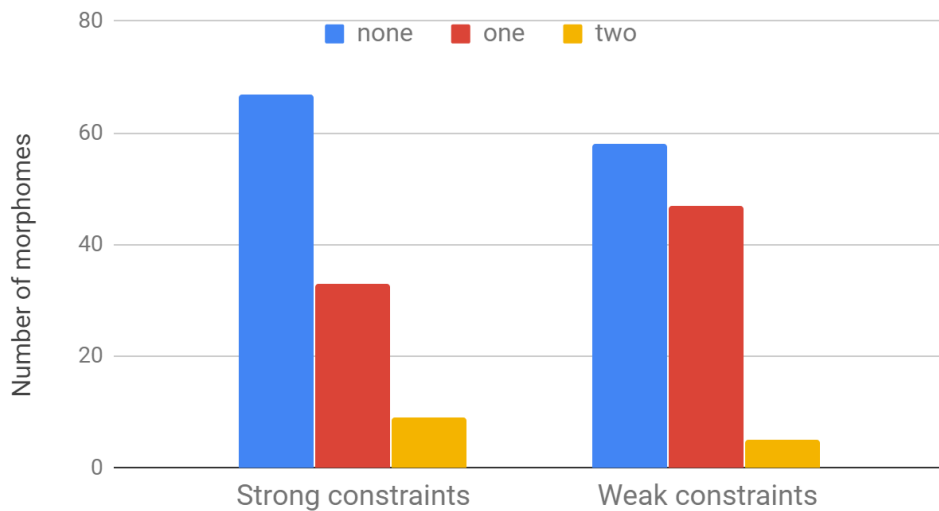


Figure 6: Morphemes and their morphosyntactic constraints

Most of the morphemes in this database (87, 65.2%) are characterized by either no constraint whatsoever or by just a single one. This is probably unsurprising. With an increased number of morphosyntactic constraints, it becomes logically more and more difficult to stay morphemic. Notice, thus, that any additional weak or strong constraint upon the distribution of diphthongization in the Asturian variety described in Table 264, would have resulted in a morphosyntactically impeccable (i.e. a morphemic) distribution.

5.4.2 Word-form recurrence

Another dimension along which morphological forms may differ dramatically is the number of distinct word forms where they appear. A morpheme, as defined here, is characterized by shared form. However, despite the sharing of segments or formatives, the cells constitutive of a morpheme can also display formal differences. On the other hand, of course, they may also be whole-word syncretic:

	<i>yatsu</i> 'brother (of a female)'		<i>yawaã</i> 'dog'	
	SG	PL	SG	PL
1	yatsu-hu	yatʃi	yawaã-hu	yawayĩ
2	yatsu-mi	yatsu-mi	yawai-mi	yawai-mi
3	yatʃi	yatʃi	yawayĩ	yawayĩ

Table 266: Possessive inflection of two Aguaruna nouns (Overall 2007:200-202)

The morphomic structure that I presented for Aguaruna in Section 5.2.2 constitutes a whole-word syncretism of 3 and 1PL. There is only one word form in all contexts and therefore, for the purposes of this variable, the word-form recurrence of the Aguaruna pattern is 1. Contrast this to the morpheme that was presented for Koasati in Section 5.2.35:

	<i>mikkon</i> 'be a chief'		<i>cákkín</i> 'catch up with'	
	SG	PL	SG	PL
1	mikko-lí	mikko-t-il-ká	cákki-l	cák-h-íl-k
2	mikko-t-is-ká	mikko-t-as-ká	cák-h-is-k	cák-h-ás-k
3	mikkó	mikkó	cák	cák

Table 267: Person-number inflection in two Koasati verbs (Kimball 1985:76, 80-81)

In contrast to the one in Aguaruna, the 2+1PL morpheme in Koasati involves different word forms in each cell, which means its word-form recurrence is 3. Of course morphomic patterns may span a much larger number of word forms (see, for example, the morphemes of Achumawi [Section 5.2.1], Icelandic [5.2.24], and Thulung [5.2.60]). There is no apparent upper-limit to how many different word-forms a particular formative may appear in.

A clarification is in order concerning how the number of different word forms has been counted here in some concrete configurations. The total of word forms in paradigms of complex agglutinative languages may number in the hundreds or thousands, which would make it difficult to retrieve an accurate estimate from descriptions. Furthermore, large paradigms are usually based on well-behaved (i.e. easily segmentable and predictable)

formatives that are simply orthogonal to the morphomic structures analyzed in this dissertation. Because of this and to simplify word-form counts, morphological distinctions orthogonal to the morphomic pattern under study have been disregarded for the purposes of this measure’s calculation here. Consider the morphome from Basque:

	Present		Past	
	SG	PL	SG	PL
1	na-bil	ga-bil-tza	nen-bil-en	gen-bil-tza-n
2	za-bil-tza	za-bil-tza-te	zen-bil-tza-n	zen-bil-tza-te-n
3	da-bil	da-bil-tza	ze-bil-en	ze-bil-tza-n

Table 268: Paradigm of Basque *ibili* ‘walk’

As shown in Table 268, this morphome (marked with the suffix *-tza* in this verb) appears in person-number values 2SG, 1PL, 2PL, and 3PL. This suffix and values are orthogonal to other morphological distinctions in the language, like tense, a fact which would multiply (from 4 to 8) the number of word-forms in the paradigm where the morphome appears. Because of this, tense morphology will be disregarded and the Basque morphome will be said here to spread only over 4 different word forms.

In the graph below I present an overview of how the morphemes in the present database (Section 5.2) classify according to this variable:

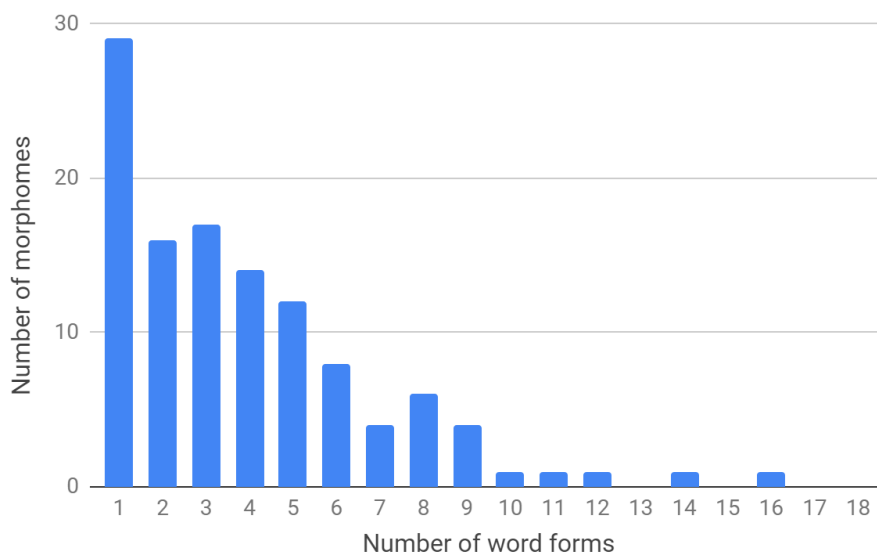


Figure 7: Morphemes and number of word-forms

As the above Figure 7 shows, whole-word syncretism is most common and has been found in 29 (26%) of the morphemes in the present study. A majority of these (19) are morphemes instantiated by affixes, and a big difference has been found here, indeed, between the word-form recurrence of stem- and affix-based morphemes, as the former span a greater number of distinct word-forms on average (5.15) compared to the latter (2.03). This suggests that, if some values are collapsed in affixes, they often remain undifferentiated in the morphological system as a whole (see e.g. Daju [5.2.16] and Ngkolmpu [5.2.46]), while the opposite is usually the case in stem-based morphemes (see e.g. the morphemes of Spanish [5.2.57], Icelandic [5.2.24], Iraqw [5.2.25], Pite Saami [5.2.54] etc.).

This may be simply a product of the fact that affixes tend to draw more and finer-grained distinctions than stems. Although this is to some extent derived from the way stems and affixes are defined, it is important to note that this is not a logical necessity. Thus, for example, in Yagaria (5.2.69) or Greek (5.2.23) we observe how person-number formal conflation of affixes are disambiguated by other morphological elements at the level of the whole word. In contrast, in Daasanach (5.2.15), the person-number formal syncretisms of the stem remain undisambiguated at the level of the word form since no other bound person-number morphology exists.

Be that as it may, from the single-word-form morpheme which has been found to be most frequent in this database, there is a downward trend according to which morphemes that include a greater number of different word forms tend to be progressively less recurrent crosslinguistically. As many as 17 (15.5%) morphemes, for example, spread over/contain 3 different word forms, 12 (10.9%) span over 5 different word forms, 4 (3.6%) extend over 9 word forms, only one (0.9%) has been found to span over 14 distinct word forms etc.

5.4.3 Paradigmatic recurrence

Morphemes, as defined here, must be instantiated by more than one allomorph. This allomorphy, however, or in other words the recurrence of a particular morphomic pattern, can take place at different levels. In the case of the morphemes that have been most

frequently discussed in the literature, the different allomorphs occur in different lexemes. Thus, the L-morpheme stem of Spanish *caber* ‘fit’ is **quep-** while that of *tener* ‘have’ is **teng-**. The different formal instantiations of the morpheme are thus found only when one looks at the paradigms of different verbs.

In other less frequently discussed cases, however, the different formal exponents appear within a single lexeme’s paradigm. We will say, thus, in these cases, that the morpheme recurs (i.e. occurs more than once) within a single lexeme’s paradigm. For this to happen, of course, each of the instantiations will need to occur in a different part of the paradigm, i.e. under a different cross-cutting value. Consider the following cases:

	Darma <i>ra</i> ‘come’ (Willis 2007:350-356)				English ‘be’			
	Non-past		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	ra-hi	ra-he-n	ra-ju	ra-n-su	am	are	was	were
2	ra-he-n	ra-he-n(i)	ra-n-su	ra-n-su	are	are	were	were
3	ra-ni	ra-ni	ra-ju	ra-ju	is	are	was	were

Table 269: Two single-paradigm-recurrent morphemes

Most of the times when a morpheme recurs in a single paradigm, its formal instantiation is affixal as in Darma. Cases like English ‘be’, however, show that this is not a logical necessity. The following Figure 8 shows the recurrence of morphemes within a paradigm:

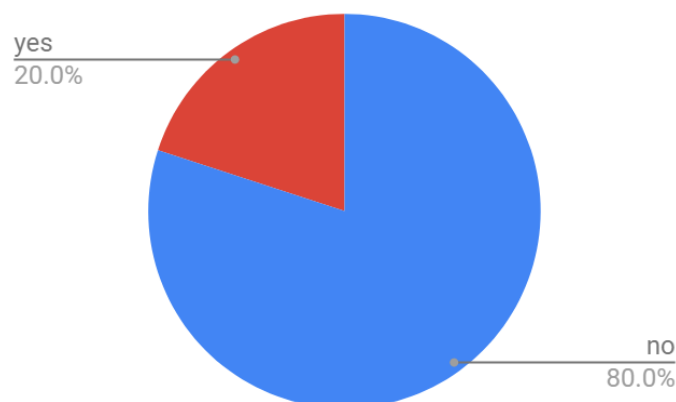


Figure 8: Morpheme recurrence within the paradigm

Overall, thus, 20% (N=22) of the morphemes in this database recur within the lexeme whereas the remaining 80% recur only across lexemes.

5.4.4 Cross-lexemic recurrence

Morphemes can also differ in their “grip” on the lexicon. Morphemes, thus, vary with respect to the number of lexical items they appear in, which can be easily measured as well as a percentage of the items in the relevant class of words. The most robust morphemes according to this variable will be a) overtly present in every single lexical item and will b) not have any exceptions. Note that these are different things:

a) Overt presence refers to those cases where the formal difference presupposed by the morpheme is present (i.e. form A appears in the cells of the morpheme and form A does not appear elsewhere). In the Spanish verb *calentar* ‘heat up’ (see Table 270), for example, the N-morpheme is overt. In *orientar* ‘orient’, however, it is covert, since the stems within the N-morpheme are indeed identical, but so are the stems in other cells of the paradigm:

	<i>calentar</i> ‘heat up’		<i>orientar</i> ‘orient’	
	SG	PL	SG	PL
1	caliente	calentamos	oriento	orientamos
2	calientas	calentáis	orientas	orientáis
3	calienta	calientan	orienta	orientan

Table 270: Present Indicative of two Spanish verbs

b) The presence of exceptions refers to those cases where the formal identity presupposed by the morpheme is not borne, i.e. the morphological identity that is supposed to hold within the cells of the morpheme conflicts with what is found, for example in the Spanish verb *venir* ‘come’:

	<i>calentar</i> 'heat up'		<i>venir</i> 'come'	
	SG	PL	SG	PL
1	caliente	calentamos	vengo	venimos
2	calientas	calentáis	vienes	venís
3	calienta	calientan	viene	vienen

Table 271: Present Indicative of two Spanish verbs

Stem vowel identity within the N-morphome (SG+3PL) is broken in *venir* and in a few other irregular verbs like *ser* 'be', *tener* 'have', and *caber* 'fit' (cf. 1SG *soy*, 2SG *eres*; 1SG *tengo*, 2SG *tienes*; 1SG *quepo*, 2SG *cabes*). Sometimes (e.g. in the case of *venir*), this happens when/because a different morphome (the L-morphome in this case) interferes with the formal identity of the 1SG with the rest of the present indicative cells.

These two variables (i.e. 'overt presence' and 'exceptions' in the lexicon) are obviously not independent because, for example, if a lexeme like *venir* constitutes an exception, this entails that the morphome is not overtly present in that lexeme. Every lexeme is classifiable, thus, as either 1) showing the morphome overtly (e.g. *calentar*), 2) abiding by the morphome without showing it overtly (e.g. *orientar*), or 3) contradicting the morphome (e.g. *venir*). For the purposes of the robustness of a morphome's presence in the lexicon, type 1) lexemes are preferred to type 2) lexemes, which are in turn preferred to type 3) ones.

When operationalizing this variable of cross-lexemic recurrence, therefore, a possibility is to measure the percentage or the number of lexemes in each of the classes. A single number seems desirable, however, and only 'overt presence' has been measured here because it is understood to be the most important of these types in the "discovery" of morphomes by either linguists or native speakers. As for the N-morphome in Spanish, for example, 426 verbs (see Herce Calleja 2016), or around 4% of the Spanish verbal lexicon, show this morphomic pattern overtly.

Everything within the range of logical possibilities has been found in the present morphome database. The most robust morphomes have been found to be present in every single lexical

item. These are overwhelmingly (but not only) affix-based morphemes. Note, however, that lexical recurrence is one of the factors that linguists rely on to assess whether something is an affix or a stem in the first place, which means that this correlation may not be particularly revealing. In the opposite pole, the least recurrent logically possible morpheme, limited to the paradigm of a single lexeme, has also been found (consider English 'be' in Table 269).

Because, even for the more thoroughly described languages, the lexicon is not (and arguably cannot be) described and measured in its entirety, the cross-lexemic recurrence of a particular morpheme will be a finer or rougher estimation depending on the evidence (i.e. source or description) available. The graph below is an overview of the extent in the lexicon of the morphemes in this database:

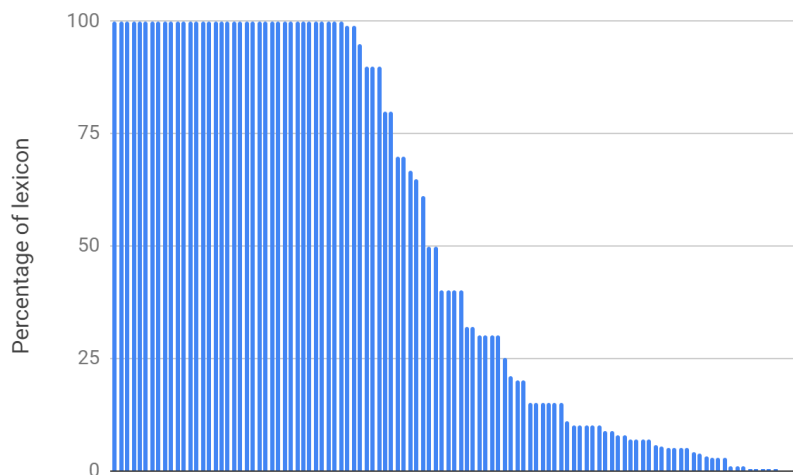


Figure 9: Cross-lexemic recurrence of the morphemes

Figure 9 shows the recurrence in the lexicon of all the morphemes in the present database, ordered from most to least recurrent. Of the 110 morphemes in this database, 37 (33.6%) appear in every single lexical item in the relevant word class. These are, I repeat, almost exclusively (but not only) affixal morphemes. On the opposite end of the scale, another 37 morphemes occur in 10% of the lexical items or less. There appears to reveal a tendency for morphemes (probably valid for formatives in general)¹⁰⁶ to either a) occur everywhere where they possibly could, or else b) be limited to a small number of (irregular?) lexemes.

¹⁰⁶ Consider the possible relationship of this finding with proposed (cognitive) principles of morphological architecture like Cartairs-McCarthy's (1994) No-Blur principle.

5.4.5 Number of exponents

Patterns of morphomic exponence may vary in their formal diversity. Some unnatural patterns are instantiated by several allomorphs and some others by just one form. Those morphemes that have several exponents, that is, several actual forms, are usually considered more systematic and robust. Consider, for example, the morphological expression of person-number agreement in Ayoreo:

	'chew'		'knock down'		'shout'		'dispatch'	
	SG	PL	SG	PL	SG	PL	SG	PL
1	yiga-se	yigaco	yiguisa-re	yiguisaco	yibi-te	yibico	yito-que	yiroco
2	baga-se	uacagaso	baguisa-re	uacaguisayo	babi-te	uacabicho	baro-que	uacarocho
3	chiga-se	chiga-se	chiguisa-re	chiguisa-re	tibi-te	tibi-te	chiro-que	chiro-que

Table 272: Person-number agreement of some verbs in Ayoreo (Bertinetto 2009)

As Table 272 above shows, the singular and third plural share form in some verbs. This suffix (or stem extension) may have a different form in different verbs. Besides the ones illustrated above we also find *-gu*, *-si*, *-ru*, *-di*, *-ra*, *-ro*, *-su*... There is a total of 28 different form(ative)s in Ayoreo associated with the contexts SG+3PL. An example with somewhat less formal diversity could be provided by the well-known L-morpheme as present in Spanish:

	<i>Caer</i> 'fall'		<i>Caber</i> 'fit'		<i>Parecer</i> 'seem'		<i>Venir</i> 'come'	
	IND	SUB	IND	SUB	IND	SUB	IND	SUB
1SG	ca-ig-o	ca-ig-a ¹⁰⁷	qu-ep-o	qu-ep-a	parez-c-o	parez-c-a	ven-g-o	ven-g-a
2SG	caes	ca-ig-as	cabes	qu-ep-as	pareces	parez-c-as	vienes	ven-g-as
3SG	cae	ca-ig-a	cabe	qu-ep-a	parece	parez-c-a	viene	ven-g-a
1PL	caemos	ca-ig-amos	cabemos	qu-ep-amos	parecemos	parez-c-amos	venimos	ven-g-amos
2PL	caeis	ca-ig-ais	cabeis	qu-ep-ais	pareceis	parez-c-ais	venis	ven-g-ais
3PL	caen	ca-ig-an	caben	qu-ep-an	parecen	parez-c-an	vienen	ven-g-an

Table 273: Present tense paradigms of four Spanish verbs

¹⁰⁷ These formal alternations are usually (but not always) considered to be part of the stem. Segmentations like *ca-ig-a*, *qu-ep-o* etc. in Table 273 are only intended to isolate the segmental material shared by the relevant paradigm cells and are not meant to convey any theoretical stance.

In a number of Spanish verbs, the present subjunctive and the 1SG present indicative share form to the exclusion of other cells. If we observe the forms of the L-morpheme in Spanish we will see that they are limited to the four above: /ig/, /ep/, /k/ and /g/.

In the lowest ranges of allomorphic variety, a morphosyntactically unnatural pattern may be instantiated by only two different forms.¹⁰⁸

	Future-potential		Hodiernal past	
	SG	PL	SG	PL
1	b-rontomo	nt-rontomo	u-rei	n-rei
2	nt-rontomo	s-rontomo	n-rei	y-rei
3M	s-rontomo	s-rontomo	y-rei	y-rei
3F	b-rontomo	s-rontomo	u-rei	y-rei

Table 274: Some durative forms of the copula in Ngkolmpu (Carroll 2016:245)

The morphomic patterns highlighted in Table 274, despite being systematic, are not revealed by a large number of different allomorphs. Often there are only two or three forms, (see e.g. s- and y-) that implement the syncretism in different tenses. The following Figure 10 represents how the morphemes in Section 5.2 fare according to their formal diversity:

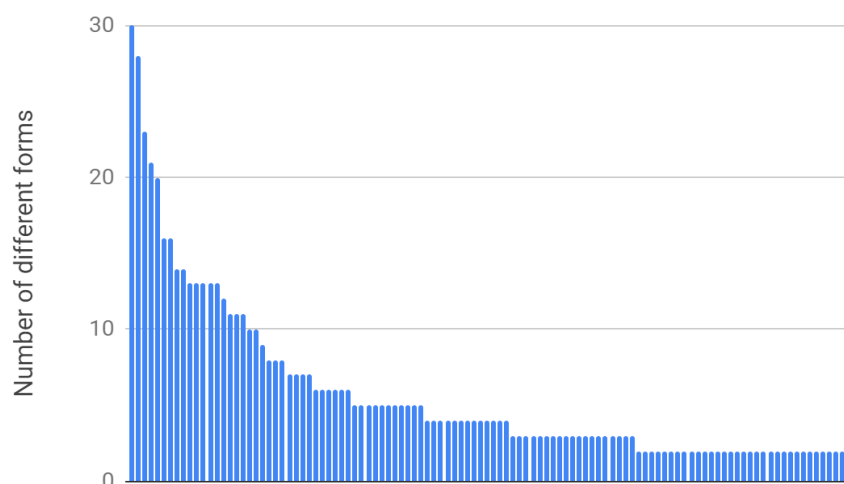


Figure 10: Formal diversity of the morphemes

¹⁰⁸ Formal instantiation with at least two different forms has been set (somewhat arbitrarily, although see Section 3.4) as the threshold to classify a pattern as systematic for the purposes of the morpheme database in Section 5.2. Single-form unnatural patterns have therefore been left out.

Figure 10 shows the formal diversity (from more to less exponents) of the morphemes in the database. Of the 110 morphemes in Section 5.2, 33 (30%) have only two different ‘allomorphs’ and, therefore, satisfy the present requirements for formal diversity only minimally. Morphemes that adopt three (19, 17%), four (13, 12%) or five (11, 10%) different forms are also relatively common. The maximum level of formal diversity found here corresponds to a morpheme in Nimboran (Section 5.2.48) that boasts up to 30 different forms. Once again, morphemes which are realized with formal alternations in the stem are characterized by a substantially higher level of formal diversity on average (6.9 allomorphs) compared to affixal morphemes (2.5 allomorphs).

5.4.6 Shared form

Another variable that will be noted in relation to morphomic variation is the “amount” of morphological substance shared between a morphome’s cells. A considerable phonological size increases our certainty that we are dealing with something significant, and must also make a pattern more ‘salient’ for language-users’ acquisition and generalization induction. A morpheme would ideally be evidenced by robust formal similarities within the cells constitutive of the morpheme and robust formal differences from the cells outside the morpheme. Those involving suppletion, for example, will be very robust (e.g. Italian SG+3PL: *vado vai va andiamo andate vanno*). The opposite pole of this would be represented by those morphological exponences where formal identity is achieved by zero, that is, where the morphological affinity results from a form being absent, rather than some form being shared (e.g. English 1SG+2SG+PL: *go go goes go go go*). Other cases will, of course, be intermediate between these two cases. For example, in Spanish *pido pides pide pedimos pedis piden*, SG+3PL cells share (only) the stem vowel to the exclusion of other cells.

This variable could be operationalized simply by counting the number of segments that the cells within a morpheme share to the exclusion of the cells outside of it. In the Italian N-morpheme, in the verb *andare* ‘go’, the shared form would be the whole stem (i.e. *va-*, i.e. 2 segments) since the stem is *and-* in other cells. In English ‘go’, by contrast, the number would

be 0, since there is no form whatsoever (in this verb)¹⁰⁹ that appears in 1SG+2SG+PL to the exclusion of the 3SG. In the Spanish example, the shared form would be 1 segment. This measure can be applied to particular forms, as in these examples, but also to whole morphemes by averaging through their different formal exponents. This will be the way this variable will be defined in this section.

The range of variation found in the present study is quite large, as morphemes have been found to range between an average of 3.7 segments for the one in Páez (that morpheme has the allomorphs *-iʔkwe*, *-kwe*, and *-we*, see Section 5.2.53) and 1 (e.g. the morpheme in Sobei, see Section 5.2.56, which has the allomorphs /o/ and /i/). All the data for this variable are presented in the graph below:

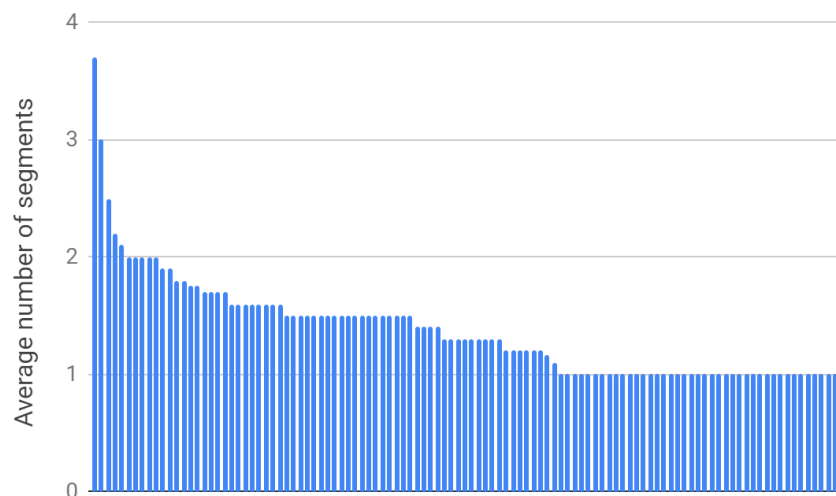


Figure 11: Average number of segments instantiating the morphemes

In a way similar to the previous variable, many (42, 38%) morphemes in this study fulfill the criteria for morphomhood only minimally and are evidenced always by just a single segment (since only segmental formal correlates were considered here, an average of 1 segment is the logical minimum). It may be worth mentioning, that, unlike in previous variables, in this one, significant differences have not been observed between stem- (1.32) and affix-based morphemes (1.42).

¹⁰⁹ Consider, however, that other verbs (namely ‘do’, ‘say’ and ‘have’) do show a segment of their own (/u:/, /ɪ/ and /v/ respectively) in these forms.

