The Typological Diversity of Morphomes

A Cross-Linguistic Study of Unnatural Morphology

BORJA HERCE

OXFORD

The Typological Diversity of Morphomes

The Typological Diversity of Morphomes

A Cross-Linguistic Study of Unnatural Morphology

BORJA HERCE

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2022921887

ISBN 978–0–19–286459–8

DOI: 10.1093/oso/9780192864598.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

Preface

This book constitutes the first typologically oriented monograph on morphomes, which is the term given to systematic morphological identities, usually within inflectional paradigms, that do not map onto syntactic or semantic natural classes like 'plural', 'past', 'third-person singular'. Chapter 1 discusses history, terminology, and the relevant literature on this unusual phenomenon, while Chapter 2 contains all necessary clarifications with respect to the identification and definition of morphomes, and their links with related phenomena like syncretism, morphophonology, homophony, defectiveness ... and theoretical notions like blocking, segmentation, and economy. Diachrony then takes centre stage, as Chapter 3 presents the different ways in which morphomic structures have been observed to emerge, change, and disappear from a language. Chapter 4 constitutes the core of this book and presents a database with 120 morphomes found across 79 languages from around the world. All these structures are presented in great detail, along with their diachrony if known. On the basis of the synchronic variation across morphomes, nine logically independent variables (and some additional ones) have been identified in the spirit of Multivariate Typology as the most relevant to describing these structures in the most fine-grained detail. These variables have been operationalized into quantitative measures; and, after establishing the values they take in all morphomes in the database, statistical analysis has been undertaken to spot some trends, correlations, and dependencies between them which are subsequently discussed.

Various findings, relevant to both proponents and detractors of Autonomous Morphology, have emerged. One is that Romance stem alternations, which have monopolized research to date, are not particularly representative of the phenomenon as a whole. Another relevant finding is that various unnatural patterns (SG+3PL, 1SG+3, 2+1PL, PL+1SG, PL+2SG, PL+3SG, and SG+1PL) are present in several genetically and geographically unrelated languages. This has theoretical implications regarding the gradient, rather than dichotomic, nature of naturalness (with a preference for more natural patterns observed even among morphomes) The database, available online, is also expected to provide morphologists and typologists with a tool to explore properties and correlations unrelated to Autonomous Morphology, for example the nature of the stem–affix distinction, the tradeoff between the lexical and grammatical informativity of morphs, or the distribution of information on the word.

Acknowledgement

This book would not have been possible without the guidance and support from the colleagues and supervisors I have had both during my PhD in Surrey (most notably Matthew Baerman) and during my postdoc in Zürich (most notably Balthasar Bickel). To them and all others I express my deepest gratitude. I am also grateful to SNF for funding the open access to this book (grant n. 10BP12_214604).

At a personal level, I could not have done this without the love and support from my family and friends. I dedicate this book in particular to Sarah and mi abuelita Josefa.

List of abbreviations

1	First person
2	Second person
3	Third person
A	Agent
ABESS	Abessive
ABL	Ablative
ABS	Absolutive
ACC	Accusative
ADJ	Adjective
ADV	Adverbial
AOR	Aorist
AP	Antipassive
APOD	Apodosis
C	Consonant
COM	Comitative
COND	Conditional
COORD	Coordinate
CORDE	Corpus Diacrónico del Español
CORPESXXI	Corpus Español del Siglo XXI
DAT	Dative
DS	Different Subject
DU	Dual
ELAT	Elative
ERG	Ergative
ESS	Essive
EVID	Evidential
EXCL	Exclusive
F	Feminine
FUT	Future
GEN	Genitive
GER	Gerund
HIAF	High affectedness
HUM	Human
ILL	Illative
IMP	Imperative
INCL	Inclusive
IND	Indicative
INESS	Inessive

INF	Infinitive
INS	Instrumental
INTER	Interrogative
IPF	Imperfect
IPFV	Imperfective
JUSS	Jussive
LAT	Lative
LOC	Locative
M	Masculine
MC	Morphosyntactic coherence
N	Neuter
NF	Non-feminine
NFUT	Non-future
NHUM	Non-human
NOM	Nominative
NSG	Non-singular
NSIB	Non-sibling
OBJ	Object
OHG	Old High German
P	Patient
PC	Paucal
PFV	Perfective
PL	Plural
PNG	Papua New Guinea
PRET	Preterite
PRF	Perfect
PRS	Present
PYTA	Pretérito y Tiempos Afines
RAE	Real Academia Española
RECPAST	Recent past
SIB	Sibling
SBJ	Subject
SBJV	Subjunctive
SG	Singular
SS	Same subject
SUB	Subordinate
TAM	Tense Aspect Mood
TNG	Trans-New Guinea
TR	Trial
V	Vowel
VOC	Vocative

1
Introduction

1.1 Initial approximation and goals

The present monograph focuses on morphomes, understood as morphosyntactically unnatural sets of paradigm cells that systematically share (some of their) morphology. The concept was introduced by Aronoff (1994) and popularized by Maiden's (e.g. 2018b) research on the diachronic behaviour of stem alternations in Romance. In this family, morphomes have been extensively studied over the last years and have even been given names of their own.

The Spanish verb 'fit', for example (Table 1.1), has a dedicated stem in 1SG.IND+SBJV. The verb 'can', in turn, has a different stem in SG+3PL.[1] These stem alternation patterns are surprising, and problematic for many theoretical morphologists, because the sets of cells that share form (a stem in this case) do not constitute natural classes. The forms, therefore, are not coextensive with any meaning/value (e.g. 'present', 'subjunctive', 'first-person') nor with a combination of values (e.g. 'present subjunctive', 'first-person plural present'). Stems like *quep-* or *pued-*, thus, seem to be morphosyntactic arbitrary in their distribution.

Table 1.1 Two morphomic stem alternations in Spanish (partial paradigms)

	caber **'fit' illustrating the L-morphome**				***poder*** **'can' illustrating the N-morphome**			
	Present indicative		**Present subjunctive**		**Present indicative**		**Present 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

These morphological affinities appear to be, however, systematic within the language, since they are repeated in hundreds of verbs, often with different formal exponents. In addition, in diachrony, these sets of disparate paradigm cells show a strong tendency to behave *en bloc* in analogical changes. These facts are well known, in Romance, from the research of linguists like Malkiel (1974), Maiden (1992, 2005, 2018b), O'Neill (2013), and Esher (2015). Maiden (2018b: 18), has mentioned that their research could be used 'to speculate on the general significance of morphomic structures in ways that should be testable

[1] The 2SG imperative (not shown in Table 1.1) also forms part of the Romance N-morphome.

The Typological Diversity of Morphomes. Borja Herce, Oxford University Press. © Borja Herce (2023).
DOI: 10.1093/oso/9780192864598.003.0001

against a wider cross-linguistic range of data'. However, and in stark contrast to the wealth of research on Romance morphomes, very few studies have explored the phenomenon at length with data from other languages families.[2] As 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 monograph is set to fill.

A typological cross-linguistic approach to the morphome 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 and Corbett (2007: 115), it might well be that '[o]f all the aspects of language, morphology is the most language-specific and hence least generalizable'. Consequently, there will be important challenges to the extrapolation of meaningful principles.

Another very significant challenge is the nature of the morphome itself. It is usually assumed that the notion is dependent on the cognitive status of the morphological associations. That is, morphomes, to be truly morphomes, must 'constitute grammatical realities for speakers' (O'Neill 2014: 32). This, however, is very difficult to ascertain in practice. The evidence that is usually presented in relation to this may be diachronic (e.g. the preservation or replication of formal allegiances) or experimental (see e.g. Nevins et al. 2015). These types of evidence are regretfully unavailable for the vast majority of the world's languages. Furthermore, even when the data are available, their interpretation is hardly ever straightforward, and disagreements abound. For this reason, alternative diagnostics will have to be explored to approach the morphome as a coherent object of analysis in a synchronic typological study.

The main contribution of this book is, thus, a typological study of morphomes, with a cross-linguistically varied sample of 120 of them, 112 from outside Romance. These plentiful data will be at the service of research questions such as: what types of morphomic structures are possible? What are the synchronic properties of morphomes? 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 morphomes? What role do frequency or morphosyntactic features play in their evolution?

This research will also contribute to the broader discussion on the phenomenon's overall place in grammatical and morphological architecture. The diachronic and synchronic evidence gathered in this book will help to answer

[2] Notable exceptions, limited in their scope, include Round (2015) and Stump (2015: 128–40). Cross-linguistically oriented research has been conducted, of course, on notions that are not unrelated to the morphome, e.g. on 'morphologically stipulated patterns of syncretism' (see Baerman et al. 2005).

the fundamental questions of the morphome debate (Luís and Bermúdez-Otero 2016): What is the function of morphomes, if any? What makes them learnable? Is there a learning bias against morphomes? And ultimately: are there any empirical properties distinguishing morphomes from morphemes?

The answers to these questions and the outcomes of this research will be hopefully relevant not only to the field of autonomous morphology (and to theoretical morphology and typology more generally), but also to language description and documentation. Because they are very different from the functionally 'sensible' structures one usually expects and looks for, and because many (field) linguists, in my experience, have not even heard of the notion and term 'morphome', these structures are undoubtedly underreported and underdescribed in descriptive grammars. A cross-linguistic exploration and typologization of morphomes, like the one I present here will hopefully contribute to put the notion on the radar of many, and provide field linguists with the tools to describe these structures more thoroughly, more coherently, and using a more homogeneous terminology, which will in turn lay the ground for better and more efficient research in the future.

1.2 History

The term 'morphome' and the adjective 'morphomic' are, as I mentioned, relatively new additions to linguists' analytical toolkit. They were 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 heterogeneous phenomena (e.g. intraparadigmatic affinities, inflectional classes) 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 linguistic 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 morphomes, and Matthews (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 the recent surge in interest in measuring and understanding the role of predictive relations within the paradigm.