5.4.7 Locus of marking

Another aspect on which different morphomic patterns may vary is the syntagmatic locus of the exponence that reveals the morpheme. The most well-known examples of morphemes (i.e. N-morpheme, L-morpheme, and PYTA) all involve changes in the stem. Some other morphomic affinities that have been presented in this dissertation (consider the case of Ngkolmpu in Table 274) are revealed by affixes instead. Some morphemes involve both:

	SG	PL
1EXC	žoxw-žwem	žoxw-žom-e-d
1INC	-	žolw-žom-e-d
2	žoxw-žwem	žoxw-žom-e-d
3	žoxw-žom-e	žoxw-žom-e-x

Table 275: Aorist tense paradigm of Svan ‘wreck’ (Tuite 1994:4)

The Svan paradigm above shows a morphosyntactically unnatural stem-vowel alternation between /o/ and /we/, and a suffix /e/ with the same distribution. This variation on the syntagmatic locus of a morphomic exponence will be the focus of this variable. To be maximally explicit, in addition to ‘stem’ and ‘affix’, a special “mixed” category will be reserved to those morphomic patterns that involve both stems and affixes (like the one in Svan), or for which classifying the relevant forms as part of the stem or as affixal in nature is arbitrary or subject to a great degree of analytical uncertainty. The morphemes in my dataset pattern in the following way as for their *loci* within the word:

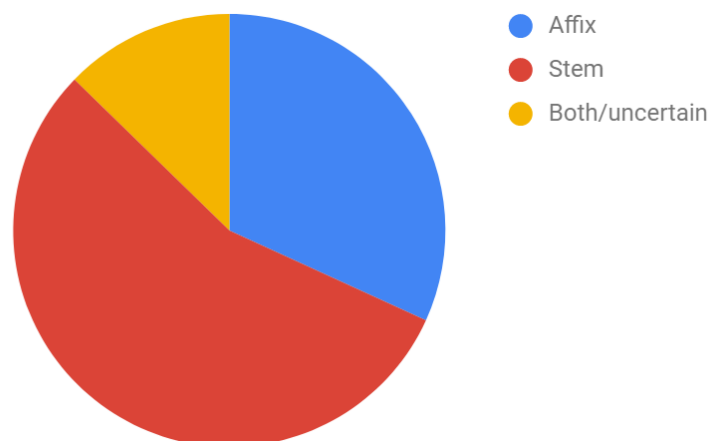


Figure 12: Proportion of affix and stem-based morphemes

A majority of the morphemes in this sample (55.5% N=61) are formal alternations in the stem, while almost a third (31.8% N=35) are affix-based. The remaining cases are either mixed or borderline cases where it is difficult to decide on the “right” segmentation.

5.4.8 Informativity

The diversity of the patterns in Section 5.2 has also revealed that morphemes (and morphemes) often differ in their informativity. By this I mean that they may differ in the extent to which they participate in the overall system of morphological contrasts in a language. I present below examples of an informative (Yagaria) and an uninformative or redundant morpheme (Jabuti), as well as an intermediate case (Jerung):

	Yagaria 'do'			Jerung 'give'			Jabuti 'get tired'	
	SG	DU	PL	SG	DU	PL	SG	PL
1	hu-ve	hu-'-ve	hu-pe	gɔk-ma	go-cum	go-kum	habä	hi-rabä
2	ha-pe	ha-'-ve	ha-ve	gɔk-nim	go-cim	go-nimme	a-rabä	a-rabä
3	hi-ve	ha-'-ve	ha-ve	gɔkt-im	gɔk-cim	gɔk-me	habä	habä

Table 276: Three morphemes with a different degree of informativity

The 2SG+1PL morpheme in Yagaria (or in other words, the alternation between *-ve* and *-pe* in this particular paradigm) may be morphosyntactically unnatural but it is as functional as it can possibly get. Because of its perfect orthogonality to the other formal distinction (*-u* vs *-a*), the morpheme in Yagaria plays a fundamental role in the expression of person-number categories in the language. Note that its presence is the only thing that distinguishes 1PL from 1SG, and 2SG from 2/3PL. It is, thus, exactly as “useful” as some of the most canonical morphemes because it is completely orthogonal to other formatives.

Contrast this to the morphomic alternation in Jabuti. The formal contrast between the stems *habä* and *rabä* is completely redundant here in that it does not increase the number of morphological distinctions in the language. More restricted affixes (*hi-* and *a-*) occur in subsets

of the 2+1PL morpheme and, because they make finer-grained distinctions, they render the stem alternation functionally superfluous.

Many morphemes, of course, are intermediate between these two kinds in that they are informative in some of its cells and redundant in others. The morphomic stem alternation in Jerung presented in Table 276, for example, is mostly redundant (e.g. the suffix *-ma* already identifies the word forms where it occurs as 1SG) but sometimes does play a decisive role in the generation of morphological contrasts. Thus, the presence of the alternant *gɔkt-* is the only feature that distinguishes 2DU and 3DU.

With regard to this variable, it may be useful to explicitly reflect on the case of morphemes where whole-word syncretism holds between the different paradigm cells (consider the verb ‘sleep’ in Alpagó below). When this happens, the morpheme is always completely informative since it is “all there is”, that is, an atomic whole as far as the paradigmatic formal contrasts in the language are concerned:

	Verb 'sleep' in Alpagó (Zörner 1997)		Verb 'sleep' in Spanish	
	IND	SUB	IND	SUB
1SG	'dɔrme	'dɔrme	'dwermo	'dwerma
2SG	'dɔrme	'dɔrme	'dwermes	'dwermas
3SG	'dɔrme	'dɔrme	'dwerme	'dwerma
1PL	dor'mɔŋ	dor'mone	dor'mimos	dur'mamos
2PL	dor'me	dor'mede	dor'mis	dur'majs
3PL	'dɔrme	'dɔrme	'dwermen	'dwerman

Table 277: Partial vs whole-word syncretism within the N-morpheme

The formal correlates of the N-morpheme in Alpagó (i.e. rhizotony and stem vowel /ɔ/) are precisely what distinguish this word form *'dɔrme* and this set of paradigm cells from other cells like e.g. the 2PL indicative *dor'me*. The morpheme is, thus, fully informative at the level of the morphological contrasts of the language. If one looks at its Spanish cognate, by contrast, one finds that the morpheme (i.e. rhizotony and the alternation /o/-/we/ in this

verb) is entirely redundant because finer-grained formal distinctions are being expressed elsewhere within the word. Only in morphemes and formatives spanning more than one word form, therefore, can we logically find that the morpheme is redundant.

Having clarified this variable and its assessment in concrete cases, this is how the morphemes in the present database pattern according to it:

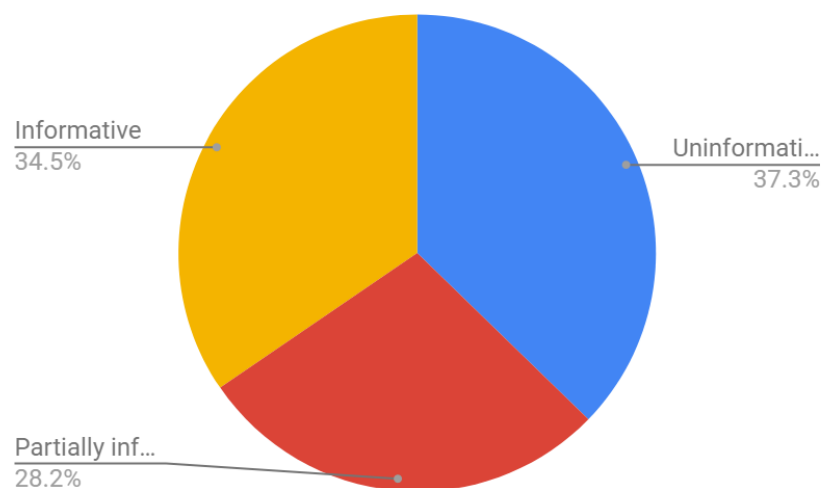


Figure 13: Informativity of the morphemes

It may be surprising to find that, although the most frequently discussed morphemes (stem alternations in Romance) are usually redundant within the broader system of morphological contrasts, the majority of morphemes (69, 63%) in the present database are at least partially informative. However, in line with other variables, it must be mentioned that there is a clear correlation between the informativity and the locus of a morpheme within the word. Thus, morphemes based on stem-alternations are most often (33, 53%) completely uninformative, whereas affixal morphemes are most often (25, 71%) fully informative.¹¹⁰ This, of course, matches well the roles usually attributed to stems and affixes respectively in the theoretical literature (but see Section 2.8 for criticism).

¹¹⁰ It would be interesting to know if these numbers are similar or different for morphemic (i.e. natural-class distributed) stem alternations and affixes. Although at the moment I cannot offer any numbers in this regard, it is my contention that, for example, “morphemic” stem alternations might well be more informative on average by having a higher degree of orthogonality to other formatives. Think, for example, of German present-past stem alternations like *sing-st* vs *sang-st* (2SG), *sing-en* vs *sang-en* (1/3PL) and *sing-t* vs *sang-t* (2PL).

5.4.9 Morphosyntactic coherence

Another dimension along which the morphosyntactic distribution of forms may vary concerns the internal morphosyntactic feature value relations within the morpheme. The overall 'shape' of a formative, as portrayed in a tabular paradigm, can be simpler (i.e. describable as the realization of some value or combination of values) or more complicated (i.e. one which necessarily has to be described disjunctively). In the first case we would probably not refer to those entities as morphomic. Consider, for example, the distribution of the following formatives in Hinuq (Nakh-Daghestanian):

	'nose'		'folk, people'	
	SG	PL	SG	PL
ABS	malu	malu	xalq'i	xalq'i
ERG	malu-y	mal-i-y	xalq'i-la-y	xalq'i-mo-y
GEN1	malu-s	mal-i-š	xalq'i-la-s	xalq'i-mo-s
GEN2	malu-zo	mal-i-žo	xalq'i-la-zo	xalq'i-mo-zo
ESS1	malu-†	mal-i-†	xalq'i-la-†	xalq'i-mo-†
ESS2	malu-qo	mal-i-qo	xalq'i-la-qo	xalq'i-mo-qo

Table 278: Two Hinuq noun paradigms (Forker 2013:55)

All non-absolute forms of the noun in Hinuq are formed on the basis of the same stem. The so-called 'oblique stem' may differ from the absolute form in many different ways: by the addition of various suffixes (*-i*, *-la*, *-mo* above), ablaut, shift of stress, deletion of the final consonant... However, the distribution of the forms is straightforward. All contexts where the same form is used share a number value and (arguably) a case value 'oblique'. Because of that, their distribution need not be described disjunctively. It displays a rectangular shape when represented in a paradigm and it does not count as a morpheme here.

However, within the distributions that cannot be described as the realization of a value or two or more values conjunctively (i.e. within morphomic distributions) there is still a great amount of variation. Some forms' distribution is such that the morphosyntactic contexts that they are associated with do not share any value whatsoever:

	SG	PL
NOM	bean	mná
GEN	mná	ban

Table 279: Declension of Irish 'woman' (Mac Congáil 2004:51)

The cells where *mná* is found (GEN.SG and NOM.PL) do not share any value. These forms are the least natural because they effectively 'mean' one thing and the opposite. Their geometrical paradigmatic shape is characterized by a lack of spatial contiguity. Other forms' distribution, by contrast, span a geometrically contiguous area of the paradigm. This means that the morphosyntactic contexts where they are used can all be "visited" by changing one feature value at a time, which makes them comparatively more 'natural'. Unlike in a morphemic distribution like Hinuq above, however, the values of more than one feature are involved and the overall paradigmatic shape is not rectangular:

	Form 1		Form 2	
	SG	PL	SG	PL
M/F	-ari	-iri	-ada	-idi
N	-ari	-ari	-ada	-ada

Table 280: Perfective positive endings in Northern Akhvakh (Creissels 2008)

You can see above how the shaded pattern of gender-number agreement in perfective positive forms is morphomic in Northern Akhvakh. The same form is used to agree with singular arguments (irrespective of their gender) and also with neuter plurals.

More complex bidimensional patterns can also be found. As presented in Section 5.2.11, gender-number agreement is morphomic in Burmeso. Particular exponents appear, depending on the noun, agreeing sometimes with a singular noun, sometimes with a plural one and yet other times with both number values:

Gender	Conjugation 1		Conjugation 2	
	SG	PL	SG	PL
II	g-	s-	n-	t-
VI	g-	g-	n-	n-
V	j-	g-	b-	n-

Table 280: Conjugations in Burmeso, partial paradigm (adapted from Donohue 2001:100-102)

All the previous examples have been 'bidimensional' in that they have involved distributions based on the possible values of just two features. This is, of course, not always the case. Inflectional paradigms are often structured around more than just two orthogonal morphosyntactic features. Consider the following morphomic pattern:

	Stem of 'sleep'						Stem of 'think'/'call'					
	Masculine			Feminine			Masculine			Feminine		
	SG	DU	PL	SG	DU	PL	SG	DU	PL	SG	DU	PL
1	e-	e-	e-	e-	e-	e-	ah-	ah-	ah-	ah-	ah-	ah-
2	e-	e-	ap-	e-	e-	e-	ah-	ah-	s-	ah-	ah-	ah-
3	ap-	e-	ap-	ap-	e-	e-	s-	ah-	s-	s-	ah-	ah-

Table 280: Stem alternation patterns in Menggwa Dla (de Sousa 2006:539-541)

Menggwa Dla verbs agree in person, number and gender. Some verbs have a pattern of stem alternation whereby the same stem is used for 3SG (for both masculine and feminine) and for 2PL/3PL masculine. Reference to all three features is thus necessary to capture the distribution of stem alternants like *ap-* or *s-*.

The variable that this section is presenting (i.e. the paradigmatic "shape" or naturalness of a pattern) can be operationalized as the average percentage of feature-values shared between its cells. In the case of the morphome of Menggwa Dla (Table 280), describing its paradigmatic distribution requires reference to four person-number-gender cells: 3.SG.M, 3.SG.F, 2.PL.M, and 3.PL.M. Between these, there is a total of six possible pairs: (2.PL.M, 3.SG.M), (2.PL.M, 3.PL.M), (2.PL.M, 3.SG.F), (3.SG.M, 3.PL.M), (3.SG.M, 3.SG.F), and (3.PL.M, 3.SG.F). The

percentage of shared values of each of these pairs is: 33.3%, 66.6%, 0%, 66.6%, 66.6%, and 33.3% respectively. The average percentage of shared values is, thus, 44.4%.

The patterns presented throughout this section vary between 0% and 50% naturalness: 0% Irish, 25% Burmeso, 33.3% Northern Akhvakh, 44.4% Menggwa Dla, 50% Hinuq. As defined here (see also Herce 2020), this variable must indeed take a number between 0% and 50%. However, because of the present morphomehood requirements (see Section 5.1.1), structures of the latter kind have not been included in this database and thus no morphome here reaches the logical maximum of 50%. Observe the morphosyntactic coherence of the morphemes in the present database:

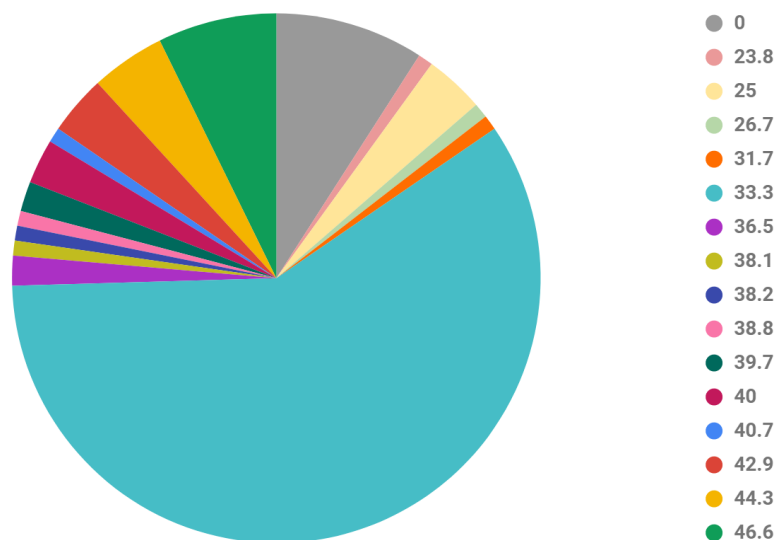


Figure 14: Morphosyntactic coherence (MC) of the morphemes

A total of 16 different values of MC have been found in the present morphome database, ranging from 0% to 46.6%. The most striking finding is the prevalence of the value 33.3%, which characterizes a majority of the morphemes (65, 59%). See Section 5.4.12 for an explanation of this fact. Another, related finding to emerge from the present research is that morphemes are not random with respect to their relation to morphosyntactic values but rather tend to be closer to naturalness (i.e. to the logical maximum of MC=50%) than to perfect “antinaturalness” (i.e. MC=0%).

5.4.10 Morphome paradigm size

Another aspect about a morphome's distribution that has not been presented so far here concerns the number of content cells required in the description of its paradigmatic distribution. Consider, for instance, the morphome of Fur:

	'tie' imperfective		'hang' imperfective		'grind' imperfective	
	SG	PL	SG	PL	SG	PL
1	ʔirg-ɛl	kirg-ɛl	ʔalg-ɛl	kalg-ɛl	ʔawan	kawan
2	ʝirg-ɛl	birg-ɛl	ʝalg-ɛl	balg-ɛl	ʝawan	bawan
3.HUM	rig-ɛl	kirg-ɛl-ɪ	ɪrg-ɛl	kalg-ɛl-ɪ	κɔɔn	kawn-ɛ
3.NHUM	rig-ɛl	rig-ɛl-ɪ	ɪrg-ɛl	ɪrg-ɛl-ɪ	κɔɔn	κɔɔn-ɛ

Table 281: Partial paradigms of three Fur verbs (Waag 2010)

Based exclusively in the presence of the morphomic exponents (e.g. κɔɔn) shaded above, and assuming orthogonality of the values of different features, one would need a content paradigm with only 4 cells to capture these formatives' distribution:

	SG	PL
HUM		
NHUM		

Table 282: Fur morphome content paradigm

For the purposes of cell-counting here, values that behave identically concerning the presence of the morphome will be combined into a single one, independently of these values' semantic content. Consider, for instance, the following morphome:

	'carry' (whole form, past)		'close' (stem+suff.)		'throw' (stem)		'bathe' (stem)	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	ni-gongo:	ni-rango:=so'	rogo	ru'gwa	grwi'ya	dri'ya	wan	na
1INCL	-	ni-rango:=lo'	-	ru'gwa	-	dri'ya	-	na
2	ni-rango:	ni-rango:=la	rugwa	ru'gwa	dri'ya	dri'ya	na	na
3	ni-gongo:	ni-rango:	rogo	ru'gwa	grwi'ya	dri'ya	wan	na

Table 283: Some inflectional forms in Malinaltepec Me'phaa (Suárez 1983: 155, 158, 160)

Because, with respect to the presence or absence of the morphomic exponence, the first person behaves like the third, the above morphomic affinity is therefore also reducible to a four-cell content paradigm:

	SG	PL
1/3		
2		

Table 284: Malinaltepec M. morphome content paradigm.

More complex morphemes will of course require reference to a greater number of features and values and will contain more cells in their content paradigm. Consider the following:

	<i>cogliere</i> 'pick'		<i>dire</i> 'say'	
	IND	SUBJ	IND	SUBJ
1SG	colgo	colga	di[k]o	di[k]a
2SG	cogli	colga	di[tʃ]i	di[k]a
3SG	coglie	colga	di[tʃ]e	di[k]a
1PL	cogliamo	cogliamo	di[tʃ]iamo	di[tʃ]iamo
2PL	cogliete	cogliate	dite	di[tʃ]iate
3PL	colgono	colgano	di[k]ono	di[k]ano

Table 285: Present tense of two Italian verbs (Maiden & Robustelli 2014)

Based on the presence or absence of the shaded stems, the Italian L-morphome is irreducible and requires a content paradigm with 12 different cells:

	IND		SBJV	
	SG	PL	SG	PL
1				
2				
3				

Table 286: Italian L-morphome content paradigm

Because of the way morphomehood has been defined here (only allowing morphomes with orthogonal values in different axes), this variable can only take a discrete number of values, namely: 4 (2x2), 6 (2x3), 8 (2x2x2), 9 (3x3), 12 (2x3x2/4x3), 16 (4x4), 18 (3x3x2) etc. Here is how the morphomes in my database classify according to this variable:

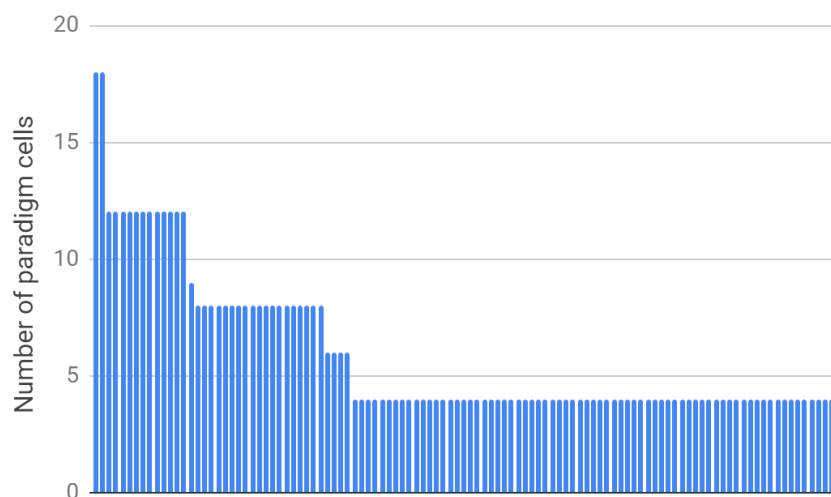


Figure 15: Morphomic paradigms' size

As Figure 15 shows, the majority of the morphomes in my dataset (65.5%, n=72) can be captured in a 2x2 (i.e. in 4-cell) paradigm and have therefore the lowest possible complexity for a morphome according to this measure. Higher complexity distributions are progressively less frequent, which suggests that the relative simplicity of a pattern may be a significant factor on a morphome's emergence and/or survival in a language (see also Section 5.4.12).

The last two variables/measures described here (Sections 5.4.9 and 5.4.10) capture different facts of a morphome’s paradigmatic distribution. Consider the following two patterns:

	<i>njorggad</i> ‘whistle’ (Feist 2015:204,210)				<i>perder</i> ‘lose’	
	Present		Past		Present	
	SG	PL	SG	PL	SG	PL
1/2	njoor[ɣ]	njorgg	njurgg	njoor[ɣ]	pierd	perd
3	njorgg	njorgg	njoor[ɣ]	njurgg	pierd	pierd

Table 287: Content paradigms of two morphemes in Skolt Saami (left) and Spanish (right)

The two morphemes above (weak stem in Skolt Saami and diphthongized stem in Spanish) have an identical morphosyntactic coherence of 33.3%. However, they constitute clearly different patterns. For the former, one needs to refer to three different features/dimensions in order to capture its distribution whereas for the latter two features suffice. Thus, the former morphome’s paradigm’s size is $2 \times 2 \times 2 = 8$ while the latter’s is only $2 \times 2 = 4$.

Although these measures are useful and complementary, they do not, by themselves, exhaust the variation found in this domain. Thus, knowing a morphome’s MS coherence (Section 5.4.9) and its paradigm size (this section) does not suffice to unmistakably identify a morphome’s paradigmatic extent, i.e. its geometrical shape. Morphome distributions also differ in the number of features required in their description (e.g. $2 \times 3 \times 2$ and 4×3 paradigms all have 12 cells but differ in this respect), and in the number of paradigm cells that the morphomic exponents span. However, even all these four variables would not suffice to capture a distribution unmistakably. Consider the following two morphemes:

	Stem of ‘drink’ in Tol					Stem of ‘fit’ in Spanish			
	Present		Future			Subjunctive		Indicative	
	SG	PL	SG	PL		SG	PL	SG	PL
1	miʔ	myis	miʔ	mis	1	quep	quep	quep	cab
2	mis	mis	miʔ	mis	2/3	quep	quep	cab	cab

Table 288: Streamlined paradigm of two morphemes

These two morphemes in my database take identical values as per MS coherence (46.6%), paradigm size (8), features involved (3), and paradigm cells spanned by the morpheme (5). However, they still constitute different configurations, by which I mean that they are not simply rotational or row/column order variants of each other. This suggests that there is still work to be done with regard to the measurement and typologization of paradigmatic distributions. I shall leave this for future research.

5.4.11 Cross-linguistic morpheme recurrence

Although typological uniqueness has sometimes been thought of as a diagnostic of morphomicity (Maiden 2018b:22), there is no reason whatsoever, in my opinion, to believe that being typologically unique should be a definitional requirement in the identification of any linguistic phenomenon (see Section 2.6). Under the approach espoused here, therefore, morphomic structures can, indeed, be found in unrelated languages. The cross-linguistic uniqueness or generality of a particular morphomic structure, thus, can be explored as a further variable of cross-linguistic morphomic research. It will be measured here by the number of non-cognate paradigmatic-distributionally identical morphemes in my database.

The morphological component tends to be extremely variable across languages (in terms of the inflectional categories encoded, number of values etc.). Therefore, we can only expect to find cross-linguistic recurrence in those inflectional domains characterized by a certain degree of universality. Grammatical cases and tenses vary quite drastically across languages in their number and the way they “chop” the functional-syntactic space. Grammatical genders can also divide up the lexicon in a quite variable number of classes by using quite heterogeneous semantic and formal criteria. Person/number features, by contrast, appear to offer a more limited gamut of choices. In their tracking of referents, all languages seem to be concerned with the same speech-act roles of speaker, addressee and non-participant. Similarly, in the domain of number, chopping the domain into one (SG) vs more than one (PL) individuals seems to be the basic distinction upon which languages may occasionally add additional ones. The following person-number morphomic patterns have been found to be recurrent in the present morpheme database:

SG+3PL

A morphological affinity involving the singular (1, 2 and 3) and the 3PL has been found in the present sample in a total of seven languages from five different stocks: Ayoreo (Zamucoan), Daju (Dajuic), Greek (IE), Jerung (Sino-Tibetan), Spanish (IE), Sunwar (Sino-Tibetan) and Tol (Jicaquean). Here is a streamlined presentation of two of these structures:

	Daju 'drink' ¹¹¹		Ayoreo 'fill up'	
	SG	PL	SG	PL
1	ur-o	ur-ciga	ɲĩ-rate	ɲĩ-ra-ko
2	ur-o	ur-cini	mã-rate	wakã-ra-tso
3	ur-o	ur-o	tɛĩ-rate	tɛĩ-rate

Table 289: Two languages showing SG+3PL morphemes

The reason for the recurrence of this structure could be related to the well known form-frequency correlation known as Zipf's Law (1935). The singular (vs the plural) and the third person (vs the second and the first) are frequently characterized by shorter or zero forms opposed to longer or non-zero forms to signal the 'marked' values (see Section 4.1.1.1). Thus, for example in pre-Daju, both SG and 3PL must have been characterized by zero, opposed to overt suffixes in 1PL and 2PL. Later sound changes would have been responsible for the later emergence of formal divergences between suffixed and unsuffixed forms and for the acquisition of overt exponents by the erstwhile zero-marked SG+3PL. Largely the same scenario applies in the diachronic emergence of the morpheme in Ayoreo.

3+1SG

The morphological affinity of third person and 1SG is very similar to the previous one both in its paradigmatic extension and probably in terms of its causes. In my sample it has been found in three different languages from three different families: Italian (IE), Chiquihuitlán Mazatec (Oto-Manguan) and in Wambisa (Chicham). The pattern is presented below:

¹¹¹ For the sources and for additional information on each of these morphemes, please consult the corresponding language's section in Section 5.2.