Another researcher whose work cast 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 these problematic facts of language before Aronoff (1994) was to extend the notion of the 'morpheme' (i.e. a form-cum-meaning sub-word unit) in ways that 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 was 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 (i.e. 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 morphome 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 also 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-iss-ons*, 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, there was already a tendency in some quarters (Hockett 1947; Harris 1942) 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.

In our journey back in time, therefore, we keep finding linguists who remained unconvinced that all morphology could be reduced to the principles of phonology or syntax/semantics. 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 declensions of conservative Indo-European languages) where various functions tend to be covered by a single formative. Bazell's 'phonomorpheme', thus, pre-dates Aronoff's 'morphome' by more than half a century but seems to have been inspired 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 morpheme[3] as a meaningful unit (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, 2011a), 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 processes of morphological analogical 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.

[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 endeavour) it is appropriate to point out that meaning has not always been part of the definition of 'morpheme'. Baudouin de Courtenay (1895 [1972]) 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).

Research around the morphome has been undertaken for over two decades now (and, arguably, with other labels, for much longer). However, there is still no consensus regarding some of the most fundamental questions such as, for example, whether morphomes have a learnability disadvantage over morphemes. Furthermore, although most research on morphomes has understandably come from morphologists that firmly believed that morphomes exist[4] and deserve attention, this is not an undisputed consensus either. Some linguists in the Morphome Debate (see e.g. Bermúdez-Otero and Luis 2016; Steriade 2016) have been very critical of the notion, worrying that morphomes may not constitute real categories for language users, but rather spurious or accidental formal resemblances. Some other concerns are more epistemological than ontological. Embick, for example, complains that the whole enterprise does 'not hold more theoretical interest than an enumeration of the facts' (Embick 2016: 299), and others like Koontz-Garboden (2016) lament the lack of positive diagnostics or empirical predictions in relation to the morphome.

Some solutions to these problems and disagreements may potentially come from quantitative research, for example, from experimental (Nevins et al. 2015) or artificial grammar learning (Saldana et al. 2022) approaches to morphology, as well as from the set-theoretic (Stump and Finkel 2013) or information-theoretic (Ackermann and Malouf 2013) exploration of paradigmatic relations. In the latter tradition, for example, Blevins (2016: 105) has proposed regarding morphomes as units of predictive value. Various other research paradigms and concepts, like 'stem spaces' (Boyé 2000, Boyé and Cabredo-Hofherr 2006, Montermini and Bonami 2013), 'niches' (Lindsay and Aronoff 2013), or 'No-Blur' (Carstairs-McCarthy 1994), also relate to the morphome in ways which are not always entirely appreciated or discussed. It will not be the focus of this book to spell out and reflect on all such connections. Let it suffice to point out here that reference to all this literature and notions and many others would be needed to present a complete picture of contemporary 'morphomics'.

1.3 Terminology

Despite the increasing appearance of the term in linguistic literature, the concept of the morphome is notoriously confusing. The noun 'morphome' and its adjectival derivation 'morphomic' have been used in the literature to refer to various linguistic objects such as meaningless stems, unnatural sets of paradigm cells, inflection classes (for a more exhaustive survey of the different uses see O'Neill

[4] It probably will not surprise anybody if I advance already here that my answer to that existence question will be positive too or else this book would not exist. I consider, however, that the existence of morphomes 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 here.

2011: 44 and O'Neill 2013: 221). These objects' only common property, as far as I can see, is that they could all be regarded as autonomous morphological phenomena. In addition to these, 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, Bermúdez-Otero and Luís 2016, Koontz-Garboden 2016: 90). This polysemy constitutes sometimes a notable hindrance to successful reasoning and dialogue. Fortunately, some contributions have recently spotted the problem and have proposed terminological remedies to some of these polysemies.

Smith (2013), for example, distinguished between what he called 'class morphomes' (i.e. inflection classes) and 'paradigm-subset morphomes'. Yet another contribution to terminological clarification is Round (2015). In his attempt at distinguishing the various senses of the terms 'morphome' and 'morphomic' in the literature, Round coined the terms 'rhizomorphome' (for inflection classes), 'metamorphome' (for sets of paradigm cells characterized by common exponents) and 'meromorphome' (for the actual forms that reveal a metamorphome). Table 1.2 illustrates the referents of these terms with an example familiar from the Romance morphome literature.

Table 1.2 L-morphome in Spanish (shaded cells)

	venir **'come'**		***nacer*** **'be born'**		***caber*** **'fit'**	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	veng-o	veng-a	naθk-o	naθk-a	kep-o	kep-a
2SG	vien-es	veng-as	naθ-es	naθk-as	kab-es	kep-as
3SG	vien-e	veng-a	naθ-e	naθk-a	kab-e	kep-a
1PL	ven-imos	veng-amos	naθ-emos	naθk-amos	kab-emos	kep-amos
2PL	ven-is	veng-ajs	naθeis	naθk-ajs	kab-ejs	kep-ajs
3PL	vien-en	veng-an	naθ-en	naθk-an	kab-en	kep-an

The lexemes *venir* and *nacer,* for example, belong to two different rhizomorphomes by virtue of their inflecting in different ways (contrast e.g. *ven-imos* vs *nac-emos*). A rhizomorphome, thus, would be a set of lexemes that inflect in the same way. Much like gender, they are partitions of the lexicon. In contrast to gender, however, they are, by definition, partitions without extramorphological effects. Because, in my opinion, inflection classes are a phenomenon quite different from the other ones referred to by the term 'morphome', the two can be explored with relative independence from one another. This book, therefore, will only be concerned tangentially with inflection classes.

More subtle is the distinction between the other two notions in Round (2015). A metamorphome, represented in Table 1.2 by the renowned L-morphome, is a set of paradigm cells which behave, within a given lexeme, in the same way regarding some morphological aspect. This particular metamorphome in Spanish encompasses the 1SG present indicative and all the present subjunctive cells. However, 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 to the stem found in other cells (i.e. /ven/>/veng/, /naθ/>/naθk/). In the case of the verb *caber*, these cells have a weakly suppletive stem alternant (i.e. /kab/>/kep/).

Distinguishing between formal elements or operations (e.g. /g/ or 'add /g/'), and the set of morphosyntactic contexts where these apply (e.g. 1SG.PRS.IND +PRS.SBJV) is sometimes necessary for clear argumentation. These two senses are, however, two sides of the same coin. The unnatural set of contexts that share a morphological affinity could be termed 'metamorphome' while the term 'meromorphome' is used to denote the actual form(ative)s which revealed 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 systematic morphosyntactic distribution we would like to account for in some principled way. The question to be asked at this point is whether we need to distinguish terminologically between a form and its distribution. To the regret of some linguists (see Haspelmath 2020), the prevalent trend in morphological literature over the last decades has been to refer to both increasingly with the same term, so that the erstwhile notions of 'morph' (a unit of form) and 'sememe' (a unit of meaning) have been increasingly replaced by 'morpheme'. Most authors in the morphomic literature (e.g. Smith 2013 or Stump 2016: 175) have also 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] In addition, I believe that the possibilities for confusion of the two senses are very limited in practice (i.e. when used in context). A terminological distinction between mero- 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, or 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. More generally, distinguishing meta- and meromorphomes would introduce new jargon into an already atomized field, unnecessarily degrading the readability of morphomic research for outsiders and for future linguists, should the terms fall into disuse. I will consequently not adopt here Round's (2015) terminology and I will continue to use the traditional terms 'inflection class' to denote a set of lexemes that inflect in the same way, and 'morphome' to refer to unnatural but systematic affinities in the paradigm, both on their form and their meaning side.

[5] Sometimes, e.g. in the Kayardild case/tense markers that Round (2015) discussed, systematic morphological affinities can be found between formatives in different word classes. In these cases, 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 but has not yet been proposed as far as I know.

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 not only a linguistic phenomenon, but also a particular formalization of this phenomenon 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 many other sciences (see Section 2.1.1.5), the two should sometimes be terminologically distinguished. To give a parallel example, the term 'syncretism' usually refers to the 'thing in the language' regardless of its formalization. The possible ways of formalizing or theoretically analysing syncretism (e.g. as 'underspecification', with a 'rule of referral') are referred to by dedicated terms, which often prevents sloppiness in argumentation and misunderstandings.

Similarly to syncretism, thus, one can simply 'observe' recurrent elements of form in a language whose domains of use are not conjunctively definable (by some measure). We may call this as we please (e.g. 'unnatural syncretism', 'morphome', 'homophony') but this descriptive 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 have the same form. A terminological disambiguation would be, therefore, most welcome in this respect since, currently, 'morphome' and 'morphomic' denote both a morphological entity and a particular theoretical stance and formalization.

1.4 A working definition

Since this monograph is mostly empirically oriented, the term 'morphome' will be used here almost exclusively in its near-observational formal-identity descriptive sense and not to refer to a higher-level theoretical or formal 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: 60), for example, defines a morphome simply as 'a systematic morphological syncretism which does not define a (syntactically or semantically) natural class'.

This is the kind of wording which I consider most appropriate for a typological investigation.[6] A definition such as this one would make it possible for us to agree

[6] In the context of more theoretically oriented disquisitions, a different definition might well be called for. Spencer (2016: 210), for example, proposes: 'An expression E is morphomic$_{\varphi\mathrm{strict}}$ iff E does not consist of a pairing of a form and a (natural) class of grammatical properties (feature–value pairs);

on the (non)morphomic status of particular exponents, provided we had clear criteria for recognizing (i) syncretisms, as well as (ii) natural classes, and that we operationalized (iii) 'systematicity' in some way. Because a consensus on these is woefully lacking, I will address these notions next, briefly in the remaining of this section and more extensively in the coming Section 1.2.

The first of these notions, **syncretism,** is one with a long tradition. As a term, it has been widely adopted by morphologists. This does not mean, however, that its usage is well established. One can actually find completely antagonistic definitions of what syncretism is. For Haspelmath and 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 and 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 highlights the need of homogeneous terminology and of agreeing upon our definitions. I believe 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 use the term 'syncretism' to refer to any total or partial morphological identity between different values (e.g. PAST and ACC) or paradigm cells (e.g. 1PL.SBJV, and 3SG.IND).

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 whatsoever. In plain terms, a natural class is one which is coextensive with a value (e.g. SG) or conjunction of values (e.g. 1SG). Unlike most extant formalisms suggest and/or allow, however, naturalness constitutes a gradient dimension (see Herce 2020a).

In Table 1.3, pattern A is unmistakably natural because it can be captured with reference to a feature value 'SG', and B is usually considered unproblematic too, although it involves more than one feature value at the same time '1SG'. Pattern F is the furthest from a natural class and thus the most unmistakably morphomic.

Table 1.3 Some paradigmatic distributions ordered for their naturalness

	(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	■									■							■	

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'.

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. My initial approach here will not be to take a particular immutable feature structure as the standard to taxonomize individual cases as morphemes or morphomes. Instead, I will try to preserve some sensitivity to the scale of variation outlined in Table 1.3, and to the plausibility or implausibility of a natural-class analysis in concrete cases.[7]

Having clarified the notions of syncretism and natural class, for our earlier working definition to 'work' it should still be clarified what exactly is meant by **systematicity**. When studying morphomes (or most other linguistic phenomena for that matter) we would like to make sure somehow that we are analysing single units/categories of some sort, that is, generalizations that the language users spot and abide by, and not instances of mere homophony. As with (un)naturalness, distinguishing between the two is not trivial, as there are many ways to understand 'systematic' and its antonym 'accidental'. The terms could apply to a pattern's diachronic origin or evolution (e.g. evidence from analogical changes), or to its synchronic status in the language. Diagnostics for synchronic systematicity can be sought in a pattern's syntactic, formal, and distributional properties. Thus, one could look at some forms' ability to resolve conflicting feature value requirements, to language users' behaviour in wug tests, to the repetition of the same unnatural pattern with different allomorphs, to the sharing of values by all the cells sharing form, to some other morphosyntactic rationale, etc. Different sources of systematicity are thus possible and widely heeded for different purposes. In Chapter 2, all these different possible sources of evidence will be discussed. Let it be mentioned pre-emptively, however, that there is no reason to believe that any of these sources should be superior or more important than the others. Availability of the information will obviously be the primary concern in a broad cross-linguistic endeavour like this book.

This chapter has provided an introduction and historical contextualization of the notion of the morphome, has clarified terminology, and provided a working definition of the object of enquiry of this book. To advance our understanding of the phenomenon of morphomicity, Chapter 2 deals at length with the most problematic issues around the definition of the morphome and their identification in specific instances.

By way of conclusion of this introduction, I would like to briefly clarify the place of this book within the broader 'Morphome Debate' (Luís and Bermúdez-Otero 2016). Readers of that volume will have undoubtedly noticed that the morphome is a strongly polarizing notion in the field. In my opinion, such a state of affairs is

[7] This will not apply for the inclusion of a morphological pattern into the synchronic morphome database in Chapter 4. In order to minimize subjectivity there, clearcut criteria will be specified to make consistent dichotomous judgements on morphomehood (i.e. morphome or not-morphome).

detrimental, and it is thus not my goal to 'take sides' in this debate. I would like to point out, however, that, although the very existence of morphomes (under some of its various senses) may still not be universally accepted, the objections raised against them tend to be mostly theoretical and philosophical (i.e. regarding what to say about them or how to best analyse them) at this point in the debate. 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. I thus hope that this book will be of interest to all morphologists and typologists, regardless of their theoretical convictions.

2

Issues in morphome identification

2.1 Systematic vs accidental

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 systematic, grammatically significant formal identities, from cases of so-called accidental homonymy, which are frequently dismissed as irrelevant to grammar and therefore uninteresting.

Some of the criteria which are 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 argued to be the case of English plural and genitive *-s* identity (e.g. Haspelmath 2003: 5).

The formal identity of plural and genitive might not seem semantically or cross-linguistically justified, and could thus be classified as accidental on the basis of these criteria. However, more sources of evidence could be brought forward to apply to the question of systematicity. As is often the case in linguistics, different diagnostics do not always converge. In English, for example, these two values share not only the same form but also an identical range and distribution of allomorphs (i.e. /s/, /z/, /iz/). Furthermore, when these values/formatives occur together, one suffices to express both meanings (e.g. 'tigers''). This constraint is not merely a phonological-identity-triggered haplology (Stemberger 1981), but can be shown to have grammatical import (consider the ungrammaticality of **the king**s** of England'**s** crown*). In addition to this, other morphs with the same form(s) and syntagmatic suffixal status occur elsewhere in the grammar, as 3SG agreement on verbs, and as clitic versions of *has* and *is*. These (also PL and GEN -s) have taken the upper hand diachronically over competing allomorphs (e.g. 3SG -th > -s) and conventions (e.g. '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(s), especially given the scarcity of morphology in the English language.

Be that as it may, from the perspective of the morphome it is obviously an unwarranted aprioristic assumption to always regard as grammatically uninteresting all morphological identities which lack cross-linguistic generality. If some

The Typological Diversity of Morphomes. Borja Herce, Oxford University Press. © Borja Herce (2023).
DOI: 10.1093/oso/9780192864598.003.0002

of those patterns, like the morphomes of Romance, are shown to be systematic, or even productive, within their language, then they must surely deserve attention and inform our morphological models.

The other main diagnostic for systematicity (semantic relatedness) means that a morphosyntactically coherent exponent (e.g. one which occurs across all 1PL verb forms, like *-mos* in Spanish) would be classified as systematic by virtue of this morphosyntactic coherence alone. No other proof would need to be offered to support the relevance of the formal identity of Spanish *va-mos*, *crezca-mos*, *ande-mos*, *tuvi-mos*, *amare-mos*, *so-mos*, etc. This diagnostic of systematicity is obviously unsuitable for morphomes because, by definition, they must lack a morphosyntactically coherent description. Evidence for the non-accidental character of a morphomic identity, therefore, will have to be sought somewhere else.

2.1.1 Assessing systematicity

It would be ideal to have some 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.

2.1.1.1 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 play.1PL/3PL against Bulgaria
'Either we or they will play in the Bulgaria match.'

2) ?Entweder Bierhoff oder ihr spielt gegen Bulgarien.
either Bierhoff or you.PL play.3SG/2PL against Bulgaria
'Either Bierhoff or you will play in the Bulgaria match.'

The above contrast, presented in Haspelmath and Sims (2010: 175), would suggest a systematic status for the formal identity of 1PL and 3PL verb forms in German (i.e. *spielen*) but an 'accidental homophony' status for the identity of 3SG and 2PL (i.e. *spielt*). This seems intuitively appealing because the former values are always whole-word syncretic whereas the latter are sometimes distinct (contrast e.g. 3SG *fährt* 'drive.3SG' and *fahrt* 'drive.2PL').

Unfortunately, it is not difficult to find limitations that severely compromise the usefulness and validity of this test. For example, one will often fail to find a construction which could be used to induce the required feature conflict. In addition, the test is obviously unsuitable for formal identities smaller than the whole word (i.e. for cases when only the stem or only an affix are formally identical). In practice, this renders this test inapplicable[1] to most morphomes, and as a result unsuitable for a broad cross-linguistic investigation like this one.

2.1.1.2 Co-occurrence restrictions

In an ideal world we would not expect the 'same' formative to appear more than once in the same word or domain. This is, patently, not always the case (the violation of this principle receives the name of 'multiple exponence', see Harris 2017); however, we might expect it to remain a very 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 freely, 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/-mış/-muş/-müş* depending on vowel harmony) that has both perfect and hearsay uses (Slobin and Aksu 1982). The two uses are very likely historically related; however, they are semantically compatible and the two can indeed co-occur synchronically 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-PRF-EVID
   '(It is said that) Kemal had come' (Slobin and Aksu 1982: 194)

Another Turkish suffix (*-lar/-ler)* is characterized by similarly related uses. It can mark both the plural of a noun and the plural of a third-person possessor. 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 from a logical perspective the two uses should be semantically 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 (see Table 2.1).

Table 2.1 Turkish noun number and possessor (Stump 2015: 176)

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

[1] For a more detailed discussion of the test and its limitations, see Johnston (1996: 13–14).

Simultaneous use of the two -lar is impossible, thus suggesting some inherent grammatical incompatibility, maybe because they are analysed by the language user as one and the same element despite its morphosyntactic–distributional complexity.

This co-occurrence test could, therefore, provide evidence to analyse these polyfunctional elements as one element with complex (co-argument-sensitive, Witzlack-Makarevich et al. 2016) semantics, or as multiple homophonous elements. It could thus help us distinguish accidental from systematic formal identities synchronically. However, there are also severe limitations to the validity and applicability of this test. First, as with the previous one, in many cases there might simply not be a word or construction in the language where the two elements could reasonably appear side by side. Second, the phenomenon of Obligatory Contour Principle, as usually portrayed (e.g. Yip 1988), constitutes an occasional, quite unpredictable obstacle to the appearance of phonologically identical contiguous sequences.[2] This may be independent of grammatical considerations. The effects of phonological and grammatical identity would be, in most cases, difficult to disentangle.