	C. Mazatec 'remember'		Italian 'do' Past	
	SG	PL	SG	PL
1	base ¹¹²	časín	feci	facemmo
2	čase	časun	facesti	faceste
3	base	base	fece	fecero

Table 290: Two languages showing 3+1SG morphemes

The reason (or one of the reasons) for the relative recurrence of this pattern in my database is, I believe, similar to the one mentioned for the previous morphomic pattern. The 1SG, 3SG and 3PL are the most frequent person-number value combinations in natural speech. This often leads to these forms having zero or short agreement formatives opposed to overt or longer markers elsewhere in the paradigm, a configuration which may result in a morphomic affinity given the right sound changes and analogical developments (see Section 4.1.1.1).

PL+1SG/2SG/3SG

A morphological affinity between the plural persons and one of the singular ones has also been found here to be common. PL+1SG has been found to occur in five languages: Barai (Koiarian), Luxembourgish (IE), Nivkh (Isolate), and Vitu and Vurës (Austronesian):

	Luxembourgish 'be', present		Barai, Future sequence		Nivkh, FUT Distant converb	
	SG	PL	SG	PL	SG	PL
1	sinn	sinn	-kuva	-kuva	-non	-non
2	bass	sidd	-kuma	-kuva	-ror	-non
3	ass	sinn	-kuma	-kuva	-ror	-non

Table 291: Three languages showing PL+1SG morphemes

¹¹² The Mazatec paradigm may not suggest at first sight that length of expression could have been responsible for the emergence of this particular pattern. Individual exceptions would not necessarily invalidate a tendency, however, there are reasons to believe that this diachronic origin may also apply to this particular case. These stem alternations in Mazatec originated from the incorporation of auxiliaries that already often showed 1SG/3 vs 1PL/2 alternations (see Baerman 2013). The shaded cells, however, were characterized by greater formal diversity in both stem-initial consonants and suffixes (see Pike 1948). This could point to a zero (i.e. *ba-* < *vʔa*) vs prefixed (e.g. *ča-* < *čʔa* < **č-vʔa*) configuration as the origin of this morpheme.

Similarly to the morphomic pattern above, a morphological affinity of PL+2SG has been found in four languages in the sample: Basque (Isolate), English (IE), Koiari (Koiarian), and Malinaltepec Me'phaa (Oto-Manguenan):

	Koiari 'see' Perfect		M. Me'phaa 'carry' Past		Basque 'be' Present	
	SG	PL	SG	PL	SG	PL
1	ereva-nu	ereva-nua	ni-gongo:	ni-rango:	na-go	ga-u-de
2	ereva-nua	ereva-nua	ni-rango:	ni-rango:	za-u-de	za-u-de-te
3	ereva-nu	ereva-nua	ni-gongo:	ni-rango:	da-go	da-u-de

Table 292: Three languages showing PL+2SG morphemes

A morphological exponence whereby PL+3SG systematically share formatives to the exclusion of other cells, in turn, is also recurrent in the present sample. In the same way as the previous morpheme, this pattern has also been found to occur in four languages: Kele (Austronesian), Nen (Yam), Svan (Kartvelian) and Wambisa (Chicham):

	Kele 'basket' (stem)		Svan 'cut' Aorist		Wambisa 'daughter'	
	SG	PL	SG	PL	SG	PL
1	dópu	dábo	o-č'k'or	o-č'k'wer-d	nauantu-ru	nauantu-rĩ
2	dópu	dábo	a-č'k'or	a-č'k'wer-d	nauantu-rumi	nauantu-rĩ
3	dábo	dábo	a-č'k'wer	a-č'k'wer-x	nauantu-rĩ	nauantu-rĩ

Table 293: Three languages showing PL+3SG morphemes

The last three person-number patterns share some very obvious similarities, which is the reason why they have been presented together here. The three of them involve, obviously, the falling-together of all the plural cells with a single singular cell. They also have been found in a similar number of languages in the present sample, which points (with all due reservations due to the small numbers involved) to a comparable cross-linguistic recurrence.

The explanation I wanted to put forward here to explain the recurrence of these patterns might be more tenuous than in the case of the other recurrent patterns described here before

as it relies partially on chance. Whole natural classes like PL will frequently share forms. Relatively ‘marked’, more infrequent classes like PL will tend to share forms to a greater extent than more ‘unmarked’/frequent natural classes like SG (see Tables 61 and 62). Diachronic accidents¹¹³ would thus more frequently result in shared forms between the PL cells and a SG cell than in shared forms between all the SG cells and one of the PL cells. I furthermore will propose here that such paradigmatic configurations, once in place, might also be somewhat more stable than patterns like SG+1PL or SG+2PL. If the deviations from naturalness (i.e. SG vs PL) occur in more frequent (i.e. SG or 3PL) paradigm cells, this may well translate into a better learnability and greater resilience of those patterns.

As further proof that this explanation might be on the right track, some other fact must be mentioned. Other morphomic patterns that involve a relatively infrequent natural class falling together with a relatively frequent cell outside of it can be found relatively often in the present database (consider e.g. the morphemes of Sobei [Irrealis+3SG.Realis], Udmurt [Future+3PL.Present], and the Spanish L-morpheme [Subjunctive+1SG.Indicative]). The opposite patterns (e.g. Present+2PL.Future, Realis+1PL.Irrealis etc.) have not been found. This fact, and the recurrence of the PL+1SG, PL+2SG and PL+3SG morphemes may therefore be the combined result of both mutational constraints, i.e. patterns like Present+2PL.Future are less likely to arise, and functional pressures, i.e. those patterns are more also likely to be lost, for example, by falling back analogically to the closest natural class distribution, in this case ‘present’ (see Section 4.2.4.1 for insights on the latter diachronic developments).

2+1PL

A pattern where 2SG, 2PL and 1PL form a class for the purposes of morphological exponence is also relatively recurrent cross-linguistically according to the data collected in this database. This morphomic class appears in five languages: Darma (Sino-Tibetan), Jabuti (Macro-Je), Karamojong (Nilotic), Koasati (Muskogean), and Mazatec (Oto-Manguenan):

¹¹³ See the sections on Basque and Malinaltepec Me’phaa for two very different kinds of “accidents”.

	Karamojong Indicative		Koasati 'be a chief'		Jabuti 'fall'	
	SG	PL	SG	PL	SG	PL
1	aka-	iki-	mikko-lí	mikko-t-il-ká	hökü	hi-nökü
2	iki-	iki-	mikko-t-is-ká	mikko-t-as-ká	a-nökü	a-nökü
3	a-	a-	mikkó	mikkó	hökü	hökü

Table 294: Three languages showing 2+1PL morphemes

The explanation I would like to propose here for the cross-linguistic recurrence of this particular pattern has to do with its closeness to a natural distribution. As I have argued elsewhere (Section 2.9.1), 2+1PL.INCL is a semantic natural class, since it is coextensive with reference to the addressee. When clusivity distinctions are lacking, the 1PL refers to a group of individuals that most often includes, rather than excludes, the addressee. This fact increases the viability of a synchronic formal allegiance of some sort between 1PL and 2. Diachronically, in turn, it probably means that changes that result in a 2+1PL paradigmatic configuration are not strongly dispreferred (e.g. when clusivity is lost, the earlier 1PL.INCL form may be the one taking over the plural exclusive meaning).

A similar explanation to this one could be offered for the fact that 2SG+1PL is the only diagonal morphomic person-number pattern which has been found here repeated in unrelated languages; namely in Ngkolmpu (Yam) and in Yagaria (TNG) (see Herce 2018b).

5.4.12 Recurrence of geometrical patterns

I have demonstrated in Section 5.4.11 that person-number agreement inflection is a morphological domain particularly appropriate to find cross-linguistically recurrent morphomic structures. Recurrence of a pattern in other domains can be found as well, however, provided some degree of flexibility is allowed with respect to the actual values and categories involved in a morpheme. Consider, for instance, the following gender-number inflectional systems:

Burmeso Conjugation 1 gender affixes			Khinalug Set 2 gender affixes		
Gender	SG	PL	Gender	SG	PL
II Female, animate	g	s	II Female	z	v
III Miscelaneous	g	j	III Animate	v	j
IV Mass nouns	j	j	IV Inanimate	j	j
I Male	j	s	I Male	j	v

Table 295: Gender-number inflections in two languages

The (6) genders of Burmeso (Isolate) are obviously different from the ones (4) in Khinalug (Nakh-Daghestanian), even if/when the numbers and/or the semantic labels given to them in descriptions might occasionally be identical. Abstracting away from that fact, however, both languages show an exponence pattern whereby the singular of one gender (I), the plural of another gender (III), and both the singular and the plural of yet another gender (IV) constitute a single class for morphological purposes. Although the values and categories vary, then, the patterns are still “the same” at some level of abstraction. This same pattern is also found in the gender-number inflectional system of Mian (see Section 5.2.44). The gender-number morphemes found in Burushaski (5.2.12), Ket (5.2.32) and Northern Akhvakh (5.2.51) are also the same in that the three of them merge the singular of some gender and the singular and plural of some other (inanimate) gender.

An abstraction away from the concrete values and categories involved in a morpheme, thus, will allow us to focus on the geometric patterns exclusively and observe another potential sort of (more abstract) cross-linguistic recurrence. All of the recurrent person-number morphemes presented in Section 5.4.11, for example, are geometrically “the same” in that they are instantiations of a pattern where some of the values of feature Y under a given value/set of values A of an orthogonal feature X are merged with a subset of those Y-values in another X-value/set of values B. This is cumbersome to explain in running text but is easy to represent geometrically:

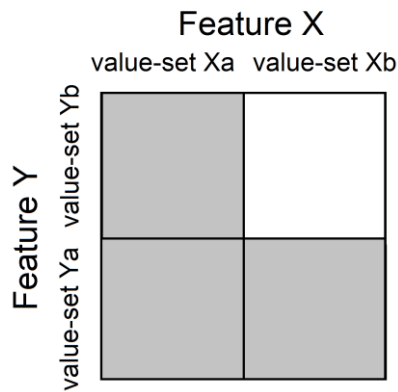


Figure 16: Schematic representation of Pattern A

Thus, if we abstract away from the concrete features and value-sets involved, the SG+3PL morpheme, the PL+2SG one, and many others (e.g. Northern Akhvakh’s M.SG+F.SG+N) will constitute instantiations of this geometrical paradigmatic pattern because they are all merely rotational or row-order variants of each other:

Koiari Perfect			Greek augment			Akhvakh disjunct		
	PL	SG		SG	PL		SG	PL
1/3	-nua	-nu	1/2	e-	-	M/F	-ari	-iri
2	-nua	-nua	3	e-	e-	N	-ari	-ari

Table 296: Some Pattern A type morphemes

This, which I will call here Pattern A, is by far the most prevalent morphomic pattern in the present sample. It is found in a total of 63 different morphemes, or 57.3% (N=110) of them.

The second most recurrent morphomic pattern is one where the Y-value sets that share a morphological affinity under X-value A and under X-value B are a disjoint set. Once again, geometrical representation helps to understand what is meant:

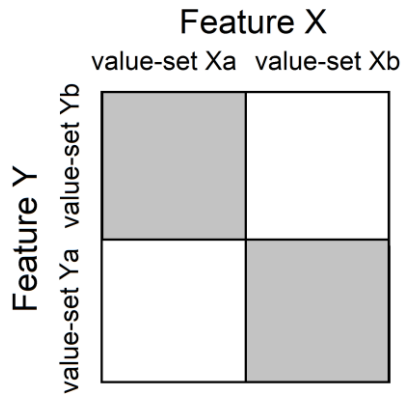


Figure 17: Schematic representation of Pattern B

This geometrical pattern is found in a total of 10 morphemes (9.1%) in the present database. This is the case of the morphemes in Khaling (Tibeto-Burman), Irish (IE), and Wutung (Sko):

Khaling 'look nice'			Irish 'woman'			Wutung 'be here'		
	SG	PL		PL	SG		SG	PL
1	bū:	bʌŋ	NOM	mná	bean	1	punga	nua
2/3	bʌŋ	bū:	GEN	ban	mná	2	mua	punga

Table 297: Some Pattern B morphemes

To present the third most recurrent morphomic pattern in this database I have to enter the realm of the three-dimensional. In a total of seven morphemes (6.4%), morphological affinities in the paradigm follow a pattern like the following:

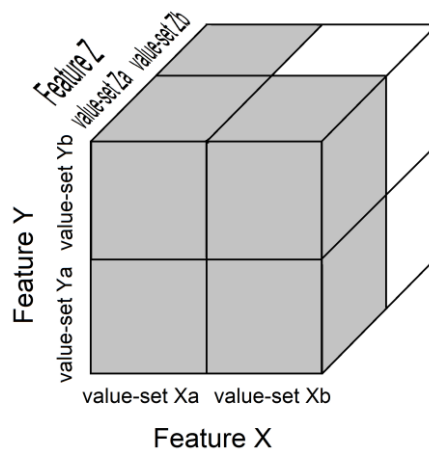


Figure 18: Schematic representation of Pattern C

Because of the obvious limitations of a three-dimensional visualization, these patterns will be represented in two dimensions from now on:

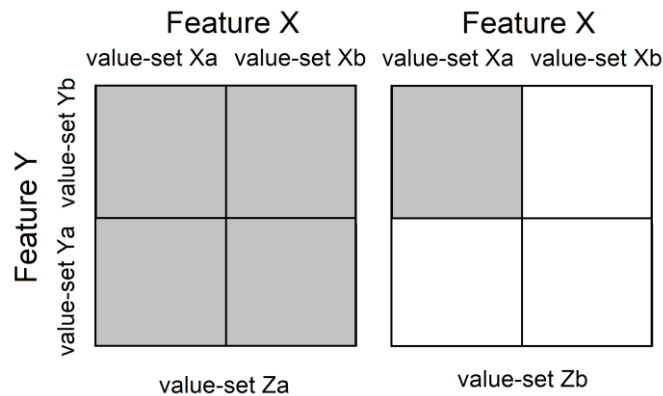


Figure 19: Schematic representation of pattern C

This morphomic pattern has been found among others in Spanish (IE) and in Udmurt (Uralic). These two morphemes are schematically represented below:

Spanish 'fall', stem					Udmurt suffix, conjugation 2				
	Subjunctive		Indicative			Future		Present	
	SG	PL	SG	PL		PL	SG	PL	SG
1	caig-	caig-	caig-	ca-	3	-lo	-lo	-lo	-
2/3	caig-	caig-	ca-	ca-	1/2	-lo	-lo	-	-

Table 298: Two pattern C morphemes

It would be quite tedious to present every single geometric pattern here one after the other. Let it, therefore, suffice to mention that a total of five other geometrical patterns have been found here represented by more than one morpheme. Observe more succinctly in Figure 20 below the recurrence of the most frequently occurring morphomic patterns:

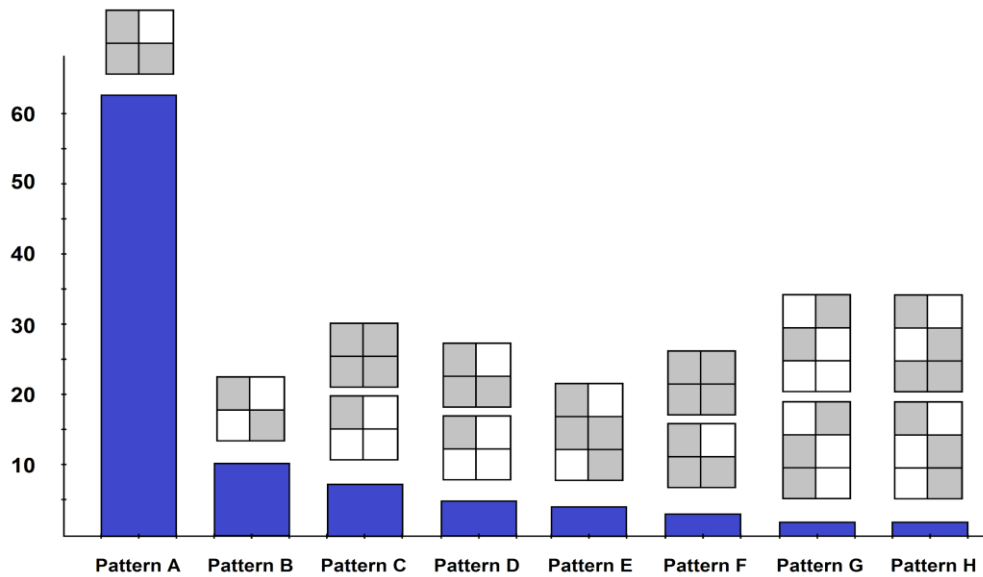


Figure 20: Geometrical morphemic patterns and their recurrence

Besides the ones above, another 13 patterns have been found represented by just a single morpheme, which brings the total number of geometrical patterns in this database to 21. It is very revealing, thus, that a single one of them, Pattern A, accounts for 57% (N=63) of all the morphemes. The skewness of this distribution is, therefore, considerable. Another circumstance that may be noted in Figure 20 above is that there appears to be a (probably unsurprising) inverse correlation between the complexity of a pattern and its cross-linguistic recurrence. In this way, the two simplest possible patterns (i.e. the ones whose shape can be captured in a 2x2 table) are the two most frequent ones and the next simplest one (2x3 pattern D) is also among the most recurrent. The most complex geometrical patterns (2x3x2 patterns like G and H, and 3x3x2 patterns) seem to be, by contrast, predominantly unique.

This complexity dimension, however, does not help with explaining the notable difference found between the cross-linguistic recurrence of patterns A and B. The reason for this dramatic difference, I believe, must be sought in the geometrical contiguity of the former pattern and the absence thereof in the latter.¹¹⁴ Paradigmatic patterns which are

¹¹⁴ That geometrical contiguity is the other main explanatory factor involved in the recurrence of different patterns is also suggested by the comparison of 2x2x2 patterns. A total of 7 patterns have been found of these kind. Two (Skolt Saami1 and Skolt Saami2) are geometrically non-contiguous (morphosyntactic coherence 33.3% and 40% respectively) and both of these are found in just one morpheme. Of the geometrically contiguous patterns, by contrast, most (3 out of 5, see C, D, and F in Figure 20) occur in at least 3 different languages. The one with the highest morphosyntactic coherence (C, at 46.6%) is also the most frequent.

geometrically contiguous (or alternatively, those with a higher morphosyntactic coherence, see Section 5.4.9) must either a) emerge more frequently in the course of language change, or b) enjoy a greater learnability and diachronic stability once they arise (see Pertsova 2011), or c) both.

We can indeed find signs of functional and diachronic pressures in this respect beyond the recurrence patterns that we are trying to explain. On the one hand, geometrically non-contiguous paradigmatic distributions must arise less frequently than contiguous ones. This is evidently the case when a morphomic pattern arises through analogical processes. As shown in Section 4.1.3.1, features and values are important structuring forces in grammar. Thus, when forms are extended analogically (or by way of secondary grammaticalization processes) to other values/paradigm cells, the source and the target are most often adjacent in that they share at least one value. Consider the following morphemes from Section 5.2:

	Basque 'walk' Past		Biak 'eat'			
	SG	PL	SG	DU	TR	PL
1EXCL	nen-bil-en	gen-bil-tza-n	y-an	nuy-an	n-k-áne	
1INCL	-		-	kuy-an	k-áne	
2	zen-bil-tza-n	zen-bil-tza-ten	w-an	muy-an	m-k-áne	
3	ze-bil-en	ze-bil-tza-n	d-an	suy-an	s-k-áne	s-an/n-an

Table 299: Two geometrically contiguous morphemes

As explained in Section 5.2.10, what must have been initially trial forms have been extended in Biak to cover greater numbers too. This extension, however, did not happen in the third person (maybe because of its higher frequency of use), which continues to make a trial vs plural distinction that has been lost elsewhere. Run-of-the-mill morphosyntactically-driven analogical processes (i.e. an incipient consolidation of the trial and plural numbers in Biak) have thus resulted synchronically in a morphomic pattern.

Similarly, in Basque, this time in a politeness-driven process similar to the one that caused the loss of *thou* in English, an originally 2PL form was extended to the 2SG, which has resulted synchronically in a morphosyntactically unmotivated pattern of formal identity between the 2SG and PL.

In these processes, language users deploy forms in contexts where they could not be used before. However, speakers are obviously sensitive to the meaning of word forms and to the morphosyntactic feature-value structure of paradigms. Because of this, the source and the target meaning are, in the majority of cases, close enough to share some value(s) and are therefore paradigmatically contiguous. Geometrically non-contiguous morphemes can, thus, usually only emerge analogically by way of an intermediate contiguous-morpheme stage:

	Pre-Pre-Twi		Pre-Twi		Twi	
	Affirmative	Negative	Affirmative	Negative	Affirmative	Negative
Past	*tò-ò	-	*tò-ò	*à-ḥ-tó	tò-ò	à-ḥ-tó
Perf.	*à-tó	*à-ḥ-tó	*à-tó	*à-ḥ-tó	à-tó	ḥ-tó-ò

Table 300: Hypothesized emergence of a non-contiguous morpheme in Twi

If the diachronic story I have proposed elsewhere (see 5.2.63) is correct, the Twi past tense descends from a completive aspect form which was initially (Pre-Pre-Twi) semantically incompatible with negation. When the semantic incompatibility disappeared, the perfective negative form would have been recruited (Pre-Twi). Only a posterior analogical paradigmatic extension would have resulted in the non-contiguous morphomic affinity we find in the language synchronically.

It is my contention, however, that, even in the case of morphemes emerging in a seemingly more accidental manner (e.g. from the morphologization of sound changes), geometrical contiguity may be more frequent an outcome than would be expected from chance alone. After all, the forms in the paradigms where the sound changes apply are far from being completely random. They are plagued with morphemes (i.e. formatives which recur in various cells with a shared value, e.g. 1PL, 2PL, 3PL), with so-called eidemic resonances (Bickel 1995),

and with the formal (i.e. Zipfian) correlates of usage-frequency differences. Consequently, even sound-change-derived morphemes/morphology cannot be completely random regarding their paradigmatic distribution, as word forms with similar content or frequency will have a higher chance of sharing forms as well.

Alongside source constraints, functional/cognitive factors may also favour geometrically contiguous morphemes. Natural classes (e.g. PL) may enjoy a learnability advantage over morphemic classes, and, in turn, contiguous morphemes (e.g. PL and/or 3) may be preferred (i.e. might be more learnable and diachronically resilient) compared to non-contiguous ones (e.g. 1PL or 3SG). This makes sense intuitively. Language users make their grammatical generalizations on the basis of both form and meaning. *Ceteris paribus*, (i.e. provided the same amount of formal evidence) ascribing grammatical relevance to the morphological identity of cells that are semantically contiguous (e.g. DU=PC=PL or 2PL=3PL=3SG) might be easier/more likely than doing so if the semantic adjacency does not hold (e.g. DU=PL≠PC or 2PL=3SG≠3PL≠2SG). This could reasonably make geometrically non-contiguous morphemes comparatively more vulnerable to change (into a contiguous pattern) or to disintegration or levelling when subsequent sound changes and/or analogical developments fiddle with the paradigmatic distribution of formatives. Some diachronic developments have been found in the present research which support this view:

	Ngkolmpu undergoer prefixes				Nen undergoer prefixes			
	SG	PL	SG	PL	SG	PL	SG	PL
1	u-	n-	b-	kn-	w-	y-n-	q-	t-n-
2	n-	y-	kn-	s-	n-	y-a-	kn-	t-a-
3	y-	y-	s-	s-	y-	y-a-	t-	t-a-

Table 301: Getting rid of a non-contiguous morpheme in Nen

In Ngkolmpu, 2SG and 1PL are always syncretic, a situation which is believed to be by-and-large inherited from the ancestral language. This diagonal morpheme, however, has been disrupted in Nen. Here, the contiguous morpheme that extended through 3+2PL has been extended to the 1PL as well, thus breaking the morphemic syncretism of 2SG and 1PL.

Something similar in its result (but quite different in its implementation) can be found in Nakh-Daghestanian. Consider the following evolution:

	Proto-Lezgian		Lak	
	SG	PL	SG	PL
I Male	*w	*b	∅	b
II Female	*r/j	*b	d	b
III Animate	*b	*d	b	b
IV Inanimate	*d	*d	d	d

Table 302: Getting rid of a non-contiguous morpheme in Lak

In the ancestral language (represented above by Proto-Lezgian even though Lak belongs to a different branch of Nakh-Daghestanian) the plural agreement morphology of human genders (I and II) was the same as the singular agreement morphology of the non-human animate gender (III). This morphological affinity is, thus, not paradigm-geometrically contiguous. Lak has seemingly remedied this by extending the inherited syncretism to the plural of gender III (to the “bridge” meaning, as it were), to achieve geometrical contiguity.¹¹⁵ The incorporation of the PL animate cell into the former pattern, thus, increases the coherence of the forms’ morphosyntactic distribution, which may make it a more “viable” meaning for a lexical entry. This may, thus, increase the chances of *b-* being conceived of by language users as a single morphological unit, i.e. a single prefix, as opposed to two homophonous prefixes.

In the present research on morphemes, thus, I have found evidence that simpler and geometrically contiguous morphemes are more frequent and therefore “preferred” in some sense. There is an ongoing debate in typology (e.g. Schmidtke-Bode et al. 2019) on whether the ultimate explanation of cross-linguistically recurrent patterns and regularities should come from mutational constraints or from functional ones. Although providing an answer to this (or even participating in the debate in any meaningful way) is beyond the purposes of the

¹¹⁵ Alternative analyses of this change are also possible, of course (e.g. a more semantically oriented extension of a human-denoting exponent to all animates). See, however, the change from Amele to Girawa (Tables 174 and 175) for a similar development that cannot easily be accounted for in the same way.

present dissertation, I consider functional/cognitive explanations superior here in that they can provide a motivation for the observed diachronic developments. Thus, if non-contiguous morphemes are less frequent, explaining this by saying that they either emerge less frequently and/or disappear more frequently is a truism. Ultimate explanation must come from establishing why this is the case. The desire of speakers to have formatives with more “sensible” or more clearly-defined meanings/distributions seems an obvious functional motivation in this respect.

5.5 Statistical analysis

Examining and discussing the data in each of the surveyed variables is interesting of and by itself, since it gives us information on the properties of morphemes cross-linguistically. The statistical analysis of their correlations promises to be another avenue for fruitful empirical discoveries that may shed light on some aspects of morphological architecture and/or linguistic cognition. Consider, therefore, the following statistical analysis in R, which shows the Pearson correlation coefficients¹¹⁶ between the variables presented in the previous section. Statistically significant ($p < .05$) correlations have been shaded, with highly significant correlations ($p < .01$) appearing in a darker shade:

¹¹⁶ This statistic is calculated as the covariance of two variables divided by the product of their standard deviations. It measures the linear correlation between two variables and takes numbers between -1 and +1, with numbers close to zero suggesting no correlation, and numbers close to -1 or +1 a strong correlation.