After surveying these two tests, thus, the conclusion is that, unfortunately, none of them can be reliably applied to obtain reliable independent evidence for the cognitive status of a morphological affinity. Other clues need to be therefore considered. Evidence for systematicity within a given language may also be plausibly sought from sources such as (i) evidence for a rationale of some sort in the morphosyntactic distribution of a form (even if this distribution falls short of complete naturalness), (ii) diachronic developments (e.g. analogical changes), and (iii) allomorphic variation, or morphophonological processes affecting all the contexts in the same way. These will be discussed next.

2.1.1.3 Morphosyntactic evidence for systematicity

Due to their morphosyntactically well-behaved nature, the systematicity of run-of-the-mill morphemes is not usually questioned. As mentioned before, /mos/ appears at the end of every 1PL verb form in Spanish, which seems by itself systematic enough that no morphologist would attempt to analyse the exponence as homophony of a *-mos1* and a *-mos2*. Morphomes are, by definition, not reducible to morphosyntactic determination. However, this is not the same as requiring complete orthogonality to morphosyntax. In fact, some of the most renown examples of morphomes (e.g. the Romance L- and N-morphomes) do abide by some soft morphosyntactic rationale, as their forms are limited to specific values (e.g. 'present' in both L and N) and in this way, they could be argued to 'mean' at least that.

[2] It may be relevant to point out here e.g. that in north Azeri (very closely related to Turkish) the two morphs in (3) are actually banned from occurring together, in what seems like a phonologically motivated dissimilation process (see Davis 2019).

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), as well as 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 *-rı̂ı̂*, for example, can code a 1PL.EXCL object in verbs or a 1PL.EXCL possessor in nouns. The morphosemantic core of the suffix is thus clear but is not enough to delimit its exact distribution in the language. A somewhat different case is that of Nuer nominal inflectional morphology (see Table 2.15 and Baerman 2012), where some suffixes have a problematic distribution which changes from one lexeme to another. Looking across all paradigms, however, 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 illustrated in the above cases, although perfect morphosyntactic determination is definitionally impossible in morphomes, a limited morphosyntactic rationale may still be offered as proof of systematicity in some cases. It may seem somewhat perverse to regard some morphosyntactic orderliness as diagnostic of a phenomenon that is defined precisely by its lack of morphosyntactic sense. However, for the reasons that will be presented in Section 2.3, this is a criterion that will be partially heeded here.

2.1.1.4 Diachronic evidence for systematicity

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 the prevalent opinion in the Romance morphomic literature. Well-known examples of formal identities which are occasionally reinforced and extended are the N-, L-, or PYTA morphomes of Romance languages (see e.g. Maiden 2011a).

While it certainly makes sense to pay close attention to analogical and diachronic changes in qualitative discussions in well documented families, this diagnosis is of limited applicability in the context of a broad cross-linguistic research like the present one. Most languages lack the historical documentation needed to access past *états de langue* with certainty. The history of a language or a pattern is also inaccessible to the naïve language user and therefore cannot be expected to play any role in linguistic cognition. Because of these limitations, diachrony will not be used here diagnostically, although morphome diachronics will, of course, still be paramount for a general understanding of the phenomenon, and will constitute the core of Chapter 3 of this book.

2.1.1.5 Allomorphic or morphophonological evidence for systematicity

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', and *seducir* 'seduce'. There is synchronically no verb with the form **ducir*, and the various verbs containing that root do not have any obvious semantic affinity. If these were all the facts, we may have had to conclude that the formal similarity between these verbs was accidental and grammatically moot. However, all of them are subject to the same phonologically unmotivated alternations in inflection and word formation: *condu**zc**o* 'I drive', *condu**j**e* 'I drove', *condu**cc**ión* '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 unit at some level despite the lack of shared semantic content.[3] Kayardild's (mero)morphomes (see Round 2015), similarly, also evidence 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 thus be observed (and may be repeated with different exponents) between lexemes (e.g. Spanish *-ducir*), between inflectional affixes in different parts of speech (Kayardild case-tense affixes), and between the different paradigms cells of a single lexeme, as in the best-known Romance morphomes (see Table 2.2).

Table 2.2 L-morphome allomorphs in Spanish (partial paradigm)

	venir 'come'		*nacer* 'be born'		*caber* '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

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

[3] This is not to say that this unity cannot sometimes be eroded, the same as any other grammatical category. The verb *seducir*, for example, appears to be more prone to losing some of these alternations (e.g. having regular *seducí* 'I seduced' instead of irregular *seduje*). This might be 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*.

have learned the pattern as two different elements: a -*t*1 triggered by 3SG subjects and a -*t*2 coextensive with the 2PL. Thus, it is often not until an unnatural distribution is replicated with different forms that morphologists recognize a morphome (see also the economy considerations in Section 2.11). Consider, for example, the morphological identity in Udmurt shown in Table 2.3.

Table 2.3 Inflectional suffixes in Udmurt (Uralic) (Csúcs 1988: 142)

	1st conjugation indicative suffixes				2nd conjugation indicative suffixes			
	PRS		FUT		PRS		FUT	
	SG	PL	SG	PL	SG	PL	SG	PL
1	-iśko	-iśkom	-o	-o-m	-śko	-śkom	-lo	-lo-m
2	-iśkod	-iśkodi	-o-d	-o-di	-śkod	-śkodi	-lo-d	-lo-di
3	-e	-o	-o-z	-o-zi	-Ø	-lo	-lo-z	-lo-zi

The sharing of form by the 3PL present and by all future forms is repeated in the two conjugations of the language with different formatives: -*o* and -*lo*. This fact provides stronger evidence for the induction of a generalization/rule 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.PRS 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 morphome. This is a criterion I will adopt here too, particularly in the systematic cross-linguistic exploration in Chapter 4.

2.1.2 On the empirical status of homophony and polysemy

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

	Future				*Genitive*	
4a)	Leihoa	ireki-ko	dut	4b)	Hiri-ko	atea
	window	open-FUT	have.1SG		city-GEN	door
	(I will open the window)				(The door of the city)	

5a)	Madrilera	joan-go	naiz	5b)	Irun-go	neska
	to.Madrid	go-FUT	be.1SG		Irun-GEN	girl
	(I will go to Madrid)				(The girl from Irun)	

6a)	Horrela	egin-en	dute	6b)	Mikel-en	aita
	thus	do-FUT	have.3PL		Mikel-GEN	dad
	(They will do it that way)				(Mikel's dad)	

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

Much effort has been devoted to answering this polysemy vs homophony question in specific cases (see e.g. Harbour 2008). However, one might wonder whether 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 $\textit{-ko}_1$ and a $\textit{-ko}_2$ 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 extratheoretical 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 a linguistic phenomenon from its analysis by (native) language users. Human beings inevitably have to analyse their language input (i.e. posit some categories, make certain analytical choices) to make 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 naïve language user constitutes the phenomenon itself, and should be the object of study.

This does not mean 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, it is unquestionably unfortunate. Consider, by way of example, the following agreement patterns with some Spanish nouns.

7a)	**la**	costa(F)	peligros-a	7b)	**las**	costas	peligros-a-s
	The	coast	dangerous-F		the	coasts	dangerous-F-PL

8a)	**el**	arma	peligros-a	8b)	**las**	armas	peligros-a-s
	the	weapon	dangerous-F		the	weapons	dangerous-F-PL

9a)	**el**	tema(M)	peligros-o	9b)	**los**	temas	peligros-o-s
	the	issue	dangerous-M		the	issues	dangerous-M-PL

The traditional account of this allomorphy of the definite article is that the form *el* in (8a) is not 'the same thing' as that in (9a), and that they just happen to be accidentally homophonous. The 'official' (see e.g. RAE-ASALE 2009: 23, 265–7) analysis argues that *el* 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 allomorph 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'), and the phenomenon must have indeed originated from some differential evolution of the form of the feminine article in these phonological contexts.

However, there is abundant synchronic evidence that this is no longer the analysis of (most) language users, which regards 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. First of all, it is just nouns, and no other grammatical category, that can trigger this allomorph (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. proper nouns do not do so: *la Ana* [*el Ana], *la A* [*el A]). Secondly, 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 adjacent to the noun (e.g. *un*(M) *hambre tremendo*(M), or *un*(M) *bonito*(M) *águila*, which is five times more frequent on the Internet than the 'correct' *una bonita águila*). These speaker practices and changes, which occur despite linguistic prescription, would make absolutely no sense if language users 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 thus have been a more or less fortuitous outcome of sound change (**ela kasa* > *la=kasa*, **ela alma* > *el=alma*). However, after this configuration emerged, language users had the understandable impulse to associate the form with other *el* rather than with other *la*, and the nouns taking this *el* with other (masculine) nouns taking *el*.

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

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 believe that it is indeed a relevant distinction, in linguistics, whether the *el* 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 morphological identity is cognitively relevant, i.e. part of the grammatical system of native speakers, or merely reflects the inert outcome of some historical accident.

Although we currently lack this type of access to the mind of language users, there seems to be experimental evidence that the homophony/polysemy distinction that has traditionally worried linguists is, indeed, a cognitively real one. Pylkkänen et al. (2006), for example, found noticeable differences in the speed at which polysemous and homophonous pairs are processed. This suggests that the difference that linguists intuitively sense between these two kinds is not a mere illusion.

Many diachronic changes can also be offered as evidence that whether or not language users make a generalization over two forms is of the utmost importance. Among the most revealing, in my opinion, are those cases where an originally single lexeme splits into two. This may happen, quite revealingly, 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.

2.1.2.1 Semantically motivated split

The Spanish verb *saber* can mean both 'know' and 'taste'.[4] 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 all language users. Under the meaning 'taste' the verb is understandably used almost exclusively in the third-person. However, when native speakers produce the rest of the forms, these are often *sepo* (e.g. *yo sepo salado* 'I taste salty'), *sabes*, *sabe*. The 1SG present form may thus differ from the one found under its sense 'know'.

It seems thus (see Table 2.4) 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 the two

[4] This section relies partially on arguments in Herce (2018).

Table 2.4 Spanish present-tense forms of *saber* 'know' and *saber* 'taste'

	saber$_1$ 'know'				*saber*$_2$ '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

forms 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*) are only ever encountered in the input under their meaning 'know', they do not count as evidence for language users' deduction of the full paradigm of *saber*$_2$ 'taste'. This means that the first and second-person forms of *saber*$_2$, when needed, have to be constructed 'online' on evidence exclusive to its sense 'taste' (i.e. third persons and non-finite forms), as well as, more generally, on the evidence of recurrent patterns of allomorphy in Spanish verbal inflection.

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.

Unlike *saber*$_1$, verbs whose stem differs between the third-person indicative and subjunctive (e.g. *tiene* vs *tenga*, *cabe* vs *quepa*) consistently have the same stem form in the 1SG indicative as in the subjunctive (Table 2.5). Knowledge of this abstract stem alternation pattern must be what leads Spanish language users to innovate a form *sepo* rather than **sabo*.

Table 2.5 Partial paradigms of some Spanish verbs

	saber$_1$ **'know'**	**saber**$_2$ **'taste'**	**tener 'have'**	**conocer 'know'**	**caber 'fit'**	**caer 'fall'**
1SG PRS IND	sé	sepo	tengo	conozco	quepo	caigo
2SG PRS IND	sabes	sabes	tienes	conoces	cabes	caes
3SG PRS IND	sabe	sabe	tiene	conoce	cabe	cae
1SG PRS SBJV	sepa	sepa	tenga	conozca	quepa	caiga
2SG PRS SBJV	sepas	sepas	tengas	conozcas	quepas	caigas
3SG PRS SBJV	sepa	sepa	tenga	conozca	quepa	caiga

The analogical reshaping operated from the paradigm of *saber*$_1$ to *saber*$_2$ suggests that these purely morphological patterns (the so-called L-morphome in this case, see Table 1.1) can 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).

2.1.2.2 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 (i.e. *deiwos* > **de:wos* > **de:os* > *deus*) meant that a stem alternation emerged in the paradigm (see Table 2.6).

Table 2.6 Expected paradigm of *deus* (Thurneysen 1887: 155)

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

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. The two forms parted ways definitively when language users analogically created the 'missing' forms to generate complete inflectional paradigms (see Table 2.7).

Table 2.7 Latin paradigms of *deus* (left) and *dīvus* (right)

	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

The two cases presented in this section suggest that whether or not language users think of two forms as part of the same grammatical category is, indeed, crucial. This even allows to make some 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 derived automatically from the loss of the former cognitive generalization.

These lexemic splits also suggest that, as will be argued throughout this book (e.g. in Sections 2.3 and 2.4), both semantic-functional distance (e.g. in *sapere*) and formal distance (e.g. 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 as significant by language users is, unfortunately (to reiterate), not directly accessible to linguists. Before any change reveals it on the surface, the lexemic unity may already have been broken in the cases presented above. Thus, we cannot always conclude that in the absence of surface morphological changes, the deeper grammatical unity still holds. 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 a more difficult one to solve, and not less of a problem. In my opinion, therefore, the fact that very often 'You can't tell' does not render the whole polysemy/homophony distinction a figment of the imagination of linguists, but simply a harder nut to crack.

2.2 Natural vs unnatural

As usually construed (e.g. Bybee 1985: 118; Haspelmath and 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 and 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), and this hierarchical structure is explicitly assumed in many models (e.g. in Functional Discourse Grammar, Hengeveld and Mackenzie 2008). In models with this overall architecture, morphology is thus considered 'post-syntactic' (e.g. in Anderson's (1992) A-morphous Morphology, and in Distributed Morphology, e.g. in Halle and Marantz (1994)), so that syntax and semantics are 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. For example, Distributed Morphology morphs realize single syntactic terminals. Most realizational models (see e.g. Matthews 1965) posit rules of grammar that spell out in surface abstract morphosyntactic properties. Thus, although it may seem 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 way of thinking) elements of form exist merely to express morphosyntactic distinctions, morphology 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 or semantic similarity and conversely, morphosyntactic differences should be signalled by differences in form. Such 'canonical' structures are not difficult to find (see Tables 2.8, 2.9, and 2.10).

Table 2.8 Teribe (Chibchan, Panama) deictic-directional verbs (Quesada 2000: 67)

	Downwards	**On the same plane**	**Upwards**
Towards EGO	ter	tek	tem
Away from EGO	jer	jek	jem

Table 2.9 Suena (Trans-New-Guinea) pronouns, INCL forms excluded (Wilson 1974: 16–17)

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

Table 2.10 Kusunda (Isolate, Nepal) verb *əm* 'eat', realis (Watters 2006: 60)

	SG	PL
1	təmən	təmdan
2	nəmən	nəmdan
3	gəmən	gəmdan

From the perspective of Canonical Typology (Corbett 2005; Brown and Chumakina 2013), the above cases can be considered canonical inflectional paradigms

(Stump 2015: 35–41). As mentioned by Round and Corbett (2017: 54), 'Canonically, 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 2.8, 2.9, and 2.10 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 never elsewhere.

The existence of structures like those of Teribe, Suena, and Kusunda points to 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 and Audring 2017). The probability of such perfectly isomorphic structures occurring by chance would be infinitesimal, and yet they are found comparatively 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. It is therefore widely agreed that '[t]here is a universal semiotic principle favouring biunique matching of lexical signata and signantia' (Maiden 2011c: 266).

The empirical evaluation of this alleged principle is, however, extremely challenging in practice. There is, in fact, widespread disagreement in the literature as to whether one-to-one mappings are the most frequent cross-linguistically: '[a] biunique relation between meaning and form is the most common relation in inflectional morphology' (Aalberse 2007: 114); 'the "one meaning–one form" principle is actually used very sparingly' (Bybee 1985: 209).

To be able to assess these claims empirically, one would need a thorough quantitative typological investigation coupled with clear criteria for segmentation (see Section 2.10), 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 the goals in the present book. It should be kept in mind at all times, however, that linguists deduce whether a morphosyntactic distinction (e.g. tense or number) is present or absent in the grammar of a language precisely by looking for morphological correlates along those lines. A unitary treatment concerning form can even lead linguists to posit a grammatical category (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 (e.g. the Spanish 'imperfect') and cases (e.g. the Latin 'ablative'). These show that, at least sometimes, formal identity leads linguists to posit language-particular grammatical categories (i.e. features or values) for which no evidence exists outside morphology itself. Similarly, if we happen to observe lexeme-dependent formal distinctions with no clear semantic correlate we just posit ad hoc features like gender[5] or inflection classes.

This *modus operandi* could be argued to be perverse in that biuniqueness becomes a self-fulfilling prophecy. Whether consciously or not, we are often building up (bi-)uniqueness into our descriptions of morphological systems. It is hardly surprising, therefore, that we should find strong parallelisms between formal/morphological and morphosyntactic/semantic structure. And yet, despite this approach, we do find many cases in which, unlike in Tables 2.8–2.10, the mapping between form and features is not canonical. Various such cases will be present throughout the remainder of this section in order of increasing deviation from the biuniqueness ideal.

Consider first cases of cumulative exponence like Albanian in Table 2.11. They 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 in Tables 2.8–2.10. Unlike in those perfectly isomorphic examples, formal elements in Table 2.11 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 given morphosyntactic feature value. That is, we cannot identify in Table 2.11 a marker for [addressee] or for [plural].

Table 2.11 Albanian *laj* 'wash' present non-active (Newmark et al. 1982: 59)

	SG	PL
1	lahem	lahemi
2	lahesh	laheni
3	lahet	lahen

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 values (first-person+plural). This cumulative exponence might

[5] This need not even have a 'semantic core'. See e.g. 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.'

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).[6] The issue boils down to the theoretical boundary between syntax and morphology and will not be discussed here further.

A different subtle deviation from the canonical isomorphic inflectional paradigm can be illustrated by the Russian past-tense inflection in Table 2.12.

Table 2.12 Russian past imperfective forms of the verb ‘work’

<table>
<tr><th></th><th>SG</th><th>PL</th></tr>
<tr><td>M</td><td>rabotal</td><td rowspan="3">rabotali</td></tr>
<tr><td>F</td><td>rabotala</td></tr>
<tr><td>N</td><td>rabotalo</td></tr>
</table>

Russian verbs in the past-tense agree in gender and number. However, gender agreement does not apply in the plural. These cases, where sensitivity to a feature is seemingly lost completely within a certain domain, are not usually considered exceedingly problematic. The form in question (i.e. *rabotali*) is usually considered simply un(der)specified for gender. This means that it is usually considered uninformative regarding gender rather than ambiguous between the different gender values. The form would still have, therefore, a clear atomic meaning [plural]. This same analysis may be (un)suitable for other syncretisms.

Manambu personal pronouns in Table 2.13, for example, distinguish gender in the second and the third-person singular but not in the plural. In the dual, the distinction between second and third-person is also missing. We thus cannot say

Table 2.13 Manambu (Ndu, New Guinea) personal pronouns (Aikhenvald 2008: 66)

<table>
<tr><th></th><th>SG</th><th>DU</th><th>PL</th></tr>
<tr><td>1</td><td>wun</td><td>an</td><td>ñan</td></tr>
<tr><td>2F</td><td>ñən</td><td rowspan="4">bər</td><td rowspan="2">gwur</td></tr>
<tr><td>2M</td><td>mən</td></tr>
<tr><td>3F</td><td>lə</td><td rowspan="2">dəy</td></tr>
<tr><td>3M</td><td>də</td></tr>
</table>

[6] In the absence (in languages like Albanian) of morphological evidence for independent features like person and number, we may wonder what the need is 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 analysing each of the six morphosyntactic entities in Table 2.11 as irreducible morphosyntactic objects.

that features such as gender or person are relevant or irrelevant in the domain of a certain number value. Finer-grained conditions are required to describe the distribution of forms and sensitivity to a particular feature.