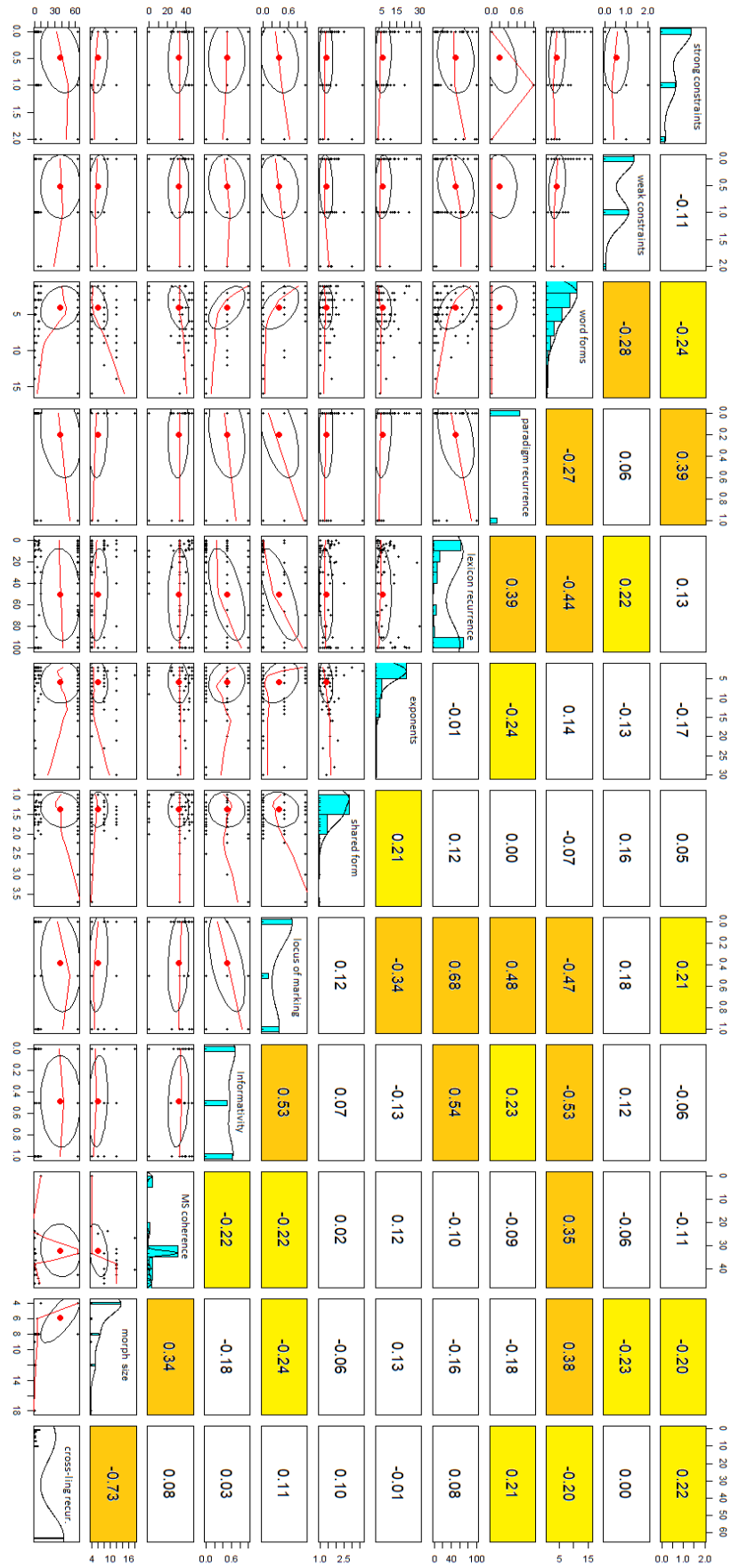


Figure 21: Pearson correlation coefficients between the variables in 5.4

-0.24 & -0.28 MS constrainedness and number of word forms: There is a statistically significant inverse correlation between the number of constraints that a morpheme is subject to and the number of different word forms that the morpheme spans. This seems to make intuitive sense since, the less constrained (i.e. the more widespread) that a formative is within the paradigm, the greater the chances that it accidentally cross-cuts or constitutes a superset of the distribution of another one. Consider the following made-up examples:

	Paradigm of *gal, System A				Paradigm of *gal, System B			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	bul-nu	gal-thi	gal-pi	gal-ho	bul-nu	bul-thi	bul-pi	gal-ho
2	bul-sa	bul-ro	gal-ju	gal-tu	bul-sa	bul-ro	bul-ju	bul-tu
3	gal-fe	gal-ky	gal-fi	gal-ki	gal-fe	bul-ky	bul-fi	bul-ki

Table 303: Biggest doubly-restricted (left) and unrestricted (right) morphemes

System A is the biggest possible morpheme subject to a strong (Present) and a weak (Non-3) constraint. With every cell characterized by a cumulative non-syncretic affix, it may contain maximally 3 different word forms. In the morphosyntactically unconstrained System B, by contrast, the maximum number of word forms is much higher. Interestingly, however, this inverse correlation between MS constrainedness and the number of word forms inside a morpheme holds only for stem-based morphemes. When one looks at affix-based morphemes separately, the Pearson coefficients are 0.07 and -0.06 for strong and weak constraints respectively, none of which come even close to statistical significance. This might point to a fundamental difference between the nature of stems and affixes in morphological architecture as the latter will usually express the finest-grained morphological distinctions in a language. Note, in this same line, the theoretical morphological operation known as blocking (see Section 2.2), according to which an affix in a subset relation to another one would prevent its appearance in the first place.

Promising as this line of thought seems to be, I believe that the failure of this correlation to show up in affixes (and for that matter the apparent success of alleged mechanisms like

blocking in accounting for empirical data) is to a large extent a byproduct of the fact that this property of “unsubsetability” is one of the criteria morphologists use in the identification of affixes and stems in the first place. Consider, to this end, the distribution of the formative /g/ in the following paradigm:

	Indicative		Subjunctive	
	SG	PL	SG	PL
1	pon-g-o	pon-emos	pon-g-a	pon-g-amos
2	pon-es	pon-éis	pon-g-as	pon-g-áis
3	pon-e	pon-en	pon-g-a	pon-g-an

Table 304: Present tense of Spanish *poner* ‘put’

The formative /g/ is, distributionally, perfectly segmentable both from the preceding segment /n/ (which occurs everywhere in the paradigm of ‘come’ and should thus probably be regarded as belonging unmistakably to the stem) and from the following segment (which can be either /o/ or /a/). On purely transitional grounds, thus, it should probably be regarded as an affix of its own right. In fact, it appears to occupy the same morphological slot as the formative /dr/ in future and conditional forms like *pon-dr-emos*, *pon-dr-án* etc. However, it has traditionally been regarded as part of the stem. Thus, morphologists will usually say that the Spanish verb ‘put’ has two different stems: *pon-* and *pong-*. The assignment of this segment to the stem is due to a number of reasons, very important among these, I believe, due to the fact that it occurs in a superset of the contexts of various other finer-grained suffixes.

This reminds us of the need to come to unified and well grounded definitions of the notions in our terminological toolkit as linguists. A well thought-out definition for a linguistic concept should, in my opinion, be concise and make reference to as few distinct variables as possible and if possible to just one. The delimitation/definition of stem and affix are particularly troublesome in this respect because it frequently (depending to some extent on the individual linguist) makes reference to i) combinatorial (i.e. transitional probabilities between segments), ii) morphosyntactic distributional (i.e. natural vs unnatural), iii) set-theoretical

relational factors (i.e. subset-superset relations), and even to iv) lexical generality (i.e. the number or proportion of lexical items that show the form).¹¹⁷ This is, obviously, very unfortunate because this definitional intertwinedness of logically different variables prevents us from analyzing their correlations in any meaningful way.

0.39 Strong constraints and paradigmatic recurrence: There is a statistically significant correlation between a morphome’s abidance to strong constraints and its recurrence within the lexeme. The explanation for this correlation is probably quite straightforward. Although a morphome can be subject to an indeterminate number of constraints and still not recur in the same paradigm, it must in principle at the very least be subject to one strong morphosyntactic constraint for it to be able to recur in another context. Consider the paradigm of the English verb ‘be’ below:

	English ‘be’				Pseudo-English ‘be’			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	am	are	was	were	*were	are	*are	were
2	are	are	were	were	are	are	were	were
3	is	are	was	were	is	are	was	were

Table 305: Paradigm of the English verb ‘be’ and a Pseudo-English variation

A morphome (e.g. PL+2SG) can hardly be found more than once in the same paradigm unless it is repeated with different formatives under two different feature values (e.g. in English, in present and past). There is, thus, a one-directional implicature here (i.e. recurrence within single paradigm implies strong MS constraint) which causes this correlation. It would be fair, I believe, to argue that, in a case like the hypothetical Pseudo-English above, we would not be dealing with two instantiations of the same morphome but rather with two different morphemes which are simply interlocked in the paradigm à la Daasanach (see Section 5.2.15).

¹¹⁷ It might be edifying to tweak each of these factors mentally in the above paradigm to realize how all of these variables can make /g/ seem intuitively more “stemy” or “affixy”.

-0.27 Number of word forms and recurrence in the paradigm: There is a statistically significant inverse correlation between the number of distinct word forms that a morpheme spans and whether or not it recurs within a single lexeme's paradigm. The reason for this correlation is, I believe, the same as for the one between MS constrainedness and number of word forms (the first correlation analyzed here). Smaller morphemes (i.e. those spanning a smaller number of cells) are more likely to be restricted to particular values (see Table 303) and only these can possibly occur more than once in a single paradigm.

0.22 Weak MS constraints and recurrence in the lexicon: The correlation between more MS constrainedness and a greater spread in the lexicon holds for both weak and strong MS constraints. It is statistically significant for weak constraints, although it does not reach significance for strong ones. This correlation appears to be due to the greater average constrainedness and spread in the lexicon of affix-based (0.69, 91.57%) compared to stem-based morphemes (0.44, 27.79%). It makes sense, intuitively, for a tradeoff of some kind to exist between a formative's spread in the paradigm (what is measured, really, by MS constraints) and its spread in the lexicon. For it to be successfully acquired, a formative must necessarily be present robustly, in naturally-occurring speech input. Greater intra-paradigmatic extension and greater cross-lexemic recurrence of a formative will both correlate to a greater frequency of occurrence. The graph below presents these correlations, with less frequent and more difficult to acquire formatives in red and easier ones in green:

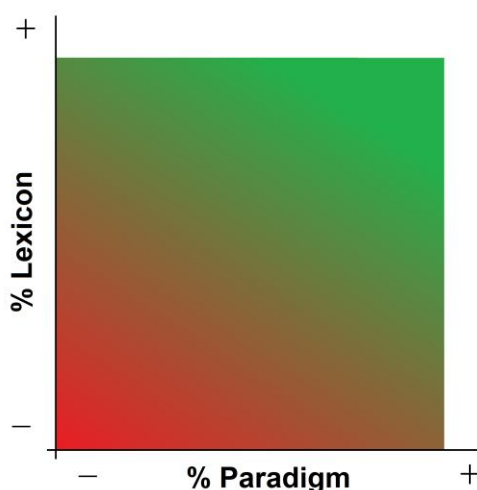


Figure 22: Some properties of a formative and their relation to frequency

It does not seem implausible, then, to suggest that those formatives that apply to a small proportion of lexical items and in a small portion of the paradigm will be more prone to dropping out of a language (e.g. by analogical replacement by more productive forms).¹¹⁸ This would result in a positive correlation between the two variables in this section.

-0.44 Number of word forms and recurrence in the lexicon: There is a statistically highly significant inverse correlation between the number of different word forms within a morpheme and its recurrence in the lexicon. Thus, a larger number of word forms usually means that a morpheme occurs in a smaller proportion of lexical items. The reasons for this correlation must be, I believe, similar to the ones mentioned for the former variable pair.

0.54 Informativity and recurrence in the lexicon: There is a highly significant statistical correlation between how grammatically informative a morpheme is and the proportion of lexemes where it appears. Informative morphemes (i.e. those which are not redundant but instead allow to discriminate different word-forms) tend to be more recurrent in the lexicon.

This and some of the previous correlations seem to show a more general connection between functionality and generality. Thus, morphology which is more “useful” (useful, that is, for the transmission of grammatical information), tends to be more widespread in the lexicon. Thus, greater MS restrictions, occurrence in a smaller number of word forms and participation in a language’s overall system of morphological contrasts are all variables which render a formative more valuable in the transfer of grammatical information. This may partially explain the greater success (i.e. greater lexical recurrence) of formatives with these properties. Consider, in the context of this discussion, the formative *-ta* in the following made up system.

¹¹⁸ Thus, only an unusually high token frequency makes it possible for forms like *am* or *is* to survive in English. This could well have been represented as a crucial third axis in Figure 22 above. Token frequency (of different cells and lexical items) would have been a most relevant variable to look at in relation to morphemes in this dissertation. The only reason why it has not been included here is the impossibility to find representative corpora of most of the languages that I will be concerned with. Based on our experience with other languages, educated guesses and approximations about token frequency can still be made (e.g. 3 is more frequent than 2, SG is more frequent than PL, realis present is more frequent than irrealis past, a lexeme meaning ‘give’ or ‘come’ will tend to be more frequent than one meaning ‘strangle’ or ‘recommend’ etc.) and will be made throughout this section because it is, I believe, one of the main sources for explanation in language.

	Paradigm of *gal				Paradigm of *bul			
	Present		Past		Present		Past	
	SG	PL	SG	PL	SG	PL	SG	PL
1	gal-ta-nu	gal-ta-thi	gal-ta-pi	gal-ta-h	bul-ta-nu	bul-ta-thi	bul-ta-pi	bul-ta-h
2	gal-ta-sa	gal-ta-ro	gal-ta-ju	gal-ta-tu	bul-ta-sa	bul-ta-ro	bul-ta-ju	bul-ta-tu
3	gal-ta-fe	gal-ta-ky	gal-ta-fi	gal-ta-ki	bul-ta-fe	bul-ta-ky	bul-ta-fi	bul-ta-ki

Table 306: An impossible formative

It would be impossible (outside word games) for formatives to appear in every possible paradigm cell of every lexeme. In the paradigms above, the formative *-ta-* fails to provide either lexical (because all lexemes have it) or grammatical information (because every single paradigm cell shows it). It would, thus, be entirely redundant and would constitute an unnecessary burden to the language's communicative efficiency (see Coupé et al. 2019). Because of this, such a formative seems impossible. The closer a formative gets to this, the less useful it is for information transfer and the more unlikely it becomes, which may partially explain some of these correlations: a broader grammatical domain (i.e. less grammatical informativity) will be associated with a narrower lexical domain (i.e. more lexical informativity) and *vice versa* (consider in this same line of thought the Uniform Information Density principle proposed by Jaeger [2010]).

0.39 Recurrence in lexicon and in paradigm: There is a statistically significant positive correlation between a morpheme's ability to recur within the paradigm of a single lexical item and its generality in the lexicon.

-0.24 Number of different exponents and recurrence within the paradigm: There is a statistically significant inverse correlation by which those morphemes characterized by greater formal diversity (most frequently stems) are less likely to recur within the paradigm.

0.21 Number of different exponents and average number of segments: There is a significant and positive correlation between the number of different formal exponents of a morpheme and their average segmental length. This is a finding which is highly relevant in that it might

not have been necessarily anticipated. It is not obvious why this correlation should exist but I will advance here one possibility: that some of the morphemes which are less formally diverse (e.g. repeated with only 2 or 3 different allomorphs) might sometimes be “spurious” in that some might be, in fact, accidental homophonies and not necessarily synchronically relevant grammatical categories. Shorter exponents (e.g. a single vowel) are more likely to be formally identical by accident. The requirement for the identity to be repeated with the exact same paradigmatic extent with a different formative (see Section 5.2.1) is intended to make it much less likely for spurious morphological identities to make it into the database. However, this risk is most probably still not zero and may only become progressively reduced the greater the number of different allomorphs.

Another, related way of explaining this correlation would be to say that those morphemes that have a higher number of formal realizations are more likely to be learned as grammatical abstractions truly independently of their concrete formal exponents (much in the way as morphemes are usually formalized with phonologically “blind” syncretic indexes named with abstract labels like e.g. λ , see Section 3.3). It may be, therefore, that only these more robust and formally diverse morphemes are really independent of their phonological instantiations and thus productive in the strictest sense of the word, that is, in a way that they can give rise to new (e.g. longer and suppletive) formal alternations.

0.21 & 0.18 MS constraints and locus of marking: There is a significant correlation by which affix-borne morphemes tend to be more morphosyntactically constrained. This matches the role traditionally attributed to affixes and stems (see also the discussion around Table 304 and about the Uniform Information Density principle in language). Note, also, however, the problems identified there concerning the incorporation of MS constrainedness as a factor relevant to morphological segmentation.

-0.47 Number of word forms and locus of marking: there is a highly significant inverse correlation between the number of different word forms that a morpheme spans and whether it has affix or stem-based exponents. The greater the number of different word forms, the greater the chances of it consisting of stem alternations. The basic explanation for this is, I believe, not dissimilar from the explanation offered for the previous correlation.

0.48 Recurrence in the paradigm and locus of marking: there is a highly significant correlation whereby it is usually affix-based morphemes that recur in a single lexeme's paradigm. This makes sense in that affixes tend to be more constricted than stems in their morphosyntactic distribution and in that a certain level of constrainedness is required for a morpheme to recur at different *loci* in the paradigm (see Table 305).

0.68 Recurrence in the lexicon and locus of marking: There is a highly statistically significant correlation between a morpheme's recurrence in the lexicon and the locus of its morphological expression in the word. Thus, morphemes based on stem alternations occur in a much smaller proportion of lexemes (27.8% on average, opposed to 91.6% for affix-based morphemes). See discussion around Table 306 for explanation.

-0.34 Number of exponents and locus of marking: There is a highly significant correlation by which morphemes which consist of stem-alternations tend to have a greater number of formal realizations compared to affixal morphemes (the averages are 6.9 vs 2.5). Again, it must be stressed that, of course, lexical recurrence is part of the traditional definition of what is a stem and what is an affix.

-0.53 Number of different word forms and informativity: there is a highly significant inverse correlation between these two variables by which a morpheme that spans a greater number of different word forms tends to be associated with lower levels of informativity. This is so in both stems and affixes and is, I believe, one of the main candidates for an empirical difference between morphemes and morphemes. Consider the paradigms below:

	Georgian morphemic system		Pseudo-Georgian morphemic system	
	SG	PL	SG	PL
NOM	buz-i	buz- eb -i	**buz-i	**buz- eb -ma
ERG	buz-ma	buz- eb -ma	**buz-ma	**buz-s
DAT	buz-s	buz- eb -s	**buz- eb -s	**buz-ad
ADV	buz-ad	buz- eb -ad	**buz- eb -ad	**buz- eb -i

Table 307: Partial paradigm of Georgian 'fly' (Aronson 1991:228-232)

It is very common (one could even say this is the default) for morphemes to be perfectly orthogonal to other formatives. Thus, in Georgian declension, for example, every single occurrence of the suffix *-eb* is informative because it consistently distinguishes singular from plural (by applying always in the plural). Orthogonality might be argued to be a desirable morphological trait, since it maximizes the number of word-form contrasts for a given number of formatives. In the Georgian partial paradigm above, for example, 5 suffixes are deployed to produce 8 different word forms.

One can easily think, however, of alternative systems where a formative with an unnatural morphosyntactic distribution could also be fully informative. Imagine, for example, if *-eb* appeared in the singular in some of the cases (e.g. in NOM and ERG) and in the plural in the others (i.e. DAT and ADV). Furthermore, all the formatives in an inflectional paradigm could potentially have an unnatural distribution in the paradigm (see Pseudo-Georgian in Table 307) without renouncing to the formal economy that comes from orthogonality.

The data gathered here, however, seems to suggest a tendency for morphemes to NOT be orthogonal to other morphological elements in the language. Although exceptions can of course be found (see Tol in Section 5.2.61, and Yagaria in Section 5.2.69), systems like the Pseudo-Georgian one above are very rare. The interpretation of this finding is up for debate but I would like to provide the following hypothesis:

The paradigmatic distribution of morphemic elements is straightforward. Thus, the element *-eb* in Georgian appears in all plural cells and only in plural cells. It is, thus not an exceedingly difficult task for Georgian language-users to correctly triangulate this formative's paradigmatic distribution even on the basis of limited input. This would be a much more difficult task in the case of Pseudo-Georgian since there is no reliable cue, formal or semantic, for when the formative must appear exactly.

A lack of orthogonality (e.g. a superset or identical-set relation) to other morphological elements in the same paradigm could be considered a way, alternative to semantics, to predict the appearance of a formative and could thus provide a coherent niche, as it were, for

its continued existence in the language. If, for example, the suffix *-eb* in Pseudo-Georgian always occurred before the suffixes *-i* and *-ad* and nowhere else, then one would be able to predict its appearance from other forms in the paradigm, thus increasing the learnability of its distribution even if this would render the suffix uninformative (i.e. redundant) as far as the morphological contrasts in the language are concerned. Thus, because morphemes, by definition, cannot rely on semantic/morphosyntactic cues for their distribution, they can only reduce or keep in check their distributional complexity with recourse to formal cues. Orthogonality to other form(ative)s prevents this and may thus be dispreferred in morphemes, but not in morphemes, which can make use of feature values instead.

0.23 Recurrence in the paradigm and informativity: there is a statistically significant correlation between a morpheme's informativity and its capacity to recur within a single lexeme's paradigm. Affixal morphemes are associated with high morphosyntactic informativity and high chances of recurrence within a paradigm whereas stem-based morphemes are characterized by the opposite properties (i.e. by low morphosyntactic informativity and small chances of recurrence within a paradigm).

0.54 Recurrence in the lexicon and informativity: There is a highly statistically significant correlation by which more informative morphemes (i.e. those that play a role in the distinction of different word-forms) are more widespread in the lexicon. This tendency is found in both stem and affix-based morphemes.

The last two correlations both imply the greater use in the language of morphemes that are more informative, and thus more "useful". This can be understood in two slightly different ways: informative morphemes tend to spread or are at least preserved because of their usefulness, or, much in line with the digression in the previous page, informative morphemes, because they lack reliable semantic and formal cues, must be more learnable/robust in other ways (e.g. with a greater lexical or paradigmatic frequency of occurrence) to survive in language.

0.53 Informativity and locus of marking: there is a statistically highly significant correlation whereby affix-based morphemes tend to be more informative than stem-based ones. Possible explanations for this correlation have already been presented before.

0.35 Number of distinct word forms and MS coherence: There is a highly statistically significant correlation between the number of different word-forms that a morpheme spans and that morpheme's MS coherence. The greater the number of word-forms, thus, the closer the morpheme's morphosyntactic distribution tends to be to naturalness. This seems to be an understandable correlation intuitively since, if a morpheme spans many different word forms, a rationale of some sort seems to be most useful as a simplifying force:

	PRES.IND	PRES.SUB	IPF	PAST	IPF.SUB I	IPF.SUB II	FUT	COND
1SG	nievo	nieve	nevaba	nevé	nevara	nevase	nevaré	nevaría
2SG	nievas	nieves	nevabas	nevaste	nevaras	nevases	nevarás	nevarías
3SG	nieva	nieve	nevaba	nevó	nevara	nevase	nevará	nevaría
1PL	nevamos	nevemos	nevábamos	nevamos	neváramos	nevásemos	nevaremos	nevaríamos
2PL	neváis	nevéis	nevábais	nevasteis	nevárais	neváseis	nevaréis	nevaríais
3PL	nievan	nieven	nevaban	nevaron	nevaran	nevasen	nevarán	nevarían

Table 308: Distribution of diphthongization in Spanish N-morpheme verbs

	PRES.IND	PRES.SUB	IPF	PAST	IPF.SUB I	IPF.SUB II	FUT	COND
1SG	nevo	nieve	nevaba	nevé	nevara	nevase	nevaré	nevaría
2SG	nievas	neves	nevabas	nevaste	nievaras	nevases	nevarás	nevarías
3SG	neva	nieve	nevaba	nevó	nevara	nevase	nievará	nevaría
1PL	nevamos	nevemos	nievábamos	nevamos	neváramos	nevásemos	nevaremos	nevaríamos
2PL	neváis	nevéis	nevábais	nevasteis	nevárais	neváseis	nevaréis	nevaríais
3PL	nievan	neven	nevaban	nievaron	nevaran	nevasen	nevarán	nevarían

Table 309: Distribution of diphthongization in Pseudo-Spanish

Even in morphemes, this rationale can be partially morphosyntactic. As evidenced in the Romance paradigm in Table 308, the N-morpheme is very far indeed from being morphosyntactically random. The 7 different word forms it spans all share the value present and, in addition, are expressible quite succinctly as the SG and/or 3 cells within that domain. This must surely aid with the functionality and learnability of the pattern. A comparison to the more unstructured hypothetical morpheme in Table 309 reminds us just how rare true morphosyntactic incoherency really is even among morphemes. The reason for this must be

related to both diachronic and learnability constraints. The need for more structure, obviously, becomes greater the larger the number of different word forms or cells involved.

-0.22 Locus of marking and MS coherence: There is a significant inverse correlation between the affix- or stem-based nature of a morphomic pattern and its morphosyntactic coherence, as stem-based morphemes are characterized by a higher MS coherence on average (33.66%) compared to affix-based morphemes (28.76%). This correlation appears to be quite perplexing at first sight. However, it is less unexpected if one keeps in mind that stem-morphemes tend to be “bigger” in general (i.e. tend to spread over a larger number of cells and word-forms, a factor which is associated with a greater need for morphosyntactic structure (consider the previous correlation).

-0.22 Informativity and MS coherence: There is a significant inverse correlation between a morphome’s capacity to participate actively in the overall system of morphological distinctions and its MS coherence. Thus, more informative morphemes tend to be less morphosyntactically coherent. Informative morphemes tend to be affixes (see previous correlation) and span over a smaller number of different word forms (see the discussion around Tables 308 and 309).

-0.20 -0.23 Constrainedness and morphome complexity: there is a statistically significant inverse correlation between a morphome’s confinement to paradigmatic subdomains and the total number of cells of the paradigm required to describe the morphome’s MS distribution. Thus, more constraints correlate, unsurprisingly, to “smaller” morphemes.

0.38 Number of word-forms and morphome complexity: There is a highly significant correlation between a morphome’s span over a number of different word forms and the number of content-cells required to describe its distribution. More word forms is, logically, associated with more paradigm cells.

-0.24 Locus of marking and morphome complexity: Stem-based morphemes are correlated in a statistically significant way to greater distributional complexity (i.e. to a greater number of paradigm cells required). This correlation is related to others that have been presented.

0.34 MS coherence and morpheme complexity: There is a highly significant correlation between a morpheme's MS coherence and its complexity as per number of features and values required to capture its distribution. More incoherent morphomic distributions tend to be simpler while more coherent morphemes can afford to be more complex in this respect.

Because both variables measure dimensions associated with morphological complexity, it might be reasonable to expect a tradeoff, as it were, between complexity in one domain and complexity in the other. Thus, one could argue that morphemes whose distribution is more easily describable (e.g. in a two-member list rather than a longer one) are easier to acquire and/or may emerge more frequently in the course of language evolution.

Two member-lists can translate into different things: i) a high MS coherence with high morpheme complexity, when the values that make up the list differ in their specificity (e.g. the Spanish L-morpheme: SBJV+1SG.IND, 46.6%, 8), and ii) a low MS coherence with low morpheme complexity when the values of the list are equally specific (e.g. the Irish diagonal syncretism of GEN.SG+NOM.PL, 0%, 4). Morphemes can only rank relatively high on both axes of complexity (e.g. Pite Saami4: 1DU.PRES+3PL+1/2SG.PAST, 26.7%, 18) when they require longer descriptions. Because of their greater complexity, these morphemes, might, understandably, be more infrequent.

0.22 Strong MS constraints and cross-linguistic recurrence: There is a statistically significant correlation by which more constrained morphomic patterns tend to be associated with a greater cross-linguistic recurrence. As explained in Section 5.4.12, a high level of paradigmatic complexity in general (MS constraints can be thought of as limits to complexity) is at odds with cross-linguistic recurrence. In a similar way that winning 4 consecutive bets in the roulette is less likely than winning only 3, a geometrical paradigmatic structure becomes less likely to have emerged several times in the life-size roulette of language evolution if it is complex and composed of more parts or independent events. This broad explanation can be offered to the next two correlations as well.