When distinctions are fewer in one domain relative to another—that is, when one form in one domain corresponds to several forms in another domain—feature structure will determine whether a form's domain constitutes a natural or unnatural class.

Kwomtari, as presented in Table 2.14, sometimes conflates the values for first and second-person plural (e.g. object suffixes), and at other times the values of first and third-person plural (e.g. subject suffixes). In both cases there is a degree of systematicity, since both patterns (i.e. 1=2 and 1=3) are found twice 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 morphological neutralizations problematic for morphological models with a rigid hierarchical feature structure.

Table 2.14 Kwomtari (Kwomtari-Nai, New Guinea) person agreement (Spencer 2008: 107)

	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		

There are also approaches to morphology, however, which are based on the 'lexicalization' or 'spelling' of 'adjacent' features (e.g. geometrical, McCreight and Chvany 1991, and nanosyntactic, Caha 2009). Because they are less restrictive, these frameworks 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), a single form could spell out any combination of adjacent values. There are syncretisms, however, that defy any such orderings.

All number values in Kiowa (Table 2.15) can be syncretic with any other number value, which makes it impossible to arrive at any fixed order such that formal identity would occur 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).

Table 2.15 Kiowa number marking (Wunderlich 2012: 178 after Wonderly et al. 1954)

<table>
<tr><th></th><th>I</th><th>II</th><th>III</th><th>IV</th></tr>
<tr><td>SG</td><td rowspan="2">che:</td><td>tósè-gɔ</td><td>k'on-dɔ</td><td rowspan="3">tʰòn</td></tr>
<tr><td>DU</td><td rowspan="2">tósè</td><td>k'on</td></tr>
<tr><td>PL</td><td>che:-gɔ</td><td>k'on-dɔ</td></tr>
</table>

Bi- or tridimensional formal conflations, in turn, also vary in the extent to which they can be analysed as the expression of a value. In Table 2.16, for example, the form *fecebil* conflates both 2 and 3, and DU and PL. Under the right feature structure, the distribution of this form can be characterized simply as non-speaker non-singular. It would thus have a morphosyntactically coherent description and could be regarded as a natural morphemic exponence.

Table 2.16 Amele (Trans-New Guinea) verb 'see' perfect switch reference (Roberts 1987)

<table>
<tr><th></th><th>SG</th><th>DU</th><th>PL</th></tr>
<tr><td>1</td><td>fecemin</td><td>fecohul</td><td>fecomun</td></tr>
<tr><td>2</td><td>fecem</td><td rowspan="2" colspan="2">fecebil</td></tr>
<tr><td>3</td><td>feceb</td></tr>
</table>

Patterns of formal identity involving L-shaped or T-shaped configurations are considered more problematic. 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 (Table 2.17).

Table 2.17 Verb 'go' in Benabena, past-tense (Young 1964: 48)

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

This category (i.e. SG and/or 1) is labelled 'monofocal' by Young, while the other cells were labelled 'polyfocal', thus hinting at the possibility of a semantic affinity of some sort between the values. Regardless of the merits of this specific analysis, L-shaped patterns like these 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.

Jackendoff (1985), for example, explains how the (or his) use of the verb *climb* is appropriate to describe actions involving upwards motion and/or those performed with the use of limbs (Table 2.18). 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. L- or T-shaped patterns can and often do (see Section 3.1.3.1) arise in one step from natural morphosyntactic distributions by means of natural morphosyntactic or semantic extensions.

Table 2.18 Meaning features of *climb* (Jackendoff 1985)

	Clambering	**No clambering**
Upward	climb	climb
Downward	climb	–

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. In Table 2.19, 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.

Table 2.19 SS NFUT medial verb agreement in Safeyoka (Angan, New Guinea) (West 1973: 10)

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

The morphosyntactic contexts where *-onji* appears still, however, constitute 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 and Chvany's (1991) geometrical approach. Other patterns (see Table 2.20) do not occupy a contiguous morphosyntactic space and are thus problematic even for these models.

The difficulty of capturing the distribution of an exponent thus increases with the number of disjoint contexts in which it appears. In addition, as will be mentioned in Section 2.8, it may also make a difference, and it is at any rate more problematic in theoretical analyses relying on defaults and blocking, whether or

Table 2.20 Skolt Saami (Uralic) *maadd* 'base', partial paradigm (Feist 2011: 146)

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

not an exponent's distribution is interlocked with that of another unnaturally distributed exponent as the paradigm in Table 2.21.

Table 2.21 Subject agreement in Yagaria, partial paradigm (Stump 2015: 128, after Haiman 1980)

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

These cases, where formatives have a distribution completely orthogonal to the assumed morphosyntactic feature structure, and where descriptions/analyses 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, troublesome for formal models that do not grant an independent status to morphology.

Different linguists would interpret in different ways the data which have been presented throughout this section.[7] However, the fact that these patterns are possible in natural languages seems to suggest that form–function isomorphism is not the only possible organizational principle for inflectional morphology. Isomorphism, thus, might well constitute a tendency for paradigmatic organization, but one which can be overridden under the right circumstances. An exhaustive typological study of those cases is likely to provide valuable information about the nature of morphological architecture and linguistic cognition.

[7] Bi-uniqueness 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 are often hypothesized to be different from the ones we see at the 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 e.g. 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 cognitively real. They may be largely formal machinery aimed simply 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 1936[1971].

2.3 Maximal domain

One of the questions that remains open (and seldom addressed) regarding morphomic structures is whether morphosyntactic or paradigmatic structure imposes any limit to them. It seems reasonable (remember the discussion related to the verb *saber* around Table 2.4) that functional similarity and feature values may play some role regarding the perception of a pattern of formal identity as grammatically (in)significant by the language user. 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 2015) believe that morphomic connections 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. which domain, if any, should be regarded as the broadest within which morphological structure is possible) is related to the acquisition of these structures by the language user. The difficulty of learning or perceiving a given formal identity as systematic is likely to increase if independently justified morphological or semantic 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. within [singular] or [present]) than across those domains.

One of the reasons why the morphome is such a controversial object of study is that a certain level of contradiction is present in its very definition. It is quite remarkable that, for us to accept some case as a genuine instance of a morphome, we usually require that a given formal identity be at once (i) 'chaotic' and (ii) 'systematic'. We are, therefore, demanding two things which are almost antagonistic: (i) morphosyntactic unnaturalness, and (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 in the 3PL possessor form of the noun would be considered very morphomic indeed according to the unnaturalness criterion. According to the second criterion, the more systematic a formal identity is, the more we should regard it as a grammatical single unit or category of some kind. The problem is that, as mentioned already in Section 2.1.1.3, one of the main sources 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 completely accidental.

This way of understanding grammar is not a theoretical whim of linguists. On the contrary, I believe it is completely justified. Language users, when making sense of their linguistic input, must also use these cues when deciding whether or not two occurrences of the same form are instantiations of the same element. It is a plausible hypothesis that the amount of evidence required to 'convince' language users that a formal identity is relevant grammatically varies as a function of the perceived distance between the 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. Naïve speakers of Basque are surprised when this formal identity is pointed out to them. In addition, the distribution of phonologically conditioned allomorphy *-ko*/*-go* is no longer identical in its two uses. For example, after stems ending in /l/, most speakers use *-go* for the genitive (e.g. *Madril-go* 'of Madrid') but *-ko* for the future (*hil-ko* 'will kill'). The different morphophonological paths taken by these formatives suggest that their formal identity might not be cognitively relevant in synchronic terms.

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 analysed 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, should indeed 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 usually inflect for different features, which 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; 2015: 229) unmistakably demands that speakers be able to make unified analyses of nominal and verbal suffixes sometimes. Languages like Tundra Nenets, for example, have sets of suffixes indexing person–number combinations in different word classes (see Ackerman and Bonami 2017). The possessor in nouns, the subject in verbs, and the object in prepositions are all marked with exactly the same markers regardless of the word class they attach to. Postulating different homophonous affixes (e.g. a *-da*$_1$ in nouns vs a *-da*$_2$ in verbs, a *-maq*$_1$ in nouns vs a *-maq*$_2$ in verbs) 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* 'mean' 1PL.

Morphological objects, therefore, seem able to 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? When different word classes do inflect for the same values, morphomic paradigmatic patterns can indeed be shared by different classes.

Consider the person–number syncretism in Khanty in Table 2.22. The same unnatural syncretism pattern (2/3DU+2PL) is found in nouns for possessor inflection and in verbs for subject inflection. The same pattern is also found in other possessee and object numbers (Table 2.22 shows only singular object/possessee), which suggests we are dealing with a systematic trans-categorial unnatural pattern of syncretism.

Table 2.22 Khanty (Uralic) possessor (left) and subject (right) inflection (Nikolaeva 1999)

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

Zooming in more, for example within a single word class, it is my contention that it should become 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 in the inflectional material, and might even be said to be the 'default'.

Table 2.23 Declension of two Russian nouns

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

Consider the two Russian declensions of Table 2.23. It would be unreasonable to regard as accidental that the oblique plural suffixes of the different inflection classes of Russian share the same form. This, in fact, was the result of an analogical levelling implemented by language users (cf. Slovene DAT.PL *-am* vs *-om*), so positing homophonous affixes (e.g. in the dative plural: *-am*$_1$, *-am*$_2$) would seem to be a misrepresentation.

When considering other formatives, however, the situation seems different. The suffix *-u* can mark the accusative singular (in *rabota*), and the dative singular (in *mesto*). Should we recognize independent homophonous suffixes -u_1, -u_2 because -*u* has different values in different inflection classes? Or should we understand -*u* as an inter-class inflectional formative (i.e. as a single operation which can map onto different values in different classes, à la Kayardild (Round 2016))? The evidence in favour of the latter analysis is, intuitively, quite weak (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 in Russian, is not unlikely to be used more than once in case–number inflection merely by chance. Much as in Basque, therefore, this morphological identity may well be moot grammatically.

There are other cases, however, where it seems more plausible that affixes in different classes might be 'the same thing' at some level despite having different morphosyntactic distributions. Consider, for example, the case of Nuer in Table 2.24. The formative -*ni* appears across different nominal inflection classes. 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 (where it is also the only possible suffix), intuitively suggest that positing many homophonous -*ni* may not be the right approach. The alternative is, inevitably, that there is a single formative with a complex morphosyntactic distribution.

Table 2.24 Some inflection classes in Nuer (Nilotic) (Baerman 2012: 470, from Frank 1999)

	Class I		Class IV		Class VIII		Class XIII		Class V	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
NOM	-Ø	-Ø	-Ø	-ni	-Ø	-Ø	-Ø	-Ø	-Ø	-Ø
GEN	-Ø	-ni	-Ø	-ni	-Ø	-ni	-Ø	-Ø	-Ø	-Ø
LOC	-Ø	-ni	-Ø	-ni	-Ø	-Ø	-Ø	-ni	-Ø	-Ø

As was the case with morphomic identities across word classes (e.g. in Basque or Kayardild), the same unnatural morphological affinity can actually be repeated with several exponents across inflection classes. This, in principle, reduces the likelihood of a morphological identity being accidental.

Consider, for example, the inflection classes in Table 2.25. The suffix of the form -*ni*, again, appears in a seemingly arbitrary set of contexts in different verbal inflection classes. However, its distribution is matched exactly by that of the formatives -*di* and -*li* in their respective inflectional classes. 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 in Table 2.25. If they have not, the forms of IPFV or PFV could 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. Any affixed IPFV form, for example, immediately entails unsuffixed AOR and PFV forms. Conversely, an affixed PFV also entails an unsuffixed IPFV, and a suffixed AOR entails an identically suffixed PFV, and an unsuffixed IPFV.

Table 2.25 Some inflection classes in Gourmanchéma (Atlantic-Congo) (Baerman et al. 2017, after Naba 1994; Ouoba 1982)

	'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

However interesting morphological affinities across classes may be, the domain within which morphological identities are usually explored in morphomic literature has tended to be smaller than this. Many researchers, in fact, have voiced objections to treating morphological affinities beyond and within the paradigm (or even beyond and within a subparadigm) in the same way. Blevins, for example, argues:

> Pairs of elements with no discernible connection, such as the agentive and comparative -er markers in English, are (...) not morphomes. 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 (e.g. Basque, Kayardild), or between different inflection classes (e.g. Nuer, Gourmanchéma), cannot ever be morphomic. Pertsova goes even further in the restriction of the window of opportunity for morphomes when she argues:

> 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 is certainly sensible. As argued by Bermúdez-Otero and Luís (2016: 337), '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 linguists are justified, as they are supported by ample evidence from the category and concept learning literature (Shepard et al. 1961; Goodman et al. 2008; Pertsova 2014). It is therefore plausible that if the Basque identity, instead of genitive and future, had involved closer functions, it might have remained a synchronically active part of the grammar.

Regardless how well-grounded these concerns are, in the absence of sensible, uniform, concrete ways of implementing them in cross-linguistic morphome identification, there is a danger that one will simply disregard morphological identities for arbitrary reasons or because they 'do not fit' into a particular theoretical framework. One could, for example, restrict what counts as an 'independently justified morphological domain' in a way that renders the possibility of morphomic exponents altogether impossible. If, for example, the present-tense sub-paradigm, or the singular sub-paradigm, constitute domains of this kind, any formative that occurs inside and outside of any of these domains will simply be reanalysed as two homophonous formatives rather than one. In this way, even the most incontrovertible morphome would be simply 'converted' into two or more homophonous morphemes. This is thus clearly not the right approach to investigate morphomicity.

A more sensible criterion could be that advocated by Blevins (2016). There is, I believe, 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 identifiable within a single lexeme's paradigm and is simply replicated in others.

As the paradigm in Table 2.26 shows, within any Burmeso 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 these. Thus, the contexts where these forms appear within a single paradigm constitute an unnatural class. The fact that the 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. Intraparadigmatic morphomes like the one in Burmeso are thus less controversial that transparadigmatic ones.

This is not meant to imply that morphological relationships beyond the paradigm are always spurious. It is hardly a far-fetched suggestion, for example, that the systematicity of Gourmanchéma verb class structure may be enhancing the learnability of the system. Its nine inflection classes can be arranged into just

Table 2.26 Conjugations in Burmeso (Isolate, New Guinea) (Corbett 2009: 9, from Donohue 2001: 100–102)

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-

three classes based on the suffix used: *-ni*, *-di*, or *-li*, and into another three classes based on the paradigmatic distribution of the affix.

If the achieved economy (in this case abstracting six categories instead of nine, see Table 2.27) is sufficient, then these generalizations may be worth making 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 morphomes.

Table 2.27 Orthogonality of Gourmanchéma inflectional classes' traits

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

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. In other words, 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 some morphological affinities are ignored instead (but see Section 2.11 on economy).

Because looking into speakers' brains is not an option in this context, an alternative strategy has to be sought to try to discard most instances of spurious morphomes like the one in Basque. Morphological affinities beyond the paradigm are necessarily weaker than those within a single paradigm. The amount of morphological evidence required to 'convince' a language user that genitive and future are marked by the same formative must therefore be more than that required to convince them of some intraparadigmatic affinity. In the context of broad cross-linguistic research, the cognitive status of individual morphomic patterns cannot be investigated in detail. Because of this, an executive decision has been taken to focus, in the remaining of this book, exclusively on those morphological structures which are apparent within the inflectional paradigm of a single lexeme.

2.4 Independence from phonology

Among the definitions of 'morphome' that circulate in the literature, one finds frequent references to the phonological component of language. O'Neill (2013: 221), for example, reports that one of the most usual senses of 'morphome' refers to elements 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 morphomic or not (e.g. Anderson 2011 vs Maiden 2011b) have also sometimes revolved around the independence of that pattern from concrete phonological environments. Morphomes, however, are precisely about form, so applying the criterion that a morphome has to be independent from phonology is difficult.

Consider the subparadigms in Table 2.28. In Russian *peč'*, the distribution of *k* vs *č* as the last consonant of the stem is perfectly correlated to the nature of the following suffix *-u* vs *-ë* (/o/ now, a front vowel originally). In Spanish *plegar*, in turn, 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 these verbs display. 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 could not possibly be morphomic (although see Maiden 2017 and Herce 2021a for a different opinion).

Table 2.28 Stem alternation patterns in a Russian and a Spanish verb

	Russian *peč'* 'bake'		**Spanish *plegar* 'fold'**	
	SG	PL	SG	PL
1	pek-u	peč-ëm	ˈpljego	pleˈgamos
2	peč-ëš´	peč-ëte	ˈpljegas	pleˈgajs
3	peč-ët	pek-ut	ˈpljega	ˈpljegan

To decide whether the alternations in Table 2.28 are morphomic, thus, one would need to assess whether they constitute productive phonological processes in these two languages synchronically. Beyond these paradigmatic alternations, there is absolutely no support for a general synchronically active rule which transforms

/k/ into /t͡ɕ/ in Russian before an /o/ (or before a front vowel, for that matter), or which turns /e/ into /je/ in Spanish in the presence of stress, as both are possible in either phonological environment. Assigning these patterns to the phonological component, for example by positing diacritics that diphthongizing /e/s 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 synchronic phenomenon. The only outcome of these approaches, as far as I can see, is to shrink 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*) is simply transferring the burden of the explanation to some other part of the system. Although there is a widespread theoretical impulse to derive the forms of stems from the forms of suffixes, there is no empirical reason why one of them would require an explanation while the other one would not. The same thing applies to diphthongization in Spanish. Explaining the paradigmatic distribution of the N-morphome (e.g. of /je/ in *plegar*) by deriving it from stress ignores the fact that the location of stress is itself unpredictable in the language. As pointed out by Esher (2015), the paradigmatic distribution of rhizotony in the Spanish paradigm is 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 to whether they undergo diphthongization or not.

If we are to remain as close as possible to the empirical data and avoid problematic assumptions, all we can note in cases like the Spanish and the Russian ones in Table 2.28 is that there is a perfect correlation between the distributions of two different formal elements which would not need to occur together synchronically but do so in these paradigms. The existence of a correlation could well point to more and not to less morphomicity for these patterns (see Herce 2021a). The morphological affinity assumed by the N-morphome, after all, is reproduced in a verb like *plegar* not once but twice, with two different exponents: presence vs absence of a glide /j/, and presence vs absence of rhizotony.

When one goes beyond the simple description of form distributions, analyses become more subjective. 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 as morphological rules or merely constitute a perpetuation of the context that historically generated the alternations. Disagreements are ubiquitous in this respect. Bermúdez-Otero and Luís (2016), for example, argue for the synchronic relevance of the correlation. They offer evidence of dialectal analogical developments rendering 1PL.SBJV ˈ*pwedamos*[8] (compared to standard Spanish *po*ˈ*damos*). O'Neill (2011), however, argued that we cannot infer causation from these cases, as in other varieties stress and stem vowel diphthongization have been found to lead separate lives (see e.g. Alta Ribagorza Aragonese in Table 4.58).

According to the division of labour between phonology, morphology, and syntax which is adopted in this book, non-automatic alternations like those of Spanish and Russian in Table 2.28 will not be considered phonological processes. Accordingly, morphological patterns will not be excluded from the ranks of morphomes just because they are coextensive to some phonological environment. Clear-cut cases of automatic phonological determination do exist. This is the case, for example, of the stem alternation in Table 2.29.