-0.20 Number of different word forms and cross-linguistic recurrence: There is a statistically significant inverse correlation by which morphemes that span a smaller number of distinct word forms tend to be more cross-linguistically recurrent.

0.21 Recurrence within a paradigm and recurrence across languages: There is a significant correlation by which morphemes that recur within the paradigm of a single lexeme tend to be also more recurrent cross-linguistically.

-0.73 Morpheme complexity and cross-linguistic recurrence: There is a highly significant inverse correlation by which morphemes which require reference to more features and values in the description of their paradigmatic distribution tend to be much less frequent cross-linguistically. See, again, Section 5.4.12 for an explanation of why this might be so.

5.6 Conclusion

The present section has presented a cross-linguistic database (the first of its kind) containing 110 morphemes from 74 genetically and geographically diverse languages. Each of these morphemes is described for a dozen different quantitative and qualitative variables which measure different aspects of each morpheme's form and distribution.

This section, thus, has contributed to the fields of morphology and typology. Regarding the latter, typological approximations to the phenomenon of morphomicity were not only lacking, but sometimes not even considered possible before. Regarding the former, this study constitutes the first and only lengthy piece of research that deals with morphomic structures beyond the Romance language family, which has nearly monopolized the literature on morphemes to date.

The findings of this study are many and varied. First of all, concerning the cross-linguistic prevalence of the phenomenon, it has been found that morphemes (as defined for this synchronic study) are present in around 15% of grammatical descriptions.¹¹⁹ This makes them relatively infrequent morphological phenomena. Although their variety, in terms of features

¹¹⁹ This is necessarily dependent on the quality and quantity of the available descriptions. Although only full, high quality grammars (i.e. not grammar sketches) have been considered for the purposes of this 15%, the proportion of languages that have these structures is likely to be, of course, higher than this.

and values involved, is outstanding, the present study has found that some morphomic structures are not typologically unique, but are rather found in several unrelated languages. These involve the following unnatural morphological allegiances of person-number values: SG+3PL, 3+1SG, 2+1PL, PL+1SG, PL+2SG, PL+3SG. Looking at geometrical patterns in the abstract, it has been found that simple and geometrically contiguous patterns (e.g. all of the above) are the most frequent. Some possible explanations for these findings have been presented throughout this section and through the diachronically oriented Chapter 4.

Regarding the cross-linguistic properties of morphemes, the present study has revealed variation across various properties: morphemes can be stems or affixes, they can be completely unconstrained or can be subject instead to multiple morphosyntactic restrictions, they can be fully informative or completely redundant, they can appear in a single lexical item or in every single one of them, and they can have distributions which range between complete morphosyntactic incoherence and near-coherence. Most of the quantitative variables analyzed (concretely: number of word forms, number of exponents, shared form, and paradigm size) follow Zipf-like distributions.

Statistical analysis shows that, although they are (mostly) logically independent, many of the analyzed variables appear to be significantly correlated. The causes for most of these correlations must be many and complex and each of them could well be the topic of a whole other dissertation. Some of the most interesting correlations found here involve the greater lexical extension of more morphosyntactically restricted morphemes (which hints at a tradeoff, as it were, between lexical and grammatical informativity, see Table 306), the tendency of morphemes to not be orthogonal to other formatives (see Table 307), and the increased redundancy and structuredness of morphemes that spread across more cells or word forms (which seems to reflect complexity limitations by which morphemes cannot be both “big” and “messy”, see Tables 308 and 309).

The present findings notwithstanding, there remains much work to be done concerning morphemes and the variables and correlations analyzed here. Most urgent, in my opinion, would be to compare these to the properties of morphemic (i.e. morphosyntactically naturally distributed) elements. A database of morphemes comparable to this one and structured along

the same variables and definitions could be used to put the present findings in a broader perspective and help find an answer to whether or to what extent morphemes and morphemes are different objects empirically. Although, in the absence of this kind of broad quantitative programme this is largely speculative, research so far suggests that differences between them could be hard to find (see Herce 2020). Although it is likely that morphemes will tend to be less informative than morphemes (see Table 307), and will tend to have different diachronic origins (e.g. sound change, see Chapter 4), my contention is that empirical differences between morphemes and morphemes might be few overall.

6 Conclusions

This final section will round up the thesis by reflecting on the importance of morphosyntactic features (Section 6.1) and of form-to-form predictive relations (Section 6.2) in the evolution of morphological patterns. Even if discussion has been understandably centred here on unnatural patterns and thus on other sources of external motivation, values and meaning also deserve to be pondered against the autonomously morphological templates that have been the topic of this dissertation. To conclude (Section 6.3), the major findings of this dissertation will be recapitulated and their implications discussed.

6.1 The importance of features

Morphological elements, whether stem alternants or affixes, whether morphemes or morphemes, owe their distribution either to their source construction or to analogical developments that subsequently modify the original distributions. Because morphology usually originates from free words in syntactic constructions, it is only to be expected that elements of form will correlate strongly to feature values or meanings, and pattern into natural classes (although see Section 4.1.5). The fact, for example, that the dental suffix *-te* in German conjugation appears in every paradigm cell of the past tense and nowhere outside of the past is probably a mere continuation of the state of affairs inherited from syntax. At some

stage before Proto-Germanic, some syntactic construction along the lines of 'ask did' must have been used to express the past. When the erstwhile free word became an affix (ask did > ask-d) it left the realm of syntax to enter that of morphology but it preserved its earlier distribution. Thus, even if the organizing principles of morphology and syntax differed substantially (and if, for example, morphology "didn't care" at all about features or values), a great deal of form-meaning correlation would be expected nonetheless in synchrony. If we believe morphology can be subject to rules of its own, we may need additional evidence to ascertain what it is that morphology cares about.

Morphosyntactic features and values are generally assumed to be an important factor to account for the distributions of morphological elements not only because of their significant synchronic correlation but also because they seem to play a big role in analogical change:

	Old Norse (Rask 1976:121)		Old Swedish (Noreen 1904:471-473)		Modern Swedish (Holmes & Hinchliffe 2003:264)	
	SG	PL	SG	PL	SG	PL
1	brenn	brennum	brenner	brennom	bränner	bränner
2	brennr	brennið	brenner	brennin	bränner	bränner
3	brennr	brenna	brenner	brenna	bränner	bränner

Table 310: Diachronic spread of the formative *-r* in Scandinavian, present tense

The fact that this dissertation has been focused on patterns at odds with morphosyntactic values cannot lead us to think that they are irrelevant in morphological architecture. The above present tense paradigms of 'burn' in several stages of Scandinavian show that in analogical extension, morphosyntactic feature values (e.g. SG.PRES in Old Swedish or PRES in Modern Swedish) often act as niches (Gause 1934, Aronoff 2016) where a single form may come to predominate. Morphosyntactic and semantic niches also constrain the expansion of formatives to other environments.

Feature values are usually assumed to be important because they are good predictors for morphological change. Thus, the paradigmatic extension of the suffix *-(e)r* from Old Norse to Old Swedish is "expected" over hypothetical extensions to other paradigm cells like, for

example, 3PL or 1PL. This is demonstrated by the fact that comparable developments can be easily found, in different morphological elements (e.g. in the stem), and in different morphosyntactic contexts (e.g. in the plural):

	<i>blōta</i> 'sacrifice' (Wurzel 1980:451-452)				<i>beran</i> 'bear' (Fertig 2016:434)			
	Pre-Old Norse		Old Norse		Pre-Old English		Old English	
	SG	PL	SG	PL	SG	PL	SG	PL
1	blōtu	blōtum	blót	blótum	bere	*berams	bere	berap
2	blótiR	blōteð	blótr	blóteþ	birest	berap	birest	berap
3	blótiR	blōta(n)	blótr	blóta	bireþ	berap	bireþ	berap

Table 311: Two similar morphological changes in Germanic

The shaded change in *blōta* illustrates how the stem vowel that arose in 2SG and 3SG regularly by i-Umlaut was generalized to the whole singular in Old Norse. Similarly, the change in *beran* shows how the syncretism of 2PL and 3PL that had resulted from regular sound change (3PL *-anþ>-ap) was extended to the remaining cell of the plural.

Feature values, therefore, constitute grammatical templates of the utmost importance. This means, in my opinion, that they should be allowed to feature prominently in morphological description, theory, and formalization. A particularly striking example of how particular feature values can act as niches in morphological change can be found in Yakkha (Kiranti):

	Data from 1984			Data from 2012		
	1SG.P	1DU.EX.P	1PL.EX.P	1SG.P	1DU.EX.P	1PL.EX.P
2SG.A	-ŋgana	-gaha	-gaha	-ŋgana	-gaha	-gaha
2DU.A	-ŋciŋaha	-ŋciŋaha	-gaha	-gaha	-gaha	-gaha
2PL.A	-ŋiŋana	-gaha	-gaha	-gaha	-gaha	-gaha
3SG.A	-ŋna	-ŋciŋaha	-ŋciŋaha	-ŋna	-ha	-ha
3DU.A	-ŋna	-ciha	-ha	-ha	-ha	-ha
3PL.A	N- -ŋna	-ciha	-ha	-ha	-ha	-ha

Table 312: Yakkha agreement suffixes, partial paradigm (Schakow 2016:223)

In a relatively brief period of time, quite dramatic changes have taken place in the agreement patterns of the language whereby many morphological distinctions and suffixes have disappeared. This process has not been random, of course. The resulting paradigm is one where, unlike in the earlier system, one-to-one form-meaning relations hold.

Diachronic changes like these suggest, thus, that morphosyntactic features and values are important in morphology. I can, therefore, not fully agree with Carstairs-McCarthy (2010:210) when he argues that morphological evolution suggests that the importance of features “has been overrated”. I do not fully agree with Maiden (2016:49) either when, based on the behaviour of stem alternation patterns in Romance, he argues that morphomic patterns are not dispreferred. This claim may fit the evidence from Romance morphemes, but may be argued to be incompatible with the paradigmatic changes that have been presented throughout this section. If morphomic patterns were not dispreferred to some extent, we would have no reason to predict that changes like the ones in Old Norse and Old English would be any more common than alternative paradigmatic extensions like these:

	Pre-Old Norse		Pseudo-Old Norse		Pre-Old English		Pseudo-Old English	
	SG	PL	SG	PL	SG	PL	SG	PL
1	blōtu	blōtum	blótu	blótum	berē	berams	**beraþ	berams
2	blótiR	blōteð	blótr	**blóteþ	birest	beraþ	birest	beraþ
3	blótiR	blōta	blótr	blóta	bireþ	beraþ	bireþ	beraþ

Table 313: Hypothetical alternative morphological changes

Probably all linguists would agree that changes of this type are not as common as morphosyntactically motivated ones. Because they play, by definition, on the same side as feature values, natural-class patterns always have this advantage over morphomic patterns and this source of external motivation that morphemes lack. There is also experimental evidence (e.g. Kirby et al. 2008, Silvey et al. 2015 etc.) that shows that, when formal distinctions are lost, the conflation of values is by no means random but rather follows largely the same tendency that was observed in Yakkha in Table 312.

Language is a system which is transmitted from one generation to the next on the evidence of only partial and incomplete data about the system itself. It only stands to reason that language users “circumvent this transmission problem by exploiting structure in the set of meanings to be conveyed.” (Kirby et al. 2008:10685). Although one might have wished ideally to evaluate preferredness on two patterns in the same language, one natural and one morphomic, matched for every single other property, the experimental findings reported above, as well as the diachronic tendencies discussed, are difficult to reconcile, in my opinion, with a theory of grammar where morphomic patterns are not dispreferred to some extent.¹²⁰

As mentioned by Maiden (2016:49), however, it is true that, in the context of Romance stem alternations, language-users usually do not seize the opportunity to align form to function (but see Section 4.2.4.1). It is interesting to find that consonant palatalization before front vowels produced stem alternations only in those conjugations where the resulting pattern was morphomic (e.g. *hacer*, *decir*). By contrast, the alternations failed to be preserved in the productive conjugation (e.g. *pagar*, *colgar*), precisely there where it would have resulted in a stem alternant isomorphic with a morphosyntactic natural class:

	<i>hacer</i> ‘do’		<i>pagar</i> ‘pay’ (expected forms)	
	IND	SUB	IND	SUB
1SG	hag-o	hag-a	pag-o	*pac-e
2SG	hac-es	hag-as	pag-as	*pac-es
3SG	hac-e	hag-a	pag-a	*pac-e
1PL	hac-emos	hag-amos	pag-amos	*pac-emos
2PL	hac-éis	hag-áis	pag-áis	*pac-éis
3PL	hac-en	hag-an	pag-an	*pac-en

Table 314: Expected paradigmatic results of velar palatalization in Romance

¹²⁰ I have presented before (Biak, Basque, Occitan, Slovene) a few cases of morphosyntactically motivated changes that gave rise to morphomic patterns. The preference for morphosyntactically motivated formal extensions could, thus, be argued to be a more localized constraint on **change** independent of the naturalness of the more general pattern that the change gives rise to as a result. A bias towards morphosyntactically motivated changes without a similar bias towards morphosyntactically-motivated patterns seems to me, however, unlikely (see also the insights on Section 5.4.12).

Velar consonant palatalization did not happen¹²¹ in first conjugation verbs like *pagar*, precisely the one, as I say, where a natural-class stem alternation would have been the result. At first sight this seems strong evidence for no-bias, or even for a bias against natural classes. However, there are many and important confounding factors. To begin with, the stem alternant *hag-* represents a greater proportion of the total use tokens of the verb compared to hypothetical **pac-*. Other asymmetries could have also favoured the pattern in *hacer*. For example, in the conjugations where it happened, the sound change affected the majority of the paradigm, whereas in the first conjugation it would only have affected a small portion of it. Other confounding factors could have been that i) maybe too few verbs ended in the “right” consonant in the *-ar* conjugation, or ii) maybe the great token frequency of a few *-ker/-kir* verbs like *decir* ‘say’ or *hacer* ‘do’, which may have played in favour of the stem alternation pattern that survived even “compensating” for its unnaturalness... With a single example (or with a few related examples in a single language family) there is simply no way to tell. This is the reason why a cross-linguistic approach to the morpheme was urgently needed.

Features and their values, cross-linguistic evidence suggests, are important in morphological structure. This, of course, does not mean that feature-value structure is the only operating force in morphology or even the most powerful one. The fact that *ceteris paribus* natural patterns are preferred over unnatural ones does not mean that other forces are irrelevant or cannot, under the right conditions, take the upper hand. Morphemes show clearly, indeed, that “the impulse toward greater isomorphism is not an irresistible one” (Stump 2015:268). It has been my goal in this dissertation to advance our understanding of precisely which conditions and forces are operating when unnatural morphosyntactic patterns do manage to get established and successfully replicated in a language.

¹²¹ Forms contain the velar in the modern languages e.g. Spanish *pa[g]e*. The sound change, thus, was either turned back analogically or resisted *ab initio* in these first conjugation verbs.

6.2 The importance of form

Morphology (i.e. the internal structure of words and paradigms) is, therefore, certainly about meaning, features, and values. It seems a lost cause to try to argue against it in all cases. In “well-behaved” agglutinative paradigms like e.g. the Turkish case-number inflection (Kornfilt 2013), for example, there is no reason not to say that particular formatives are there to convey semantic information like ‘plural’. Diachrony shows us that semantic values (e.g. PL, PAST) can become associated with particular morphological forms even when the ancestral language lacked any such exponents. This happens in run-of-the-mill grammaticalization processes where a formerly independent word (e.g. a pronoun) may accrete to another word (e.g. a verb) and simply preserve its original meaning. Morphology-internal processes also bear witness to the architectural importance of natural-class distinctions and morphosyntactic and semantic values. Consider the discussion in Section 6.1 and the emergence, in many Romance or Germanic languages, of plural markers (e.g. *-s -i -er -en*”) from former états de langue (e.g. Latin or OHG) that had no number-dedicated morphology whatsoever. Morphology is, thus, often about conveying meaning. This is hardly new or surprising considering the functions and communicative needs that language has to serve.

As this dissertation has shown, however, morphology is also about something else. It is about trying to preserve the inherited system as faithfully as possible even when this is communicatively superfluous. Developments of many kinds (e.g. sound change, grammaticalizations, the loss of morphosyntactic distinctions, semantic drift etc.) can result in morphological affinities that do not match natural classes. These structures can be acquired and can provide a model in processes of analogical change. This is because morphology is also about being able to produce forms one may never have heard before. This means that, along with shared meanings, formal predictabilities within and across words are registered and used by speakers to cover the gaps that a Zipfian input does not fill. This leads to formally-driven analogies that either perpetuate or reinforce the paradigmatic results of former historical accidents, or even create new categories (see sections on formally motivated analogy [4.1.3.2] and pattern interactions [4.1.4]) based on more-or-less accidental formal affinities.

Similarity and covariation in formal exponence, therefore, attracts more similarity. This could hardly be otherwise. When predicting and producing forms online on the basis of an imperfect input, language users may sometimes overgeneralize and change/regularize the grammatical system handed down to them. In this way, formal implicational patterns tend to reinforce themselves at both the paradigmatic and the syntagmatic level:

Old Norse			Spanish 'put'			
Lexeme	Infinitive	3PL.PRES.IND	Cell	Stem	Suffix1	Suffix2
'drive'	<i>fara</i> ↔ <i>fara</i>		3PL.PRET	<i>pus</i> ← <i>ie</i>		<i>ron</i>
'must'	<i>skulu</i> ↔ <i>skulu</i>		1SG.IMP.SBJV	<i>pus</i> ← <i>ie</i>		<i>ra</i>
'owe'1	<i>eiga</i>	<i>eigu</i>	Gerund1	<i>pon</i>	<i>ie</i>	<i>ndo</i>
'owe'2	<i>eiga</i> ↔ <i>eiga</i>		Gerund2	<i>pus</i> ← <i>ie</i>		<i>ndo</i>

Table 315: Some predictability-driven morphological changes

The Old Norse verbal inflectional system (see Table 87 and 88) led one to expect formal identity between the infinitive and the 3PL present indicative forms. For the vast majority of verbs, thus, one could correctly predict infinitive *fara* from 3PL *fara* and vice versa. This vast generalization was perceived by language users, who then had the capacity to overgeneralize this rule (see the change from owe1 (Old Norse) to owe2 (Icelandic) whenever an exception was not successfully acquired from the input).

Implicative relations can be bidirectional and one-directional. They can also apply at the paradigmatic level (i.e. between different word forms) and at the syntagmatic level (i.e. between different parts of a single word). The Spanish verbal inflectional system leads one to expect that the stress-bearing suffix /je/ will co-occur with the PYTA root. This is so because it is the case in the vast majority of cells where the formative appears (in 13 out of 14 word forms). This syntagmatic implicative pattern is perceived by language users, who may then strengthen it further when they occasionally overextend it to the exceptional context where the rule did not use to apply before (consider the morphological change from Gerund1 to Gerund2 in some Spanish varieties).

Despite diachronic changes like these, it is the point of departure of most models of morphology that the main and sometimes only reason for the existence of a morphological module in language (whether autonomous or not) is the expression of meaning or morphosyntactic functions. Morphological structure, therefore, is most of the times interpreted and explained exclusively with reference to morphosyntactic features and their interaction. Formal identities that correlate well with morphosyntactic values are deemed to be significant while those which do not are either straightjacketed into better behaviour (e.g. by underspecification and blocking) or dismissed as 'accidental homophonies'. Yet there is abundant evidence that formal differences do not always correspond to differences in morphosyntactic values (e.g. inflection class distinctions, overabundance) and conversely, that differences in morphosyntactic values do not always align with formal differences (e.g. syncretism, deponency). These are examples of morphological structures that exist at odds with meaning and morphosyntactic values, which undermines to some extent the traditional way of understanding and modelling inflectional morphology only with reference to them.

Noticing identities (also partial identities and similarities) in both form and meaning and integrating those patterns into the fabric of grammar is the only cogent account of how speakers learn and use their language. Perceiving a formal similarity and knitting it into grammatical structure will be surely facilitated by the existence of some overarching meaning or morphosyntactic affinity as this provides 'extra evidence' for the importance of the formal pattern and for predicting its distribution (see Section 2.4). However, doing the same thing with morphosyntactically unrelated forms is likely to optimize cognitive resources too and allow language-users to solve the paradigm cell-filling problem (Ackermann et al. 2009):

	nacer 'be born'		hacer 'do'	
	IND	SBJV	IND	SBJV
1SG	nazco	nazca	hago	haga
2SG	naces	nazcas	haces	hagas
3SG	nace	nazca	hace	haga

	nevar 'snow'		sentar 'sit'	
	SG	PL	SG	PL
1	nievo	nevamos	siento	sentamos
2	nievas	neváis	sientes	sentáis
3	nieva	nievan	siente	sientan

Table 316: Partial paradigms of some Spanish verbs

For obvious reasons, the verb ‘be born’ is only seldom used in the present tense in persons other than 3. The 1SG.IND *nazco* appears in the 286 million-word corpus CORPESXXI only 12 times. The form, thus, must be produced online, not stored. However, when this happens, it resembles the 3 subjunctive (/naθk/-), rather than the 3 indicative (/naθ/-). This is so because the forms of those lexemes with comparable alternations whose paradigms are more “complete” in the input (e.g. *hacer*, *conducir*) create the expectation that this should be so.

Something similar happens with other verbs and stem alternations. In weather verbs like *nevar*, only the 3SG and nonfinite forms are regularly present in natural speech. These forms, however, are enough to establish that alternation (compare infinitive *nevar* vs 3SG *nieva*) is indeed present in the verb. On the basis of other verbs with comparable alternations, then, the whole paradigm can be filled out online if necessary, even when this results in forms that do not align well to morphosyntactic values.

It seems, therefore, that morphological entities and productive implicative patterns do not always need to have a morphosyntactically coherent description. Formal affinities alone can also prompt language users to construct grammatical categories like the morphemes in Table 316. As expressed by Hockett (1987:88), therefore, sometimes “it is the resonances that induce the grammatical structure.”

6.3 Summary and results

The present dissertation has been the first, as far as I am aware, to approach the concept of the morpheme from a typological and cross-linguistic perspective. Chapter 1 presented briefly the phenomenon and our knowledge of it, clarified the terminology, and presented the overall goals of this research.

To make the morpheme a workable comparative concept suited for the purposes of a typological investigation, Chapter 2 dealt with definitional issues: how to distinguish morphemes from accidental homophonies, how to identify an unnatural class, what is the role attributed to blocking or zeroes... as well as other issues that may come on the way when

deciding on the morphomicity of some morphological structure: segmentability, the intra- or extraparadigmatic domain of the pattern, its cross-linguistic recurrence etc. The overall *modus operandi* has been to set a high bar for unnaturalness and systematicity but at the same time to avoid reference to metaempirical factors (e.g. theoretical analyses and controversial units like blocking or zeroes) in the present definition of the phenomenon, thus remaining as close as possible to the data.

Having left the overall approach clear and having settled potentially controversial issues, Chapter 3 started to deal with more intriguing topics. It explored the connection between morphomicity and other morphological phenomena to find out that syncretism, deponency, heteroclisis etc. can all adopt morphomic distributions. It explored the gradient nature of the natural vs unnatural class distinction. It presented evidence that casts doubt on the possibility to define morphemes as independent from phonology let alone to have a clean cut between phonologically-derived and morphological patterns. It also explored the economic advantage of morphomic vs non-morphomic analyses of different patterns.

The diachronically-oriented Chapter 4 explored the different ways in which morphemes can arise, change, and disappear from a language. Sound change was found to be the most frequent source of morphemes (of the kind analyzed in this dissertation at least). Sound change, however, has been found to be an internally heterogeneous route to morphomhood, as the locus and result of sound changes can differ in nontrivial ways. The main finding of Chapter 4 is that not only sound change but also every other process that can result in a change to the forms in a paradigm (e.g. grammaticalization, analogy, pattern interactions, borrowing etc.) may become a source of morphemes under the right conditions.

The next section, Chapter 5 constituted the core of the dissertation. It presents a multivariate typological deconstruction of cross-morphomic variation. Morphemes in different languages have been found to vary along several different dimensions, among others their degree of unnaturalness, their number of exponents, their generality across the lexicon, the number of word-forms they span, how informative they are, etc. A synchronic database was introduced where 110 morphemes from languages all over the world have been painstakingly described, presented in their genetic and diachronic context when possible, and quantified for the

above-mentioned variables. An exploration of the data and variable correlations follows. Some of the most interesting findings are the recurrence of some unnatural patterns and the prevalence of geometrically contiguous morphemes in general. Frequency, and functional and mutational constraints have been proposed as explanations. Other interesting findings of this synchronic typological section are the greater lexical generality of more paradigmatically constrained morphemes (which points to a tradeoff between lexical and grammatical informativity), the greater structuredness (i.e. near-naturalness) of bigger morphemes (which points to some upper limit to how complex a morpheme can get), and the tendency for morphemes not to be orthogonal to other formatives within the paradigm (which argues for a form-based rationale of some sort to their distribution).

Elaborating on the findings of Section 5, it has found that, even when setting a high bar for morphomehood, **morphemes are present across the world's languages**. They have been found here in as many as 30 genetically entirely independent stocks both big (e.g. Austronesian, Indo-European, Otomanguean, Sino-Tibetan etc.) and small (e.g. isolates like Basque, Burmeso, Nivkh etc.). This suggests that the phenomenon cannot be dismissed lightly as an accidental quirk of a few languages and has to be explored in detail instead. It deserves, therefore, the systematic cross-linguistic treatment that has been missing so far.

Previous morphomic literature has highlighted **the importance of formal predictability relations within the paradigm**, which seem to constitute the synchronic *raison d'être* of morphemes, as well as the source of their purported diachronic resilience. This has received additional confirmation in this dissertation (see e.g. Section 6.2). Speakers notice and use these predictability relations because they need to produce unknown forms. They need to solve the paradigm cell-filling problem and overcome the Zipfian nature of linguistic input to induce a complete system. Because of this, as previous literature has found (e.g. Maiden 2018), pre-existing forms can serve as templates for the distribution of new formatives. This dissertation has provided many clear examples (beyond the Romance ones most often discussed, see e.g. the sections on Luxembourgish [Tables 118 and 119], Yakkha [Tables 122 and 123], and Svan [Table 231]) of the power of forms to act as niches or templates for other forms. Morphemes and unnatural implicative patterns, therefore, can constitute productive grammatical categories and steer morphological change.

That predictability must lie at the core of morphemes is thus clear. There is, however, a fundamental fact that morphomic literature has not engaged with so far, which is that predictability relations exist also outside morphemes and in the absence of formal. As shown by Herce (forthcoming-b), for example, the +g stem-augment in the L-morpheme cells and the +dr stem-augment in the future and conditional tenses always appear together in Spanish (cf. *venir, tener, poner, salir, valer*). The presence of one (e.g. 1SG.IND *ven-g-o*) allows one to predict the other (e.g. 3PL.FUT *ven-dr-emos*) and vice versa. This perfect predictability has come about analogically and so it appears that systematic differences can also steer morphological change. It remains to be investigated whether predictable identities (i.e. morphemes) and predictable contrasts are different in any empirically meaningful way.

Another property that previous morphomic literature has usually ascribed to morphemes is, as I mentioned before, that they are **diachronically resilient**. That is, even though these structures often constitute what might seem to be a gratuitous complication, it is not the case that language users (or analogical changes) get rid of them within a few generations. As far as I can tell, the identification of resilience as a characteristic property of morphemes was based to date exclusively on the evidence of Romance, which is unfortunate. This dissertation has confirmed that this is not a parochial feature of Romance morphemes, as comparable evidence has been found in various other language families, most notably East Kiranti (see Tables 120 to 123 and Herce forthcoming-a), Saami (see Sections 5.2.54-55 and Herce 2020), Chinantec (see Sections 5.2.13-14), and Daghestanian (see Section 5.2.35).

Here too, it must be noted that these findings are subject to some caveats and limitations. On the one hand, one has to take so-called ‘survivor(ship) bias’ into account (see e.g. Mangel & Samaniego 1984). Since this dissertation has focused on existing morphomic patterns, “deceased” and unstable morphemes and their characteristics must necessarily be under-represented. Thus, whereas the evidence from Saami or East Kiranti has been extensively discussed in this dissertation, the patterns in closely-related Finnic and West Kiranti respectively have been completely ignored. The morphological affinities in the latter families, in contrast to the former, show a very notable variability from one language to another. This “mess”, of course, invites less comparative and diachronic work in general but must be associated to the greater instability of (some of) those morphomic patterns.

A second caveat with respect to the diachronic stability of morphemes is more ontological in nature. Even looking at the patterns that did manage to survive in a language or language family, it is difficult to say whether they are more stable or less stable than other traits. Stability and resilience are relative, not absolute concepts. Two millennia may be very long indeed compared to the duration of human life events, but very little time on biological evolutionary timescales. The evolution of human language is likely to fall somewhere in-between. An assessment of whether morphemes are resilient or not must necessarily involve a comparison with various other linguistic traits such as the lifespan of morphemes, ergativity, the phoneme /x/ etc. Future research could be aimed at systematically assessing the relative stability of these and other traits in language.

Morphemes are defined as systematic formal identities that do not map onto syntactic or semantic natural classes. The present cross-linguistic research has also shown, however, that, beyond this definitionally shared property, morphemes can differ dramatically in many respects: in their syntagmatic location (in prefixes, stems, or suffixes), their formal diversity (i.e. number of allomorphs), their restriction to particular morphosyntactic environments, their generality across the lexicon, the number of different word forms they span, their informativity in the overall system of morphological contrasts, their geometrical “shape” within the paradigm etc. **This dissertation has identified what exactly those dimensions are along which morphemes may be different and has proposed novel ways to operationalize and measure this variation** in the most finely grained way possible. Adopting methodologies like Canonical or Multivariate Typology, thus, wide typological and comparative research is possible even on such idiosyncratic entities as morphemes. As already hinted at in Section 5.4.10, the variables surveyed in this dissertation do not exhaust all variation. The token frequency of a morpheme, for example, (e.g. operationalized as the combined usage frequency of a morpheme’s cells as a proportion to the total frequency of the lexeme) is likely to be a factor of the utmost importance but was not registered in this database because token frequency is usually not reported, unfortunately, in grammatical descriptions.

After assembling a big enough sample of morphemes and finding ways of measuring different logically independent aspects about their form and distribution, we have now a better knowledge of what morphemes tend to be like (see the general properties of morphemes in

Section 5.4) and, by way of statistical analysis, what logically independent properties tend to occur together (see Section 5.5). This gives us insights into linguistic cognition and the properties of morphological architecture. Future research should be aimed at contrasting the properties of morphemes with those of morphemes. Assembling a big database of morphemes (i.e. of natural-class-distributed formatives) which is structured along the same variables as the morpheme database in this dissertation would provide the resource that is missing to quantify whether or to what extent morphemes and morphemes are different types of morphological signs in a deep meaningful way (see Herce 2020).

This dissertation has also provided evidence for various cognitive and diachronic pressures on morphemes. Most of the earlier literature has tended to regard these entities as accidental, unique, idiosyncratic structures that, because of their very nature, were largely incompatible with the extraction of meaningful generalizations. Here it has been found, that, quite on the contrary, various regularities can indeed be observed. In the domain of person-number agreement, for example, **some unnatural patterns (namely SG+3PL, 1SG+3, 2+1PL, PL+1SG, PL+2SG, and PL+3SG) have been found to be recurrent** and represented by 3 or more morphemes each in my sample. A cogent explanation of why these particular morphemes are more frequent than the other logically possible combinations (e.g. SG+1PL, 2+3SG, 1SG+2PL etc.) must involve a variety of factors. Among these, I have highlighted the importance of Zipf's law and the tendency of more frequent values (SG, 3) to be unmarked relative to more infrequent ones. I have shown (see Section 4.1.1.1) how a vague accidental split between marked and unmarked values can be transformed by sound change into a more robust morphomic split. Token frequency may also favour morphomic patterns where deviations from naturalness occur in more frequent cells (e.g. PL+3SG rather than SG+2PL).

Another related finding of this dissertation is that **morphemes tend to span a geometrically contiguous set of cells** (see all the recurrent person-number patterns above) rather than a discontinuous paradigmatic space (e.g. 1SG+2SG+3PL, 1PL+3SG, etc.). This seems to agree with some proposed cognitive biases in categorial learning (e.g. Pertsova 2011), which might make "discontinuous" morphological affinities harder to acquire (consider also the so-called *ABA constraints, see e.g. Bobaljik & Sauerland 2018, and morphological models like that of McCreight & Chvany 1991).

Alternatively to (or in conjunction with) a learning bias, it could be proposed that geometrically contiguous patterns may also arise more frequently in the course of language change. Both hypotheses (i.e. that geometrically contiguous morphological patterns emerge more frequently and/or that they are more learnable) are explained, in any case, with recourse to the feature value relations between the cells involved, which demands, in my opinion, that the importance of morphosyntactic and semantic structure in morphology is acknowledged (see section 6.2). Thus, **even in the realm of morphemes, morphosyntactic values and distinctions seem to constitute an important constraining factor.**

The above discussion connects to the ongoing one in the field about what the best way is to explain synchronic typological universals and trends. The literature (e.g. Schmidtke-Bode et al. 2019) has identified a dichotomy between so-called source-oriented and result-oriented constraints and explanations. Future research on the morpheme could be aimed at contributing to this literature by ascertaining which of these forces is (more) responsible for the synchronic prevalence of geometrically contiguous morphological patterns. An artificial language acquisition experiment (à la Kirby et al. 2008, Nevins et al 2015 etc.) where one compares the learnability of natural, morphemic-contiguous, and morphemic-noncontiguous patterns will help place this preference for contiguous, more natural patterns in the synchronic or in the diachronic sphere.

One of the major findings of this dissertation has been, therefore, that **the properties of morphemes are not random.** Nowhere is this clearer than under statistical analysis. The various logically independent variables identified in Section 5.4 have often been found to be significantly correlated. It has been found, among others, that more paradigmatically restricted morphemes tend to be more recurrent in the lexicon. There seems to be a trade-off, thus, between lexical and grammatical informativity. Such a divide between these two types of information may also lie at the core of the asymmetry found here between the properties of morphemic stems and morphemic affixes. Another interesting trade-off found here appears to involve different aspects of morpheme complexity, as bigger morphemes will tend to be relatively orderly, and messy morphemes will tend to be comparatively small, thus seemingly establishing an upper boundary to a morpheme's overall distributional complexity.

The findings of this dissertation, both incremental and novel, argue thus in favour of the view that **morphology cannot be reduced to either morphosyntactic values and their expression, nor to formal resonances and the abstraction of exclusively morphological patterns.** Both morphosyntactic (Section 6.1) and formal (Section 6.2) templates and niches must be allowed to be an active part of morphological architecture. Furthermore, the strength of one or the other will most likely vary from one part of the paradigm to another. While in the most frequent areas of the paradigm (e.g. SG, 3, PRES) formal resonances are likely to be strong due to their robust presence in the input, in relatively infrequent values (e.g. DU, SBJV, FUT) morphosyntactic structure is likely to prevail as the main organizational principle of morphological contrasts.

Different types of patterns will also plausibly demand different analyses, not only on the part of the linguist, but also on the part of the language user. There is no reason, thus, to believe that one size must fit all. In a canonical morphosyntactically well-behaved inflectional system that abides by the principle of one-form one-meaning (e.g. Turkish nominal declension), analyzing concrete exponents as expressions of particular values (e.g. DAT, PL) seems to be the easiest thing to do. By contrast, in a deeply morphomic system like many of the ones that have been presented in this dissertation (e.g. Daasanach, Chinantec, Murrinh-Patha, Ngkolmpu, Saami, Yagaria etc.), autonomously morphological rules, and using forms to predict other forms (see Table 315) might be the best solution.

Even to narrow down the paradigmatic distribution of one single formative, however, simultaneous reference to form and function might sometimes be needed. Consider, for example, the distribution of the Yakkha suffixes *-wa* and *-me* in Tables 122 and 123. As explained there, reference to the morphomic stem alternation pattern coextensive to them is unavoidable. However, these are still present tense suffixes, and are consequently found everywhere through the present and nowhere outside the present. Form-provided and feature value-provided templates are thus being used in one and the same exponent.

This account of how grammar works (i.e. one with form and meaning-provided niches for morphological elements) is cognitively realistic and is grounded on abundant evidence on how homo sapiens and other species make sense of their daily experience. Categorical

perception (e.g. Harnad 2005) will often lead language users to form discrete grammatical categories even in the presence of gradient evidence. There is, however, no reason to think that only one source of evidence (e.g. meaning, feature values) will be used for this purpose while all others (e.g. form) are completely ignored. It seems more likely that all the possible different sources of evidence will be used to some extent when making sense of linguistic input (compare to the renowned McGurk effect in the domain of phonemic perception).

Thus, as mentioned by Silvey et al. (2015), “a language can be seen as a dynamic system where the meanings of individual words adapt to, as well as themselves contributing to, the salience of particular dimensions in contexts of learning and use”. Similarly, in the domain of grammar and of inflectional morphology in particular, formal (i.e. acoustic or visual) along with various sorts of semantic and syntactic information can all serve as the basis for language-users to form their linguistic categories. It may well be the case that some kinds of evidence (e.g. morphosyntactic values like ‘speaker’, ‘plural’, or ‘past’) are more salient than others (e.g. formal similarity, or predictability) and that linguistic categorization tends to be aligned preferably to those dimensions. This, however, should be subject to empirical testing and not adopted as the initial axiom of our models of how speakers structure their grammars.

Appendix

Morphome identifier	S-co nstr	W-co nstr	Word- forms	Prdgm recur.	Lexic. recur.	Expon ents	Avg. form	Locus mark.	MS Infor.	MS Coh.	Morph compl	Pattern recur.
Achumawi	0	1	7	0	90	16	1.8	0	0	33.3	4	63
Aguaruna	0	1	1	0	100	4	1.5	0.5	1	33.3	4	63
Aragonese	1	0	9	0	1	2	1.5	0	0	42.9	8	3
Athpariya	2	2	3	0	100	2	3	1	0.5	33.3	4	63
Ayoreo	0	0	3	0	21	28	1.9	0.5	0	33.3	4	63
Bantawa1	1	0	8	1	5.5	5	1	0	0	33.3	4	63
Bantawa2	0	1	6	0	61	8	1	0	0	33.3	4	63
Bantawa3	1	1	5	1	11	5	1	0	0	33.3	4	63
Barai	1	0	1	1	100	2	1	1	1	33.3	4	63
Basque	0	0	4	0	0.1	3	2	0.5	0.5	33.3	4	63
Benabena	0	1	5	0	66.6	2	1	0.5	0	33.3	4	63
Biak	0	1	4	0	100	2	1.5	1	1	33.3	4	63
Burmeso1	0	1	1	0	100	2	1	1	1	25	6	4
Burmeso2	0	1	1	0	100	2	1	1	1	25	6	4
Burushaski	0	1	1	0	100	3	1.2	1	1	33.3	4	63
Chinantec L.1	0	2	6	0	5.1	10	1.7	0	0	44.3	8	5
Chinantec L.2	0	0	14	0	3.3	8	1.75	0	0	38.2	12	1
Chinantec P.	0	1	8	0	4.2	11	1.9	0	0	44.3	8	5
Daju M.	1	0	1	1	100	5	1.4	1	1	33.3	4	63
Darma	1	1	1	1	100	2	1	1	1	33.3	4	63
Daasanach1	0	1	1	0	100	23	1.2	0	1	40	8	2
Daasanach2	0	0	1	0	100	21	2.1	0	1	39.7	12	1
Ekari1	2	0	3	1	100	2	1.5	1	0	33.3	12	2
Ekari2	2	0	5	1	100	2	1.5	1	0	36.5	12	2
English1	1	0	1	1	0.01	2	1.5	0	1	33.3	4	63
English2	1	0	1	0	0.03	3	1	0	0	33.3	4	63
French	0	1	6	0	7	7	1.3	0	0.5	33.3	4	63
Fur1	1	0	2	0	99	20	1.6	0	0.5	33.3	4	63
Fur2	0	0	5	0	99	14	1.5	0	0.5	33.3	4	63
Girawa	0	1	4	1	30	6	1.2	0.5	0	33.3	4	63
Greek	1	0	4	0	40	2	1	1	0	33.3	4	63
Icelandic	1	0	6	0	5.8	13	1.6	0	0.5	33.3	4	63

Morphome identifier	S-co nstr	W-co nstr	Word- forms	Prdgm recur.	Lexic. recur.	Expon ents	Avg. form	Locus mark.	MS Infor.	MS Coh.	Morph compl	Pattern recur.
Iraqw1	0	1	2	0	100	5	1.2	0	0.5	46.6	8	7
Iraqw2	0	1	4	0	90	6	1.16	0	0.5	42.9	8	3
Irish	0	0	1	0	40	4	1.6	0.5	1	0	4	10
Italian1	1	0	4	0	9	5	1.4	0	0	40	12	2
Italian2*	1	1	2	0	9	6	1.3	0	0	44.3	8	5
Italian3	1	1	3	0	1	11	1.5	0	0.5	33.3	4	63
Jabuti	0	1	2	0	7	2	1	0	0	33.3	4	63
Jerung	0	0	5	0	20	7	1.6	0	0.5	33.3	4	63
Karamojong	2	1	1	1	100	2	1.5	1	1	33.3	4	63
Kele	0	0	9	0	4	6	1.6	0	0	33.3	4	63
Ket	2	0	1	0	70	2	1	1	1	33.3	4	63
Khaling1	1	1	3	0	15	4	1	0	0	0	4	10
Khaling2	0	1	5	0	7	4	1	0	0	33.3	4	63
Khinalug1	0	1	1	0	30	3	1	1	1	0	4	10
Khinalug2	0	1	1	0	25	2	1	1	1	25	6	4
Koasati	0	1	3	0	40	4	2.5	1	0	33.3	4	63
Koiari	1	0	2	1	100	2	1.5	1	1	33.3	4	63
Kosena	1	2	2	1	100	2	1.5	1	1	33.3	4	63
Luxembourgish	1	0	2	0	5	13	1.1	0	0.5	33.3	4	63
Maijiki	1	1	1	0	100	4	2	1	1	33.3	4	63
Malinaltepec M.	0	0	5	0	3	10	2.2	0.5	0.5	33.3	4	63
Mazatec C.1	0	1	2	0	32	13	1.5	0.5	0.5	33.3	4	63
Mazatec C.2	0	1	4	0	32	5	1.8	0	0.5	33.3	4	63
Mehri	2	1	2	0	50	3	1	0	0	0	4	10
Menggwa Dla	0	2	3	0	1	4	1.75	0	0	44.3	8	5
Mian	1	1	1	1	100	3	1.3	1	1	25	6	4
Murrinh-Pathal	0	1	2	0	100	12	1.4	0.5	1	33.3	4	63
Murrinh-Patha2	0	1	3	0	100	14	1.6	0.5	1	33.3	4	63
Nen	0	0	3	1	100	3	1	1	0.5	33.3	4	63
Ngkolmpu1	0	1	1	1	100	3	1.7	1	1	0	4	10
Ngkolmpu2	0	1	1	1	100	2	1	1	1	33.3	4	63
Nimboran1	0	0	9	0	80	30	1.4	0	0	31.7	12	1
Nimboran2	0	1	6	0	70	3	1.7	0	0	46.6	8	7
Nivkh	1	0	1	1	100	2	1	1	1	33.3	4	63
North Saami	0	0	1	0	100	2	2	1	1	0	4	10

Morpheme identifier	S-co nstr	W-co nstr	Word- forms	Prdgm recur.	Lexic. recur.	Expon ents	Avg. form	Locus mark.	MS Infor.	MS Coh.	Morph compl	Pattern recur.
N. Akhvakh	2	0	4	1	100	6	1.5	0.5	1	33.3	4	63
Old English	1	0	3	0	7	13	1.6	0	0.5	42.9	8	63
Paez	1	1	1	1	100	3	3.7	1	1	33.3	4	63
Pite Saami1	0	1	3	0	90	5	1.3	0	1	33.3	4	63
Pite Saami2	0	1	5	0	20	3	2	0	0	33.3	4	63
Pite Saami3	0	0	8	0	80	5	1.3	0	0.5	39.7	12	1
Pite Saami4	0	0	3	0	10	4	1	0	0	26.7	18	1
Pite Saami5	0	0	9	0	10	4	1	0	0	40.7	18	1
Skolt Saami1	0	0	5	0	30	7	1	0	0	33.3	8	1
Skolt Saami2	0	0	7	0	30	7	1	0	0	40	8	1
Skolt Saami3	1	0	3	0	10	9	1.2	0	0	0	4	10
Sobei1	1	0	5	0	0.5	2	1	0	0.5	33.3	4	63
Sobei2	0	0	7	0	0.5	2	1	0	0.5	46.6	8	7
Spanish1	1	0	4	0	3	2	1.5	0	0	33.3	4	63
Spanish2	1	0	6	0	3	4	1.5	0	0	46.6	8	7
Spanish3*	2	1	3	0	0.3	2	1.5	0	0	33.3	4	63
Sunwar	2	0	5	0	5	3	1	0	0	33.3	4	63
Svan	2	0	5	0	95	11	1.2	0.5	0.5	33.3	4	63
Thulung1	0	0	8	0	5	3	1	0	0.5	23.8	9	1
Thulung2	0	0	12	0	8	3	1	0	0.5	46.6	8	7
Tol1	0	1	4	0	65	5	1.5	0	0.5	33.3	4	63
Tol2	0	0	11	0	40	13	1.6	0.5	0.5	38.1	12	1
Tol3	0	0	7	0	15	16	1.7	0	0.5	38.8	12	1
Tol4	0	1	4	0	10	5	1.3	0	0.5	33.3	4	63
Tol5	0	0	10	0	0.4	2	1	0.5	0.5	42.9	12	1
Tol6	0	1	8	0	0.6	3	1	0	0.5	46.6	8	1
Tol7	0	0	4	0	0.2	3	1	0	0	33.3	4	63
Turkana1	0	1	3	0	100	2	1	1	1	44.3	8	5
Turkana2*	0	1	3	0	100	2	1	1	1	46.6	8	7
Twi	1	0	2	0	100	2	2	1	1	0	4	10
Udmurt	0	1	6	0	100	2	1.5	1	0.5	46.6	8	7
Vitu	1	0	2	0	100	2	1	1	0.5	33.3	4	63
Vures	0	0	8	0	50	5	1	0	0	33.3	4	63
Wambisa1	0	0	1	0	100	2	1	1	1	33.3	4	63

Morphome identifier	S-co nstr	W-co nstr	Word-forms	Prdgm recur.	Lexic. recur.	Expon ents	Avg. form	Locus mark.	Infor.	MS Coh.	Morph compl	Pattern recur.
Wambisa2	1	1	2	1	100	2	1	1	0	33.3	4	63
Wutung1	0	1	2	0	10	4	1.3	0	1	0	4	10
Wutung2	0	0	3	0	15	3	1.3	0	1	33.3	12	2
Wutung3	0	0	4	0	15	3	1.3	0	1	36.5	12	2
Wutung4	0	0	2	0	15	3	1	0	1	33.3	4	63
Yagaría	1	2	2	1	100	4	1	1	1	0	4	10
Yorno-So	1	0	4	1	100	6	1.5	1	0.5	33.3	4	63
Zapotec1	1	0	6	0	15	8	1	0	0	33.3	4	63
Zapotec2*	0	0	16	0	8	4	1	0	0	33.3	4	63

Descriptive statistics

	n	mean	sd	median	trim	mad	min	max	range	skew	kurt	se
Strong constr	110	0.48	0.66	0.00	0.36	0.00	0.00	2.00	2.00	1.02	-0.17	0.06
Weak constraints	110	0.52	0.59	0.00	0.47	0.00	0.00	2.00	2.00	0.61	-0.63	0.06
Word forms	110	4.05	2.98	3.00	3.64	2.97	1.00	16.00	15.00	1.35	2.14	0.28
Paradigm recurr	110	0.20	0.40	0.00	0.12	0.00	0.00	1.00	1.00	1.48	0.19	0.04
Lexicon recurr	110	50.5	42.7	40.0	50.59	57.82	0.01	100.0	99.99	0.11	-1.82	4.07
Exponents	110	5.81	5.48	4.00	4.65	2.97	2.00	30.0	28.00	2.22	5.32	0.52
Shared form	110	1.38	0.44	1.30	1.31	0.44	1.00	3.70	2.70	2.03	6.86	0.04
Locus marking	110	0.38	0.45	0.00	0.35	0.00	0.00	1.00	1.00	0.48	-1.63	0.04
Informativity	110	0.49	0.43	0.5	0.48	0.74	0.00	1.00	1.00	0.05	-1.63	0.04
MS Coherence	110	32.2	11.4	33.3	34.22	0.00	0.00	46.6	46.6	-1.81	3.11	1.09
Morph complex	110	5.94	3.17	4.00	5.28	0.00	4.00	18.0	14.00	1.68	2.43	0.30
Cross-ling rec	110	38.1	29.1	63.0	39.6	0.00	1.00	63.0	62.0	-0.31	-1.90	2.77

Some of the variables above are not numerical but have been coded as such for ease of presentation. Thus, paradigm recurrence (i.e. whether or not a morpheme occurs more than once in a single lexeme's paradigm) is a binary, yes/no variable and has been coded as 1 (for "yes" and 0 for "no". Locus of marking, in turn, has been coded as 0 for "stem", 1 for "affix" or 0.5 when a morpheme includes affix and stem formatives or when its assignment to stem or affix is analytically uncertain. Informativity has been coded as 0 "uninformative", 0.5 "partially informative" or 1 "fully informative".

References

Aalberse, Suzanne Pauline. 2007. The typology of syncretisms and the status of feature structure. Verbal paradigms across 355 Dutch dialects. *Morphology* 17, 1: 109-149.

Ackerman, Farrell; James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (Eds.), *Analogy in Grammar: Form and Acquisition*: 54-81. Oxford: Oxford University Press.

Ackerman, Farrell & Olivier Bonami. 2017. Systemic polyfunctionality and morphology-syntax interdependencies. In Gisborne, Nikolas Gisborne & Andrew Hippisley (Eds.) *Defaults in Morphological Theory*. Oxford: Oxford University Press.

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

Aikhenvald, Alexandra Y. 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.

Aikhenvald, Alexandra I. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.

Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78, 4: 684-709.

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In Gina Garding and Mimu Tsujimura (Eds.), *West Coast Conference on Formal Linguistics 22 proceedings*:1-14. Somerville, MA: Cascadilla Press.

Albright, Adam. 2010. Base-driven leveling in Yiddish verb paradigms. *Natural Language & Linguistic Theory* 28.3: 475-537.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90, 2: 119-161.

Ali, Mohammed, and Andrzej Zaborski. 1990. *Handbook of the Oromo language*. Cracow: Polska Academia Nauk-Oddzial W. Krakowie.

Amado, Rosane de Sá. 2004. *Aspectos morfofonológicos do Gavião-Pykobjê*. PhD thesis. University of São Paulo.

Anceaux, Jean C. 1965. *The Nimboran Language: Phonology and Morphology*. 's Gravenhage: Martinus Nijhof.

Anderson, Judi Lynn. 1989. *Comaltepec Chinantec syntax*. *Studies in Chinantec languages* 3. SIL.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2011. Stress-conditioned allomorphy in Surmiran (Rumantsch). In Maria Goldbach, Marc-Olivier Hinzelin, Martin Maiden & John Charles Smith (Eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*: 13–35. Oxford: Oxford University Press.

Anderson, Stephen R. 2015. The morpheme: Its nature and use. *The Oxford handbook of inflection*: 11-33.

de Angulo, Jaime & Lucy Shepard Freeland. 1930. The Achumawi language. *International Journal of American Linguistics* 6.2: 77-120.

Arkadiev, Peter & Maria Klamer. 2018. Morphological theory and typology. In Jenny Audring and Francesca Masini (Eds.) *Oxford Handbook of Morphological Theory*. Oxford: Oxford University Press.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge (MA): MIT press.

Aronoff, Mark. 2016. Competition and the lexicon. In Annibale Elia, Claudio Iacobini & Miriam Voghera (Eds.). *Livelli di Analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della società di linguistica italiana*. 39-52. Roma: Bulzoni Editore.

Austing, John F. & Randolph Upia. 1975. Highlights of Ömie morphology. In: Tom E. Dutton (Ed.), *Studies in Language of Central and South-East Papua*: 513-598. *Pacific Linguistics* 29. Canberra: Australian National University.

Avilés, Arthur J. 2008. *The Phonology and Morphology of the Dar Daju Daju Language*. MA Dissertation. University of North Dakota.

Avril, Yves. 2006. *Parlons komi*. Paris: L'Harmattan.

Axenov, Serge. 2006. *The Balochi Language of Turkmenistan: A corpus-based grammatical description*. PhD Dissertation. *Acta Universitatis Upsaliensis*.

Baerman, Matthew. 2007a. Syncretism. *Language and Linguistics Compass*. 1, 5: 539-551.

Baerman, Matthew. 2007b. Morphological reversals. *Journal of Linguistics* 43.1: 33-61.

Baerman, Matthew. 2012. Paradigmatic chaos in Nuer. *Language* 88, 3: 467-494.

Baerman, Matthew. 2013. Inflection class interactions. In Nabil Hathout, Fabio Montermini, and Jesse Tseng (Eds.), *Morphology in Toulouse: Selected proceedings of Décembrettes 7 (Toulouse, 2–3 December 2010)*:11–33. Munich: Lincom.

Baerman, Matthew & Dunstan Brown. 2013. Syncretism in verbal person/number marking. In: Matthew S. Dryer & Martin Haspelmath (Eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Plank Institute for Evolutionary Anthropology.

Baerman, Matthew & Greville G. Corbett. 2007. Linguistic typology: morphology. *Linguistic Typology* 11,1: 115-117.

Baerman, Matthew & Greville G. Corbett. 2013. Person by other means. *Languages across boundaries: Studies in memory of Anna Siewierska*: 1-14.

Baerman, Matthew; Dunstan Brown & Greville G. Corbett. (Eds.). 2017. *Morphological Complexity*. Cambridge: Cambridge University Press.

Baerman, Matthew; Greville G. Corbett & Dunstan Brown (Eds.). 2010. *Defective paradigms: Missing forms and what they tell us*. Oxford: Oxford University Press.

Baerman, Matthew; Dunstan Brown & Greville G. Corbett. 2005. *The syntax-morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.

Baerman, Matthew; Dunstan Brown & Greville G. Corbett. 2017. External Typology of Inflection Classes. In *Morphological Complexity*: 8-33. Cambridge: Cambridge University Press.

Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *The Oxford Handbook of Language Typology*, 90–127. Oxford: Oxford University Press.

Bank, Sebastian & Jochen Trommer. 2012. Paradigm learning and subanalysis complexity. *Proceedings of the 48th Annual Meeting of the Chicago Linguistic Society*.

Barcos, Miguel Ángel. 2007. *El aragonés ansotano: estudio lingüístico de Ansó y Fago*. Institución Fernando el Católico.

Baudouin de Courtenay, Jan. 1895 [1972]. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (Ed.), *A Baudouin de Courtenay anthology*. Bloomington, IN: Indiana University Press

Bauer, Laurie; Rochelle Lieber & Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Bazell, Charles E. 1938. Archimorpheme and Phonomorpheme. *Modern Language Notes* 53.5: 363-366.

Bazell, Charles E. 1960. A question of syncretism and analogy. *Transactions of the Philological Society*. 1-12.

Beard, Robert. 1995. *Lexeme-morpheme base morphology: a general theory of inflection and word formation*. Albany: Suny Press.

Beaton, Arthur Charles. 1968. *A grammar of the Fur language*. Sudan Research Unit, Faculty of Arts, University of Khartoum.

van den Berg, Helma. 1995. *A grammar of Hunzib (with texts and lexicon)*. Munich: Lincom.

van den Berg, René; & Peter Bachet. 2006. *Vitu grammar sketch*. Ukarumpa: SIL.

Bermúdez-Otero, Ricardo & Ana R. Luís. 2016. A view of the morpheme debate. In Ana R. Luís & Ricardo Bermúdez-Otero (Eds.) *The morpheme debate*: 309-40. Oxford: Oxford University Press.

Berthiaume, Scott. 2004. *A Phonological Grammar of Northern Pame*. PhD Dissertation. University of Texas at Arlington.

Bertinetto, Pier Marco. 2009. *Ayoreo (Zamuco). A grammatical sketch*. Quaderni del Laboratorio di Linguistica 8. Pisa: Scuola Normale Superiore.

Bickel, Balthasar. 1995. In the vestibule of meaning: transitivity inversion as a morphological phenomenon. *Studies in Language* 19, 1: 73-127.

Bickel, Balthasar. 2010. Capturing particulars and universals in clause linkage. In Isabelle Brill (Ed.) *Clause-hierarchy and clause-linking: the syntax and pragmatics interface*: 51-101. Amsterdam: John Benjamins.

Bickel, Balthasar & Martin Gaenszle. 2015. First person objects, antipassives, and the political history of the Southern Kirant. *Journal of South Asian Languages and Linguistics* 2, 1: 63-86.

Bickel, Balthasar & Johanna Nichols. 2002. Autotypologizing databases and their use in fieldwork. In *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics, Las Palmas*, vol. 2627. Nijmegen: MPI for Psycholinguistics.

Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (Ed.) *Language typology and syntactic description* 3, 2: 169-240. Cambridge: Cambridge University Press.

Bittner, Dagmar. 1995. Affixhomonymie in der Natürlichkeitstheorie. In: Norbert Boretzky et al. (eds.) *Natürlichkeitstheorie und Sprachwandel: Beiträge zum internationalen Symposium über Natürlichkeitstheorie und Sprachwandel an der Universität Maribor, 1993*: 123-144. Bochum: Brockmeyer.

Blevins, James P. 1995. Syncretism and paradigmatic opposition. *Linguistics and Philosophy* 18. 113-152.

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.

Blevins, James P., Farrell Ackerman, Robert Malouf & Michael Ramscar. 2016. Morphology as an adaptive discriminative system. In Daniel Siddiqi & Heidi Harley (Eds.) *Morphological Metatheory*. Amsterdam: John Benjamins.

Blevins, James P. & Juliette Blevins (Eds.). 2009. *Analogy in grammar: Form and acquisition*. Oxford: Oxford University Press.

Blevins, James P; Petar Milin & Michael Ramscar. 2017. The Zipfian Paradigm Cell Filling Problem. In F. Kiefer, J. P. Blevins & H. Bartos (Eds.), *Perspectives on Morphological Structure: Data and Analyses*. Leiden: Brill.

Bloomfield, Leonard. 1926. A set of postulates for the science of language. *Language* 2, 3: 153-164.

Bloomfield, Leonard. 1943. Meaning. *Monatshefte für deutschen Unterricht* 35.3/4: 101-106.

Bobaljik, Jonathan David & Uli Sauerland. 2018. ABA and the combinatorics of morphological features. *Glossa: a journal of general linguistics* 3, 1.

Bonami, Olivier. 2017. Predictability in inflection and word formation. Talk delivered in ParadigMo 2017.

Bonami, Olivier, & Pollet Samvelian. 2009. Inflectional periphrasis in Persian. HPSG 2009 Conference. CSLI Publications.

Bonami, Olivier & Gregory T. Stump. 2016. Paradigm Function Morphology. In Hippisley, Andrew and Stump, Gregory T. (Eds.), *Cambridge Handbook of Morphology*: 449-481. Cambridge: Cambridge University Press.

Booij, Geert. 2016. Morphology: the structure of words. In Keith Allan (Ed.) *The Routledge Handbook of Linguistics*: 104-117. London: Routledge.

Booij, Geert & Jenny Audring. 2017. Construction morphology and the parallel architecture of grammar. *Cognitive science* 41, 2: 277-302.

Borchers, Dörte. 2008. A grammar of Sunwar: descriptive grammar, paradigms, texts and glossary. Leiden: Brill.

Boyé, Gilles. 2000. Problèmes de morpho-phonologie verbale en français, en espagnol et en italien. PhD Dissertation, University Paris VII.

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.

Boyé, Gilles & Gauvain Schalchli. 2016. The status of paradigms. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology*: 206–234. Cambridge: Cambridge University Press.

Braune, Wilhelm & Ingo Reiffenstein. 2004. *Althochdeutsche Grammatik I: laut-und Formenlehre*. Tübingen: Niemeyer.

Brown, Dunstan & Marina Chumakina. 2013. What there might be and what there is: an introduction to Canonical Typology. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (Eds.) *Canonical morphology and syntax*: 1-19. Oxford: Oxford University Press.

Burzio, Luigi. 2001. Zero derivations. *Linguistic Inquiry* 32.4: 658-677.

Butler, Inez M. 1980. Gramática zapoteca: Zapoteco de Yatzachi el Bajo. Serie de gramaticas de lenguas indigenas de Mexico, 4. México: Instituto Lingustico de Verano.

Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Caha, Pavel. 2009. *The Nanosyntax of Case*. Ph.D. thesis, University of Tromsø.

Camproux, Charles. 1958. *Étude syntaxique des parlers gévaudanais*. Montpellier: Université de Montpellier Faculté des lettres.

Cano González, Ana M. 1981. El habla de Somiedo (Occidente de Asturias). Santiago de Compostela: Universidade de Santiago de Compostela.

Cardinaletti, Anna & Giuliana Giusti. 2001. Semi-Lexical Motion Verbs in Romance and Germanic. In Norbert Corver & Henk van Riemsdijk (Eds.), *Semi-Lexical Categories: The Function of Content Words and the Content of Function Words*: 371-414. Berlin: de Gruyter.

Carlin, Eithne B. 2006. Feeling the need: The borrowing of Cariban functional categories into Mawayana (Arawak). In Alexandra Y. Aikhenvald and R. M. W. Dixon (Eds.), *Grammars in contact: A cross-linguistic typology (Explorations in linguistic typology 4)*: 313-32. Oxford: Oxford University Press.

Carroll, Matthew J. 2016. The Ngkolmpu Language with special reference to distributed exponence. PhD Dissertation. Australian National University.

Carstairs-McCarthy, Andrew. 1994. Inflection classes, gender, and the principle of contrast. *Language*: 737-788.

Carstairs-McCarthy, Andrew. 1998. How lexical semantics constrains inflectional allomorphy. *Yearbook of Morphology 1997*: 1-24. Amsterdam: Springer.

Carstairs-McCarthy, Andrew. 2010. *The evolution of morphology*. Oxford: Oxford University Press.

Cassidy, Frederic G. & Richard N. Ringler (Eds). 1971 (3rd ed). *Bright's Old English grammar and reader*. Fort Worth: Harcourt Brace Jovanovich College Publishers

Chumakina, Marina & Corbett, Greville G. 2015. Gender-number marking in Archi: small is complex. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (Eds.), *Understanding and measuring morphological complexity*. Oxford: Oxford University Press.

Ciucci, Luca. 2016. Inflectional morphology in the Zamucoan languages. Asunción: CEADUC.

Ciucci, Luca & Pier Marco Bertinetto. 2017. Possessive inflection in Proto-Zamucoan. *Diachronica* 34.3: 283-330.

Coats, Herbert S. 1973. Old Russian declension: a synchronic analysis. In Ladislav Matejka (ed.) *American Contributions to the Seventh International Congress of Slavists: I: Linguistics and Poetics*: 67-99. The Hague: Mouton.

Cook, Dorothy M. & Linda L. Criswell. 1993. *El idioma koreguaje (tucano occidental)*. SIL.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (Eds.) *Linguistic diversity and language theories*: 25-49. Amsterdam: John Benjamins.

Corbett, Greville G. 2009. Canonical inflectional classes. *Selected proceedings of the 6th Décembrettes: Morphology in Bordeaux*: 1-11.

Corbett, Greville G. 2016. Morphomic splits. In Ana R. Luís & Ricardo Bermúdez-Otero (Eds.). *The morphome debate: Diagnosing and analysing morphomic patterns*: 64-88. Oxford: Oxford University Press.

Corbett, Greville G. & Matthew Baerman. 2006. Prolegomena to a typology of morphological features. *Morphology* 16.2: 231-246.

Corbett, Greville G. & Sebastian Fedden. 2016. Canonical gender. *Journal of Linguistics* 52.3: 495-531.

Coupé, Christophe, Yoon Mi Oh, Dan Dediu, & François Pellegrino. 2019. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances* 5.9.

Cowan, Hendrik Karel J. 1965. Grammar of the Sentani language. *Verhandelingen van het koninklijk instituut voor taal-, land- en volkenkunde*.

Creissels, Denis. 2008. Remarks on so-called “conjunct/disjunct” systems. Paper delivered at the conference *Syntax of the World’s Languages III*.

Crowley, Terry. 1998. An Erromangan (Sye) Grammar. *Oceanic Linguistics Special Publication 27*. University of Hawai’i Press.

Csúcs, Sándor. 1988. Die wotjakische Sprache. In Denis Sinor (Ed.) *The Uralic Languages: Description, History and Foreign Influences 1*: 131-146. Leiden: Brill.

Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press.

Cysouw, Michael. 2005. A typology of honorific uses of clusivity. In Elena Filimonova (Ed.) *Clusivity: Typology and case studies of the inclusive exclusive distinction*: 213-230. Amsterdam: John Benjamins.

Dawkins, Richard. 1996. *Climbing Mount Improbable*. London: Penguin.

Dennis, Ronald K. 1992. *Conjugación del verbo en idioma tol (jicaque)*. Vol. 8. Instituto Hondureño de Antropología e Historia.

Derksen, Rick. 2007. *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.

Dienst, Stefan. 2014. *A grammar of Kulina*. Amsterdam: de Gruyter.

Diertani, Chaya Eliana Ariel. 2011. Morpheme boundaries and structural change: Affixes running amok. PhD thesis, University of Pennsylvania.

Dimmendaal, Gerrit J. 1991. The geometry of verb paradigms in Teso-Turkana. In Frans Plank (Ed.) *Paradigms: the economy of inflection*: 275-306. Amsterdam: de Gruyter

Dixon, R. M. W. 2004. *The Jarawara language of Southern Amazonia*. Oxford: Oxford University Press.

Doble, Marion. 1987. A description of some features of Ekari language structure. *Oceanic Linguistics* 26.1/2: 55-113.

Donohue, Mark. 2001. Animacy, class and gender in Burmeso. In: Andrew Pawley, Malcolm Ross & Darrell Tryon (eds) *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* (Pacific linguistics 514), 97-115. Canberra: Pacific Linguistics.

Donohue, Mark. 2004. *A Grammar of the Skou Language of New Guinea*. Ms., National University of Singapore.

Donohue, Mark. 2015. Verbal Inflection in Iha: A Multiplicity of Alignments. In Matthew Baerman (Ed.), *The Oxford Handbook of Inflection*: 405–418. Oxford: Oxford University Press.

Doornenbal, Marius. 2009. *A grammar of Bantawa*. Utrecht: LOT.

Doyle, Andrew. 2001. *Irish*. Munich: Lincom

Drabbe, Peter. 1952. *Spraakkunst van het Ekagi*. 's-Gravenhage: Martinus Nijhoff.

Dunn, Michael. 1999. *Chukchi*. PhD dissertation, Australian National University.

Dutton, Thomas Edward. 1975. A Koita grammar sketch and vocabulary. *Pacific Linguistics*: 281-412.

Dutton, Thomas Edward. 1996. *Koiari*. Munich: Lincom.

Dutton, Thomas Edward. 2003. *A dictionary of Koiari, Papua New Guinea, with grammar notes*. Pacific linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Döhler, Christian. 2018. *A grammar of Komnzo*. Berlin: Language Science Press.

Ebert, Karen H. 1997. *A grammar of Athpare*. Munich: Lincom.

Elbourne, Paul. 2011. *Meaning: a slim guide to semantics*. Oxford: Oxford University Press.

Embick, David. 2013. On the distribution of stem alternants: Separation and its limits. In Ana R. Luís & Ricardo Bermúdez-Otero (Eds), *The morpheme debate*. Oxford: Oxford University Press.

Esher, Louise. 2014. Autonomous morphology and extramorphological coherence. *Morphology* 24, 4: 325-350.

Esher, Louise. 2015. Morphemes and predictability in the history of Romance perfects. *Diachronica* 32, 4: 494-529.

Evans, Nicholas. 2015. Inflection in Nen. In Matthew Baerman (Ed.), *The oxford handbook of inflection*. Oxford: Oxford University Press.

Evans, Nicholas; I Wayan Arka; Matthew Carroll; Christian Döhler; Eri Kashima; Emil Mittag; Kyla Quinn; Jeff Siegel; Philip Tama & Charlotte van Tongeren. 2017. The languages of Southern New Guinea. In Bill Palmer (Ed.), *The languages and linguistics of the New Guinea area*: 641–774. Berlin: de Gruyter.

Farr, Cynthia. 1999. The interface between syntax and discourse in Korafe: a Papuan language of Papua New Guinea. Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Fast, Gerhard & Ruby Fast. 1981. Introducción al idioma Achuar. Lima: Ministerio de Educación

Fedden, Sebastian. 2011. A grammar of Mian. Berlin: de Gruyter.

Fedden, Sebastian, and Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa* 2.1: 1-47.

Feist, Timothy Richard. 2011. A grammar of Skolt Saami. PhD Dissertation. University of Manchester.

Feist, Timothy. 2015. A grammar of Skolt Saami. Helsinki: Suomalais-Ugrilainen Seura.

Feist, Timothy & Enrique L. Palancar. 2015. Oto-Manguean Inflectional Class Database. University of Surrey. <http://dx.doi.org/10.15126/SMG.28/1>

Fertig, David. 2016. Mechanisms of paradigm leveling and the role of universal preferences in morphophonological change. *Diachronica* 33.4: 423-460.

Floyd, Simeon; Elisabeth Norcliffe & Lila San Roque, (Eds.). 2018. Egophoricity. John Benjamins.

Foley, William A. 1986. The Papuan languages of New Guinea. Cambridge: Cambridge University Press.

Foreman, Velma. 1973. Grammar of Yessan-Mayo. Santa Ana: SIL.

Forker, Diana. 2013. A grammar of Hinuq. Berlin: Walter de Gruyter.

François, Alexandre. 2001. Contraintes de structures et liberté dans l'organisation du discours: Une description du mwotlap, langue océanienne du Vanuatu. PhD diss., Univ. Paris Sorbonne.

François, Alexandre. 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics*: 443-504.

François, Alexandre. 2009. The languages of Vanikoro: Three lexicons and one grammar. In Bethwyn Evans (Ed.) *Discovering history through language: Papers in honour of Malcolm Ross*: 103-126. Canberra: Australian National University.

Frodl, Martin. 2013. Aspect marking in Udmurt. PhD dissertation, Masarykova univerzita.

Gaeta, Livio. 2010. Analogical change. In Silvia Luraghi and Vit Bubenik (Eds.), *The Continuum companion to historical linguistics*: 147-160. London: Continuum.

Gamkrelidze, Thomas V. 1966. A typology of Common Kartvelian. *Language*. 42, 1: 69-83.

Gasaway, Eileen & Heather Sims. 1977. *Girawa grammar*. Ukarumpa: SIL.

Gause, Georgij F. 1934. *The struggle for existence*. Baltimore: Williams and Wilkins.

Georg, Stefan. 2007. *A Descriptive Grammar of Ket (Yenisei-Ostyak): Part 1: Introduction, Phonology and Morphology*. Global Oriental.

Giraud, Hélène; Serena Dal Maso; & Sabrina Piccinin. 2016. The role of stem frequency in morphological processing. *Mediterranean Morphology Meetings*, 10.

Gordon, Matthew K. 2009. A fortition-based approach to Balto-Fennic-Sámi consonant gradation. *Folia Linguistica Historica* 18, 1-2: 49-79.

Gruzdeva, Ekaterina. 1998. *Nivkh*. Vol. 111. Munich: Lincom.

Gönczöl, Ramona. 2007. *Romanian: an essential grammar*. London: Routledge.

Haensch, Günther. 1958. Las hablas de la Alta Ribagorza. *Archivo de filología aragonesa* 10: 57-194.

Haiman, John. 1980. Hua, a Papuan language of the eastern highlands of New Guinea. Amsterdam: John Benjamins.

Halle, Morris and Marantz, Alec. 1994. Some key features of Distributed Morphology. In A. Carnie, H. Harley & T. Bures (Eds.), *Papers on Phonology and Morphology*: 275–288. Cambridge, MA: MIT Working Papers in Linguistics.

Hansson, Gunnar Oláfur. 2007. Productive syncretism in Saami inflectional morphology. In Ida Toivonen & Diane Nelson (eds.) *Saami linguistics*: 91-136. Amsterdam: John Benjamins.

Harbour, Daniel. 2007. *Morphosemantic Number: From Kiowa Noun Classes to UG Number Features*. Berlin: Springer.

Harbour, Daniel. 2008. On homophony and methodology in morphology. *Morphology* 18.1: 75-92.

Harnad, Stevan. 2005. To Cognize is to Categorize: Cognition is Categorization. In Lefebvre & Cohen (Eds.), *Handbook of Categorization in Cognitive Science*. New York: Elsevier.

Harris, Alice C. 2017. *Multiple exponence*. Oxford: Oxford University Press.

Harris, Zellig S. 1942. Morpheme alternants in linguistic analysis. *Language* 18, 3: 169-180.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language* 2: 1-30.

Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Daniël Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of Linguistic Variation*, 83–113. Berlin: De Gruyter.

Haspelmath, Martin. To appear. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability.

Haspelmath, Martin & Andrea Sims. 2010. *Understanding Morphology*. Oxford: Oxford University Press.

Hayward, Richard J. 1991. Concerning a vocalic alternation in North Omotic verb paradigms. *Bulletin of the School of Oriental and African Studies* 54.3: 535-553.

Healey, Alan; Ambrose Isoroembo & Martin Chittleborough. 1969. Preliminary notes on Orokaiva grammar. *Pacific Linguistics. Series A. Occasional Papers* 18: 33.

Heath, Jeffrey. 2011. A Grammar of Togo Kan. Unpublished grammar ms., August 2011, available on www.dogonlanguages.org.

Heath, Jeffrey. 2014. A Grammar of Yorno So. Unpublished grammar ms. available on www.dogonlanguages.org

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Hengeveld, Kees, & J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press.

Herce Calleja, Borja. 2016. Why frequency and morphological irregularity are not independent variables in Spanish: A response to Fratini et al.(2014). *Corpus Linguistics and Linguistic Theory* 12.2: 389-406.

Herce, Borja. 2018a. Information-theoretic approaches to (verbal) inflection classes: some limitations. In Pilar Guerrero Medina, and Roberto Torre Alonso (Eds.), *Verbs, clauses and constructions: Functional and typological approaches*. Cambridge: Cambridge Scholars Publishing.

Herce, Borja. 2018b. 2SG/1PL morpheme? In Papuan languages. Talk delivered in APLL10.

Herce, Borja. 2019. Morpheme interactions. *Morphology* 29.1: 109-132.

Herce, Borja. 2020. On morphemes and morphemes: exploring the distinction. *Word Structure* 13, 1.

Herce, Borja. Forthcoming-a. Stem alternations in Kiranti and their implications for morphology.

Herce, Borja. Forthcoming-b. Alignment of forms in Spanish verbal inflection: the *gang poner, tener, venir, salir, valer* as a window into the nature of paradigmatic analogy and predictability.

Herrity, Peter. 2000. *Slovene. A comprehensive grammar*. London: Routledge.

van den Heuvel, Wilco. 2006. *Biak: description of an austronesian language of papua*. Utrecht: LOT.

Hockett, Charles F. 1947. Problems of morphemic analysis. *Language* 23, 4: 321-343.

Hockett, Charles F. 1958. *A course in modern Linguistics*. New York: Macmillan

Hockett, Charles F. 1987. *Refurbishing our foundations*. Amsterdam: John Benjamins.

Holmes, Philip, and Ian Hinchliffe. 2003. *Swedish: A comprehensive grammar*. London: Routledge.

Holt, Dennis. 1999. *Tol (Jicaque)*. Munich: Lincom.

Holton, David; Peter Mackridge & Irene Philippaki-Warbuton. 2012. *Greek: A comprehensive grammar of the modern language*. London: Routledge.

Inkelas, Sharon. 1993. Nimboran position class morphology. *Natural Language & Linguistic Theory* 11, 4: 559-624.

Jackendoff, Ray S. 1985. Multiple Subcategorization and the Theta-Criterion: The Case of Climb. *Natural Language and Linguistic Theory* 3: 271-295.

Jacques, Guillaume. 2017. A reconstruction of Proto-Kiranti verb roots. *Folia Linguistica* 51: 177-215.

Jacques, Guillaume, Aimee Lahaussais, Boyd Michailovsky & Dhan Bahadur Rai. 2011. An overview of Khaling verbal morphology. *Language and linguistics*, 13(6): 1095-1170.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology* 61, 1: 23-62.

Jakobson, Roman. 1936 [1971]. Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus, in *Idem, Selected Writings 2*, The Hague: Mouton, 23–71.

Jamieson, Carole. 1988. *Gramática mazateca del municipio de Chiquihuitlan, Oaxaca*. Mexico: Instituto Linguístico de Verano.

Janda, Richard D. 1982. Of formal identity and rule-(un) collapsibility: On lost and found generalizations in morphology. *Proceedings of the 1st west coast conference on formal linguistics*.

Janda, Richard D. & María Sandoval. 1984. *"Elsewhere" in morphology*. Vol. 288. Indiana University Linguistics Club.

Jensen, John Thayer. 1990. *Morphology: Word structure in generative grammar*. Amsterdam: Benjamins.

Johnson, Orville E. & Stephen H. Levinsohn. 1990. *Gramática secoya*. SIL.

Johnston, Jason Clift. 1996. *Systematic homonymy and the structure of morphological categories: some lessons from paradigm geometry*. PhD dissertation. University of Sydney.

Jones, Wendell & Paula Jones. 1991. *Barasano syntax*. Summer Institute of Linguistics & University of Texas at Arlington.

Joseph, Brian D. 1997. *How general are our generalizations? What speakers actually know and what they actually do*. Eastern States Conference on Linguistics (ESCOL'96).

Jung, Ingrid. 1989. *Grammatik des Paez: Ein Abriss*. PhD dissertation, Universität Osnabrück.

Jörg, Christina. 1989. *Isländische Konjugationstabellen: Icelandic Conjugation Tables-Tableaux de Conjugaison Islandaise-Beygingatöflur Islenskra Sagna*. Hamburg: Buske.

Kahn, Lily & Riitta-Liisa Valijärvi. 2017 *North Sámi: An Essential Grammar*. London: Routledge.

Kamil, Mohamed Hassan. 2004. *Parlons afar: langue et culture*. Paris: L'Harmattan.

Kaye, Steven J. 2013. *Morphomic stems in the Northern Talyshi verb: diachrony and synchrony*. In Silvio Cruschina, Martin Maiden, & John Charles Smith (Eds.), *The boundaries of pure morphology: Diachronic and synchronic perspectives*. Vol. 4. Oxford: Oxford University Press.

Khalilova, Zaira. 2009. A grammar of Khwarshi. PhD dissertation, LOT Netherlands Graduate School of Linguistics.

Kibrik, Aleksandr E. 1994. Khinalug. *The indigenous languages of the Caucasus* 4: 367-406.

Kibrik, Aleksandr E., Sandro V. Kodzasov & Irina A. Muravyova. 2004. Alutor Grammar. In Megumi Kurebito (Ed.), *Language and Folklore of the Alutor People*: 203-337.

Kießling, Roland. 1994. *Eine Grammatik des Burunge*. Vol. 13. Cologne: Köppe.

Killian, Don 2015. *Topics in Uduk Phonology and Morphosyntax*. PhD dissertation, Helsinki: University of Helsinki.

Kimball, Geoffrey D. 1985. *A descriptive grammar of Koasati*. PhD dissertation. New Orleans: Tulane University.

King, Copland. 1927. *Grammar and dictionary of the Binandere language, Mamba River, Northern Division, Papua*. With an introduction by Sidney H. Ray. Sydney: D. S. Ford.

Kirby, Simon; Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105, 31: 10681-10686.

Klein, Devorah E. & Gregory L. Murphy. 2001. The representation of polysemous words. *Journal of Memory and Language* 45.2: 259-282.

Koontz-Garboden, Andrew. 2016. Thoughts on diagnosing morphomicity: A case study from Ulwa. In Ana R. Luís & Ricardo Bermúdez-Otero (Eds.) *The morphome debate*: 309-40. Oxford: Oxford University Press.

Kornfilt, Jaklin. 2013. *Turkish*. London: Routledge.

Kossmann, Maarten. 2015. Contact-induced change. In Matthew Baerman (Ed.) *The Oxford Handbook of Inflection*. Oxford: Oxford University Press.

Krishan, Shree. 2001. A sketch of Chaudangsi grammar. *New research on Zhangzhung and related Himalayan languages*: 401-424.

Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic language of Zaire*. Cologne: Köppe.

Lahaussais, Aimée. 2002. *Aspects of the grammar of Thulung Rai: an endangered Himalayan language*. PhD dissertation, University of California Berkeley.

Larsen, Bud. 1977. *Orokaiva grammar*. Unpublished Ms. 87pp.

Leiss, Elisabeth. 1997. Synkretismus und Natürlichkeit. *Folia Linguistica* 31: 133-160.

Lepori, Antonio. 2001. *Gramàtica sarda po is campadinesus*. Quartu Sant'Elena: Edizioni C.R.

Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67.3: 475-509.

Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. *Morphology in Toulouse*: 133-153.

Lockwood, William Burley. 1977. *An introduction to modern Faroese*. Copenhagen: Munksgaard.

Loporcaro, Michele; Nadja Kägi & Francesco Gardani. 2018. Morfomi sommersi in Pantesco o dell'arte di arrangiarsi in morfologia. In Annamaria Chilà & Alessandro De Angelis (Eds.), *Capitoli di morfosintassi delle varietà romanze d'Italia: teoria e dati empirici*. Palermo: Centro di studi filologici e linguistici siciliani.

Lupardus, Karen Jacque. 1982. *The Language of the Alabama Indians*. Ph.D. dissertation, University of Kansas.

Luís, Ana R. & Ricardo Bermúdez-Otero (Eds). 2016. *The morpheme debate*. Oxford: Oxford University Press.

Lynch, John; Malcolm Ross & Terry Crowley. 2002. *The oceanic languages*. Vol. 1. Psychology Press.

Mac Congáil, Nollaig. 2004. *Irish grammar book*. Inverin: Cló Iar-Chonnachta.

Maiden, Martin. 1992. Irregularity as a determinant of morphological change. *Journal of Linguistics* 28, 2: 285-312.

Maiden, Martin. 2001. A strange affinity: 'perfecto y tiempos afines'. *Bulletin of Hispanic Studies* 78, no. 4: 441-464.

Maiden, Martin. 2005. Morphological autonomy and diachrony. *Yearbook of morphology* 2004: 137-175.

Maiden, Martin. 2011a. Morphological persistence. In Martin Maiden, John Charles Smith & Adam Ledgeway (Eds.), *Cambridge history of the romance languages*: 155-215. Cambridge: Cambridge University Press.

Maiden, Martin. 2011b. Morphemes and 'stress-conditioned allomorphy' in Romansh. In Maria Goldbach, Marc-Olivier Hinzelin, Martin Maiden & John Charles Smith (Eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*: 36–50. Oxford: Oxford University Press.

Maiden, Martin. 2011c. Morphophonological innovation. In Martin Maiden, John Charles Smith & Adam Ledgeway (Eds.), *Cambridge history of the romance languages*: 216-267. Cambridge: Cambridge University Press.

Maiden, Martin. 2012. A paradox? The morphological history of the Romance present. In S. Gaglia & M.-O. Hinzelin (Eds.), *Inflection and word formation in Romance languages*. Amsterdam: John Benjamins.

Maiden, Martin. 2013. The Latin 'third stem' and its Romance descendants. *Diachronica* 30.4: 492-530.

Maiden, Martin. 2016. Morphemes in diachrony. In Ana R. Luís & Ricardo Bermúdez-Otero. (Eds.), *The Morpheme Debate*. Oxford: Oxford University Press.

Maiden, Martin. 2017. Romansh allomorphy (Again!). In Claire Bower, Laurence Horn and Raffaella Zanuttini (Eds.), *On Looking into Words (and Beyond): Structures, Relations, Analyses*: 189–210. Berlin: Language Science Press.

Maiden, Martin. 2018a. New thoughts on an old puzzle: The Italian alternation type *dissi, dicesti; feci, facesti*. *Revue Romane* 53, no. 2: 217-260.

Maiden, Martin. 2018b. *The Romance Verb: Morphomic Structure and Diachrony*. Oxford: Oxford University Press.

Maiden, Martin & Paul O'Neill. 2010. Morphomic Defectiveness. In M. Baerman, G. Corbett, & D. Brown (Eds.), *Defective Paradigms: Missing Forms and What They Tell Us*: 103-124.

Maiden, Martin & Cecilia Robustelli. 2014. *A reference grammar of modern Italian*. London: Routledge.

Maiden, Martin; John Charles Smith; Silvio Cruschina; Marc-Olivier Hinzelin; & Maria Goldbach. 2010. *Oxford Online Database of Romance Verb Morphology*. University of Oxford. Online at <http://romverbmorph.clp.ox.ac.uk/>

Malau, Catriona. 2016. *A grammar of Vurës, Vanuatu*. Vol. 651. Berlin: de Gruyter.

Malkiel, Yakov. 1974. New Problems in Romance Interfixation (I): The Velar Insert in the Present Tense (with an Excursus on "-zer/-zir" Verbs). *Romance Philology* 27.3: 304-355.

Mangel, Marc & Samaniego, Francisco. 1984. Abraham Wald's work on aircraft survivability. *Journal of the American Statistical Association*. 79, 386: 259–267.

Mansfield, John. n.d. Murrinhpatha finite verb inflection paradigms. Unpublished Ms. Online at http://langwidj.org/Murrinhpatha-inflection/Intersecting-formatives-MP_Appendix.pdf

Marks, Doreen. 1974. Kosena grammar. Ukarumpa: SIL.

Marmion, Douglas Edric. 2010. Topics in the phonology and morphology of Wutung. PhD dissertation, Australian National University.

Matras, Yaron. 2015. Why is the borrowing of inflectional morphology dis-preferred? In Francesco Gardani, Peter Arkadiev & Nino Amiridze (Eds.), *Borrowed Morphology*: 47–80. (*Language Contact and Bilingualism* 8). Berlin: De Gruyter.

Matthews, Peter H. 1965. The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics* 1.2: 139-171.

Matthews, Peter H. 1991. *Morphology*. Cambridge: Cambridge University Press.

McCreight, Katherine & Catherine V. Chvany. 1991. Geometric representation of paradigms in a modular theory of grammar. In *Paradigms: The economy of inflection*: 91-112.

McPherson, Laura. 2013. *A grammar of Tommo So*. Vol. 62. Walter de Gruyter.

Meier-Brügger, Michael. 2013. *Indo-European Linguistics*. Walter de Gruyter.

Mel'čuk, Igor A. 2002. Towards a formal concept 'zero linguistic sign'. Applications in typology. In Bendjaballah, Sabrina; Wolfgang U. Dressler; Oskar E. Pfeiffer & Maria D. Voeikova (Eds.), *Morphology 2000*: 241-258. Amsterdam: Benjamins.

Merlan, Francesca C. 1994. *A Grammar of Wardaman*. Berlin: Mouton de Gruyter

Merrifield, William R. 1965. *Palantla Chinantec grammar*. Ph.D. thesis. Cornell University.

Meul, Claire. 2010. The intra-paradigmatic distribution of the infix -I/ESC- from Latin to Modern Romance: morphomic patterning and beyond. *Morphology* 20.1: 1-40.

Michailovsky, Boyd. 1975. Notes on the Kiranti verb (East Nepal). *Linguistics of the Tibeto-Burman Area* 2,2: 183-218.

Michailovsky, Boyd. 1985. Tibeto-burman dental suffixes: Evidence from Limbu (Nepal). *Linguistics of the Sino-Tibetan area: The state of the art. Papers presented to Paul K. Benedict for his 71st birthday*: 363-375.

Michailovsky, Boyd. 1988. *La langue hayu*. Éd. du Centre national de la recherche scientifique.

Montermini, Fabio & Olivier Bonami. 2013. Stem spaces and predictability in verbal inflection. *Lingue e linguaggio* 12, 2: 171-190.

Morin, Yves-Charles. 1988. Morphological Conditioning in Phonologically Transparent Processes: Evidence from the Evolution of Vowel Reduction in Vinzelles Occitan. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 33.4: 431-442.

Morrison, Michelle Elizabeth. 2011. *A reference grammar of Bena*. PhD dissertation, Rice University.

Mosonyi, Esteban Emilio & Jorge Carlos Mosonyi. 2000. *Manual de lenguas indígenas de Venezuela*. Caracas: Fundación Bigott.

Mous, Maarten. 1992. *A Grammar of Iraqw*. PhD Dissertation. University of Leiden.

Müller, Gereon. 2005. Syncretism and iconicity in Icelandic noun declensions: a distributed morphology approach. In *Yearbook of morphology 2004*: 229–271. Berlin: Springer.

Murane, Elizabeth. 1974. *Daga grammar: From morpheme to discourse*. Norman, OK: Summer Institute of Linguistics.

Naba, Jean-Claude. 1994. *Le Gulmancema: essai de systématisation; phonologie-tonologie-morphophonologie nominale-système verbal*. PhD Diss. Köppe Köln.

Nagore Lain, Francho. 1986. *El aragonés de Panticosa: gramática*. Huesca: Instituto de Estudios Altoaragoneses.

Nedjalkov, Vladimir P., and Galina A. Otaina. 2013. *A syntax of the Nivkh language: The Amur dialect*. Amsterdam: John Benjamins.

Nevins, Andrew, Cilene Rodrigues & Kevin Tang. 2015. The rise and fall of the L-shaped morpheme: diachronic and experimental studies. *Probus* 27,1: 101-155.

Newmark, Leonard; Philip Hubbard & Peter Prifti. 1982. *Standard Albanian, A reference grammar for students*. Stanford: Stanford University Press.

Nikolaeva, Irina. 1999. *Ostyak*. Munich: Lincom.

Noonan, Michael. 2011. *A grammar of Lango*. Vol. 7. Berlin: Walter de Gruyter.

Nordlinger, Rachel. 1998. *A Grammar of Wampaya, Northern Territory, Australia*. Canberra: Pacific Linguistics.

Nordlinger, Rachel. 2015. Inflection in Murrinh-Patha. In Matthew Baerman (Ed.). *The Oxford handbook of inflection*. Oxford: Oxford University Press.

Noreen, Adolf. 1904. *Altschwedische grammatik*. Tübingen: Niemeyer.

Novelli, Bruno. 1985. *A Grammar of the Karimojong Language*. Berlin: Dietrich Reimer Verlag.

Nübling, Damaris. 2000. *Prinzipien der Irregularisierung. Eine kontrastive Analyse von zehn Verben in zehn germanischen Sprachen*. Tübingen: Niemeyer.

O'Neill, Paul. 2011. *The Ibero-Romance Verb: allomorphy and the notion of the morpheme*. PhD dissertation.

O'Neill, Paul. 2013. The morpheme and morphosyntactic/semantic features. In Cruschina, Silvio; Martin Maiden & John Charles Smith (Eds.), *The Boundaries of Pure Morphology*. 221-46. Oxford: Oxford University Press.

O'Neill, Paul. 2014. The morpheme in constructive and abstractive models of morphology. *Morphology* 24.1: 25-70.

O'Neill, Paul. 2015. The Origin and Spread of Velar Allomorphy in the Spanish Verb. *Bulletin of Hispanic Studies* 92(5):489-518.

O'Neill, Paul. 2018. Velar allomorphy in Ibero-Romance: Roots, endings and clashes of morphemes. In Miriam Bouzouita, Ioanna Sitaridou & Enrique Pato (Eds.) *Studies in historical Ibero-Romance morpho-syntax: a descriptive and prescriptive analysis*: 13-46. Amsterdam: John Benjamins.

Obata, Kazuko. 2003. *A grammar of Bilua: A Papuan language of the Solomon Islands*. Research School of Pacific Studies. Australian National University.

Olson, Michael L. 1973. Barai sentence structure and embedding. Santa Ana, CA: Summer Institute of Linguistics.

Operstein, Natalie. 2002. First-person plural and the aspect morphology of Zapotec. In Proceedings of the 5th Annual Workshop on the American Indigenous Languages, Santa Barbara Papers in Linguistics, pp. 53-64. 2002.

Opgenort, Jean Robert. 2004. A grammar of Wambule: grammar, lexicon, texts, and cultural survey of a Kiranti tribe of eastern Nepal. Leiden: Brill.

Opgenort, Jean Robert. 2005. A grammar of Jero: With a historical comparative study of the Kiranti languages. Leiden: Brill.

Osam, E. Kweku. 1994. Aspects of Akan grammar: A functionalist perspective. PhD dissertation, University of Oregon.

Ouoba, Bendi Benoît. 1982. Description systématique du gulmancema: phonologie, lexicologie, syntaxe. Doctoral dissertation, Université de la Sorbonne Nouvelle (Paris 3).

Overall, Simon E. 2007. A grammar of Aguaruna. PhD dissertation, La Trobe University.

Overall, Simon E. 2017. *A Grammar of Aguaruna (liniá Chicham)*. Berlin: de Gruyter.

Palancar, Enrique. 2001. Emergent markers: The relation of instruments, causes, and agents in grammatical constructions. STUF-Language Typology and Universals 54.4: 365-384.

Panman, Otto. 1982. Homonymy and polysemy. *Lingua* 58.1-2: 105-136.

Parlier, James. 1964. Managalasi verb inflection. *Te Reo* 7: 28-35.

Paster, Mary. 2010. The verbal morphology and phonology of Asante Twi. *Studies in African Linguistics* 39.1: 77-120.

Pato, Enrique & Paul O'Neill. 2013. Los gerundios analógicos en la historia del español (e iberorromance). *Nueva Revista de Filología Hispánica* 61, 1: 1-27.

Paudyal, Netra Prasad. 2013. *Aspects of Chintang Syntax*. PhD dissertation, University of Zurich.

Pertsova, Katya. 2007. *Learning form-meaning mappings in presence of homonymy: a linguistically motivated model of learning inflection*. Doctoral dissertation, UCLA.

Pertsova, Katya. 2011. Grounding systematic syncretism in learning. *Linguistic Inquiry* 42, 2: 225-266.

Peña, Jaime. 2016. *A grammar of Wampis*. PhD Dissertation, University of Oregon.

Pike Kenneth L. 1948. *Tone Languages*. Ann Arbor: University of Michigan Press.

Pires, Nadia N. 1992. *Estudo da gramática da língua Jeoromitxi (Jabuti)*. MA Thesis, Universidade Estadual de Campinas.

Pittau, Massimo. 1972. *Grammatica del sardo-nuorese, il più conservativo dei parlari neolatini*. Bologna: Pàtron.

Plank, Frans & Elena Filimonova. 2000. The Universals Archive: A Brief Introduction for Prospective Users. *STUF-Language Typology and Universals* 53.1: 109-123.

Pylkkänen, Liina; Rodolfo Llinás; & Gregory L. Murphy. 2006. The representation of polysemy: MEG evidence. *Journal of cognitive neuroscience* 18,1: 97-109.

Quesada, Juan Diego. 2000. *A grammar of Teribe*. Vol. 36. Munich: Lincom.

Quint, Nicolas. 1998. *Le Parler occitan alpin du Pays de Seyne*. Paris: L'Harmattan.

RAE-ASALE (= Real Academia Española / Asociación de Academias de la Lengua Española). 2009. Nueva Gramática de la Lengua Española. Madrid: Espasa-Calpe.

Rapold, Christian. 2006. Towards a grammar of Benchnon. PhD thesis. University of Leiden.

Rask, Rasmus. 1776. A grammar of the Icelandic or Old Norse tongue. Amsterdam: John Benjamins.

Real Academia Española. n.d. CORPES XXI: Corpus del Español del Siglo XXI. <http://www.rae.es/recursos/banco-de-datos/corpes-xxi> (Accessed December 2016).

Rensch, Calvin R. 1989. An etymological dictionary of the Chinantec languages. Summer Institute of Linguistics.

Rießler, Michael. Forthcoming. Kildin Saami. In Marianne Bakró-Nagy, Johanna Laakso & Elena Skribnik (Eds.), Oxford Guide to the Uralic Languages. Oxford: Oxford University Press.

Roberts, John R. 1987. Amele. London: Croom Helm.

Ronjat, Jules. 1930. Grammaire istorique des parlers provençaux modernes. Montpellier: Société des langues romanes.

Ross, Malcolm. 1980. Some elements of Vanimo, a New Guinea tone language. Pacific Linguistics. Series A. Occasional Papers 56.

Ross, Malcolm. 2002. Kele. The Oceanic languages: 123-147.

Round, Erich R. 2011. Morphemes as a level of representation capture unity of exponence across the inflection-derivation divide. *Linguistica* 51.1: 217.

Round, Erich R. 2013. Rhizomorphemes, meromorphemes, and metamorphemes. In Matthew Baerman, Dunstan Brown & Greville Corbett (Eds.), *Understanding and Measuring Morphological Complexity*. 29-52. Oxford: Oxford University Press.

Round, Erich R. 2016. Kayardild inflectional morphotactics is morphomic. In Ana R. Luís & Ricardo Bermúdez-Otero (Eds.), *The Morpheme Debate*: 228-247. Oxford: Oxford University Press.

Round, Erich R., & Greville G. Corbett. 2017. The theory of feature systems: One feature versus two for Kayardild tense-aspect-mood. *Morphology* 27.1: 21-75.

Rubin, Aaron. 2010. *The Mehri Language of Oman*. Leiden: Brill.

Rueter, Jack M. 2010. Adnominal person in the morphological system of Erzya. Helsinki: Société finno-ougrienne.

Rupp, James E. 1989. *Lealao Chinantec Syntax*. No. 88. Summer Inst of Linguistics.

Rupp, James. 1996. *Diccionario chinanteco de San Juan Lealao, Oaxaca*. Tucson: SIL.

Saad, George Michel. 2014. *A grammar sketch of Shuar: With focus on the verb phrase*. MA thesis. Radboud University Nijmegen.

Sachse, Frauke. 2010. *Reconstructive description of eighteenth-century Xinka grammar*. Utrecht: LOT-Netherlands Graduate School of Linguistics.

Saeed, John. 1999. *Somali*. Amsterdam: John Benjamins.

Sasse, Hans Jurgen. 1976. Dasenech. In M. Lionel Bender (Ed.), *The non-Semitic languages of Ethiopia*: 196-221. East Lansing: African Studies Center, Michigan State University.

Sauerland, Uli & Jonathan Bobaljik. 2013. Syncretism distribution modeling: Accidental homophony as a random event. Proceedings of GLOW in Asia IX 2012.

Saura Rami, José Antonio. 2003. Elementos de fonética y morfosintaxis benasquesas. Zaragoza: Gara d'Edicions-Institución «Fernando el Católico».

Saxena, Anju. 1992. Finite verb morphology in Tibeto-Kinnauri. PhD dissertation. University of Oregon.

Schackow, Diana. 2016. A grammar of Yakkha. Berlin: Language Science Press.

Schalchli, Gauvain; & Gilles Boyé. 2017. Realistic paradigms for derivational morphology. Talk delivered at ParadigMo, Toulouse.

Schanen, François. 2004. Parlons luxembourgeois: langue et culture linguistique d'un petit pays au coeur de l'Europe. Paris: L'Harmattan.

Schmidtke-Bode, Karsten; Natalia Levshina; Susanne Maria Michaelis & Ilja Seržant (Eds.). 2019. Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence. Berlin: Language Science Press.

Schulze, Wolfgang. 1997. Tsakhur. Vol. 133. Munich: Lincom.

Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91.3: 511-532.

Sharma, Narayan Prasad. 2014. Morphosyntax of Puma, a Tibeto-Burman language of Nepal. PhD dissertation. SOAS.

Sharma, Suhnu Ram. 2001a. A Sketch of Byangsi Grammar. In Yasuhiko Nagano & Randy J. LaPolla (Eds.), *New Research on Zhangzhung and Related Himalayan Languages*. Osaka: National Museum of Ethnology.

Sharma, Suhnu Ram. 2001b. A sketch of Rongpo grammar. In Yasuhino Nagano & Randy J. LaPolla (Eds.), *New research on Zhangzhung and related Himalayan languages*. Osaka: National Museum of Ethnology.

Siewierska, Anna. 2010. From third plural to passive: incipient, emergent and established passives. *Diachronica* 27, 1: 73-109.

Silva, Wilson de Lima. 2012. A descriptive grammar of Desano. Doctoral dissertation. University of Utah.

Silvey, Catriona; Simon Kirby; & Kenny Smith. 2015. Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive science* 39, 1: 212-226.

Silzer, Peter James. 1983. Ambai: An Austronesian Language of Irian Jaya, Indonesia. PhD dissertation. Australian National University.

Slobin, Dan I. & Ayhan A. Aksu. 1982. Tense, aspect and modality in the use of the Turkish evidential. In Paul J. Hopper (Ed.), *Tense-aspect: Between semantics and pragmatics*: 185-200. Amsterdam: John Benjamins.

Smith, John Charles. 2013. The morpheme as a gradient phenomenon: evidence from Romance. In S. Cruschina, M. Maiden, & J. C. Smith (Eds.), *The boundaries of pure morphology*: 247–261. Oxford: Oxford University Press.

Smye, Graham. 2004.. A grammar of Anufo. Institute of African Studies, University of Ghana.

So-Hartmann, Helga. 2009. A descriptive grammar of Daai Chin. Berkeley: Sino-Tibetan Etymological Dictionary and Thesaurus Project.

de Sousa, Hilario. 2006. The Menngwa Dla Language of New Guinea. PhD dissertation. University of Sidney.

Speck, Charles H. 1984. The phonology of the Texmelucan Zapotec verb. *International Journal of American Linguistics* 50, no. 2: 139-164.

Spencer, Andrew. 2012. Identifying stems. *Word Structure* 5, 1: 88-108.

Spencer, Andrew. 2016. Stems, the morpheme and meaning-bearing inflection. In Ana Luís & Ricardo Bermúdez-Otero. (Eds.), *The Morpheme Debate*. Oxford: Oxford University Press.

Spencer, Katharine. 2008. Kwomtari grammar essentials. In Murray Honsberger, Carol Honsberger & Ian Tupper (Eds.) *Kwomtari phonology and grammar essentials*: 53-183. Ukarumpa: SIL-PNG Academic Publications.

Steriade, Donca. 2016. The morpheme vs. similarity-based syncretism: Latin t-stem derivatives. In Ana R. Luís & Ricardo Bermúdez-Otero. (Eds.). *The Morpheme Debate*. Oxford: Oxford University Press.

Sterner, Joyce K. 1987. Sobei verb morphology reanalyzed to reflect POC studies. *Oceanic Linguistics* 26.1/2: 30-54.

Stump, Gregory. 2001. *Inflectional Morphology: a Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory. 2006. Heterocclisis and paradigm linkage. *Language* 82, 2: 279-322.

Stump, Gregory. 2014. Polyfunctionality and inflectional economy. *LiLT (Linguistic Issues in Language Technology)* 11.

Stump, Gregory. 2015. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface*. Cambridge: Cambridge University Press.

Stump, Gregory. 2016. Morphomic categories and the realization of morphosyntactic properties. In Ana Luís & Ricardo Bermúdez-Otero. (Eds.), *The Morpheme Debate*. Oxford: Oxford University Press.

Stump, Gregory; & Raphael A. Finkel. 2013. *Morphological typology: From word to paradigm*. Vol. 138. Cambridge: Cambridge University Press.

Sturtevant, Edgard H. 1947. *An Introduction to Linguistic Science*. New Haven: Yale University Press.

Suárez, Jorge A. 1983. *La Lengua Tlapaneca de Malinaltepec*. México: Instituto de Investigaciones Filológicas, Universidad Nacional Autónoma de México.

Söderholm, Eira. 2017. *Kvensk grammatikk*. Cappelen Damm Akademisk.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5.2: 183-207.

Thurneysen, Rudolf. 1887. Lateinischer lautwandel. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 28, 1/2: 154-162.

Tosco, Mauro. 2001. *The Dhaasanac Language*. Cologne: Köppe.

Trommer, Jochen. 2016. A postsyntactic morpheme cookbook. In Daniel Siddiqi & Heidi Harley (Eds.), *Morphological metatheory*: 59–93. Amsterdam: John Benjamins.

Trommer, Jochen & Sebastian Bank. 2017. Inflectional learning as local optimization. *Morphology* 27, 3: 383-422.

Tuite, Kevin. 1994. Aorist and pseudo-aorist for Svan atelic verbs. *NSL 7: Linguistic studies in the non-Slavic languages of the Commonwealth of Independent States and the Baltic Republics*: 319-340.

Tuite, Kevin. 1995. Svan and its speakers. Unpublished Ms. Online at <http://mapageweb.umontreal.ca/tuiteki/publications/Tuite-1997-SvanSpeakers.pdf>.

Tuite, Kevin. 1998. A short descriptive grammar of the Svan language. Université de Montréal.

Uhlenbeck, Eugenius M. 1952. The study of wordclasses in Javanese. *Lingua* 3: 322-354.

Velie, Daniel & Virginia Velie. 1981. *Vocabulario orejón*. Lima: Ministerio de Educación, Instituto Lingüístico de Verano.

Völlmin, Sascha. 2017. *Towards a Grammar of Gumer-Phonology and Morphology of a Western Gurage Variety*. PhD Dissertation, University of Zurich.

Voorhoeve, Clemens L. 1975. Central and Western Trans-New Guinea Phylum languages. In Stephen Wurm (Ed.), *Papuan Languages and the New Guinea Linguistic Scene*, New Guinea Area Languages and Language Study 1: 345-459. Australian National University.

van der Voort, Hein. 2007. Proto-Jabutí: um primeiro passo na reconstrução da língua ancestral dos Arikapú e Djeoromitxí. *Bol. Mus. Para. Emílio Goeldi. Ciências Humanas* 2(2): 133-168. Belém.

Waag, Christine. 2010. *The Fur verb and its context*. Cologne: Rüdiger Köppe.

Walsh, Michael. 1976. *The Murinypata language of north-west Australia*. PhD dissertation. Australian National University.

Watters, David. 2006. *Notes on Kusunda grammar: A language isolate of Nepal*. Himalayan Linguistics Archive 3.

West, Birdie. 1980. *Gramática popular del Tucano*. Bogotá: Instituto Lingüístico de Verano.

West, Dorothy. 1973. *Wojokeso sentence, paragraph, and discourse analysis*. Australian National University.

Wheeler, Alva. 1970. *Grammar of the Siona language, Colombia, South America*. PhD dissertation. University of California at Berkeley.

Wilbur, Joshua. 2014. *A grammar of Pite Saami*. Berlin: Language Science Press.

Van der Wilden, Jaap. 1976. Simplicity and detail in Kemtuk predication. *Irian* 5, no. 2: 59-84.

Willis, Christina Marie. 2007. *A descriptive grammar of Darma: an endangered Tibeto-Burman language*. PhD Dissertation, University of Texas at Austin.

Wilson, Darryl. 1974. *Suena Grammar*. *Workpapers in Papua New Guinea Languages* 8, Summer Institute of Linguistics.

Winkler, Eberhard. 2001. *Udmurt*. Munich: Lincom.

Wonderly, William L., Lorna F. Gibson & Paul L. Kirk. 1954. Number in Kiowa: Nouns, demonstratives, and adjectives. *International Journal of American Linguistics* 20, 1: 1-7.

Wunderlich, Dieter. 2012. Polarity and constraints on paradigmatic distinctness. In Jochen Trommer (Ed.) *The morphology and phonology of exponence*: 121-159. Oxford: Oxford University Press.

Wurzel, Wolfgang U. 1980. Ways of morphologizing phonological rules. *Historical morphology* 62: 443-462

Wurzel, Wolfgang U. 1989. *Inflectional morphology and naturalness*. Amsterdam: Springer.

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic inquiry* 19, 1: 65-100.

Yoshioka, Noboru. 2012. A reference grammar of Eastern Burushaski. PhD thesis, Tokyo University of Foreign Studies.

Young, Robert A. 1964. The primary verb in Bena-bena. In Alan R. Pence (Ed.) *Verb studies in five New Guinea languages*: 45-83. Summer Institute of Linguistics of the University of Oklahoma.

Zhirkov, Lev I. 1955. *Lakskij jazyk: Fonetika i morfologija*. [The Lak Language: Phonetics and Morphology.] Moscow: AN SSSR.

Zipf, George Kingsley. 1935. *The psycho-biology of language*. Boston: Houghton Mifflin Company

Zwicky, Arnold M. 1991. Systematic versus accidental phonological identity. *Paradigms: the economy of inflection*: 113-131.

Zoëga, Geir T. 1910. *A Concise Dictionary of Old Icelandic*. Oxford: Clarendon Press

Zörner, Lotte. 1997. *Il pagotto. Dialetto dell'Alpago: descrizione fonologica, storico-fonetica e morfologica*. Padova: Unipress.