Table 2.29 Declension of the adjective *mraj*- 'lucky' in Alutor (Chukotko-Kamchatkan) (Kibrik et al. 2004: 287)

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

As explained by Kibrik et al. (2004: 287) 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. 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[9] and will not be discussed here.

Another issue that has to be settled in relation to the independence of morphomes 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 morphome, 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:

[8] This change would still leave the direction of causation possibly undetermined (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.

[9] Note that the non-morphomic character of even these patterns is not unarguable. 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.PRS.IND /ˈamə/ vs 2SG.PRS.IND /ɐˈmaː/) was apparently not analysed as such by language users since, when they analogically levelled stressed within the present-tense, the allomorphy was preserved (i.e. 1SG.PRS.IND /ˈamə/ vs 2SG.PRS.IND /ˈɐmaː/). Similarly, research in East Kiranti (Herce 2021) suggests that phonologically derivable patterns of stem alternation are acquired redundantly, since they show otherwise unexpected diachronic resilience and influence on affixal allomorphy.

> the clearest and most predictive aspects of the L-morphome 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 behind these assertions appears to have been the following: to be sure that an unnatural morphological identity is systematic and not an instance of accidental homophony, morphologists have usually required that an identity be repeated with various different forms. Because of the multi-allomorphic nature of these morphomic patterns, those cases have often been conceived and formalized, in turn, 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 their particular instantiations. This is intuitively sensible, since it is forms (i.e. concrete exponents) 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 morphological identity (morphemic or morphomic) could plausibly be generalizable (e.g. in wug tests) to unattested forms.

Consider the case of the Italian past indicative stem allomorphy in Table 2.30. Many Italian verbs have two stem forms in the past indicative, distributed in the way indicated above. The formal differences between the forms are varied: *fec-fac* 'do', *coss-cuoc* 'cook', *rupp-romp* 'break', *vid-ved* 'see', *ebb-av* 'have', etc. If the formal differences between the two forms 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 pair of wug-forms (e.g. *mef-i* vs *pal-esti* for 1SG and 2SG respectively) would plausibly lend themselves to being mapped into unnatural morphosyntactic domains by adhering to the abstract pattern of Table 2.30 despite the total novelty of the alternation involved.

Table 2.30 Pattern of stem allomorphy in the Italian *passato remoto*

	SG	PL
1	A	B
2	B	B
3	A	A

However, most cases of morphomes (and most morphemic oppositions too for that matter) are not instantiated with such a wide array of forms. Consider the case of the Romance N- or L-morphomes. The number of forms associated with each of the patterns is usually relatively small. The Spanish N-morphome, 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 identified as 'N-morphomic' by language users, whereas other kinds of alternations, because of their exceptional nature within the system, would face a greater difficulty in fitting into the N-morphome. Consider the history of the Spanish verb *llevar* 'take'. In Old Spanish (Table 2.31), the verb was a diphthongizing one *levar–lievo*, in line with hundreds of other verbs in the language.

Table 2.31 Old Spanish verb *levar* in two different stages

	levar 'lift/take'		*levar* (after sound change)	
	IND	SBJV	IND	SBJV
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

At some point, however, a sound change /lje/>/ʎe/ occurred by which the former monophthong-to-diphthong alternation (/e/-/je/) was replaced by a consonant1–consonant2 alternation (/l/–/ʎ/). 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 high frequency of use of the verb, the alternation 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).

The history of these verbs (Table 2.32) shows that, sometimes (I would argue most of the time), the actual phonological instantiation of a morphome does matter a great deal. If a lexeme does not have the 'right' formal alternation,

Table 2.32 Modern Spanish outcomes

	llevar 'take'		*levar* 'lift' (an anchor)	
	IND	SBJV	IND	SBJV
1SG	llevo	lleve	levo	leve
2SG	llevas	lleves	levas	leves
3SG	lleva	lleve	leva	leve
1PL	llevamos	llevemos	levamos	levemos
2PL	lleváis	llevéis	leváis	levéis
3PL	llevan	lleven	levan	leven

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 displayed alternations between a mid and a high stem vowel (e.g. *pedir/pido* 'request', *cobrir/cubro* 'cover').

Both the *e/i* and the *o/u* alternating verbs followed the paradigmatic template shown in Table 2.33. It is, however, revealing, that, while the *e/i* alternation has been preserved robustly in the modern language, the *o/u* alternation has largely disappeared.

Table 2.33 Distribution of the high vowel stem in Spanish raising verbs

	PRS.IND	PRS.SBJV	IPF	PAST	IPF.SBJV I	IPF.SBJV II	FUT	COND
1SG	pido	pida	pedía	pedí	pidiera	pidiese	pediré	pediría
2SG	pides	pidas	pedías	pediste	pidieras	pidieses	pedirás	pedirías
3SG	pide	pida	pedía	pidió	pidiera	pidiese	pedirá	pediría
1PL	pedimos	pidamos	pedíamos	pedimos	pidiéramos	pidiésemos	pediremos	pediríamos
2PL	pedís	pidáis	pedíais	pedisteis	pidierais	pidieseis	pediréis	pediríais
3PL	piden	pidan	pedían	pidieron	pidieran	pidiesen	pedirán	pedirían

Figure 2.1 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* alternation was preserved, *o/u* alternations were lost to paradigm levelling. Largely in the 16th century, the high-vowel stem was generalized throughout the

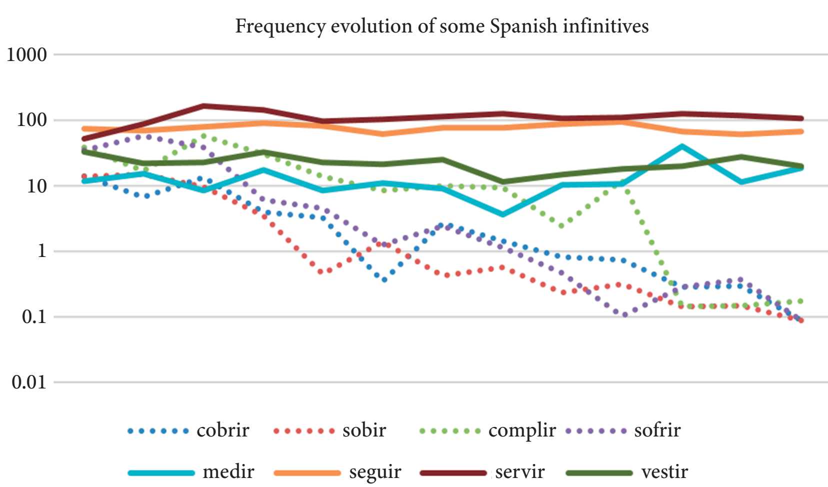

Figure 2.1 The demise of the o/u alternating verbs in Spanish

paradigm (i.e. *cubrir* 'cover', *subir* 'go up', *cumplir* 'fulfil', *sufrir* 'suffer' are the contemporary forms). As Figure 2.1 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 morphome have involved dissociating paradigmatic distributions from their concrete exponents, this is often[10] just a convenient fiction.

This, I believe, explains experimental results like those reported by Nevins et al. (2015), who 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-morphome verbs. Because the formal alternations presented to the Portuguese speakers did not match those of the L-morphome verbs in their language, they did not know what to make of a completely alien alternation. This is not, I believe, very surprising. In the same way as the history of *llevar*, and of *o/u* alternating verbs in Spanish, it reminds us that morphomic paradigmatic patterns (probably also morphemic and 'regular' patterns, see Albright 2002 and Albright and Hayes 2003) are most likely not independent from their actual formal instantiations.

2.5 Stem spaces

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 (see Trommer's definition in Section 1.4) is not enough when taken literally. Consider the 1SG.PAST and the 3PL.PAST in German in Table 2.34.

Table 2.34 Present and past-tense inflection of two German verbs

	machen 'do'				*singen* 'sing'			
	PRS		PST		PRS		PST	
	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

[10] 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 had 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 juegamos juegades jugan* to Spanish *juego juegas juega jugamos jugáis juegan*.

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. Common sense tells us, however, that we are clearly 'cheating' by analysing the exponence of the 1SG.PST and the 3PL.PST separately from its neighbouring cells. The other PST 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 singling out the 1SG.PST and 3PL.PST from other PST cells.

In cases like this, it is intuitively clear that the correct unit of analysis is the whole of the past-tense sub-paradigm. However, it is not always so straightforward. In some cases, cross-paradigmatic evidence can indeed single out a set of cells (e.g. because they, and only they, always share form across every single lexical item) without necessarily 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 cells in Spanish (see Table 2.35).

Table 2.35 Five selected paradigm cells from five Spanish verbs

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

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. In 'go', the stem in the infinitive/2PL.imperative is also used in the future. In 'have', by 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.

At the same time, there is an inescapable generalization, however, that these two cells, and only these, behave always in the same way regarding stem alternation. This is the reason why they are regarded as forming a so-called stem-space in Spanish (see Boyé and Cabredo-Hofherr 2006). Stem-spaces like this one are obviously closely related to the notion of the morphome and very interesting objects of morphological analysis. Unfortunately, they will be excluded, for definitional reasons, from any further consideration in this book. As mentioned in Section 2.3, the requirement will be upheld here that the set of cells constituting an alleged morphome must be identifiable by overt morphology within a single lexeme's paradigm. Cross-paradigmatically identified stem spaces, thus, will not be further examined in this book.

A clarification note seems appropriate, however, in relation to this. The distinction between morphomes and stem spaces is one that could well turn out to be relatively superfluous if the two phenomena/notions share most empirical properties apart from the definitional one. There is, for example, some evidence that, in the same way as morphomes, stem spaces can also constitute cognitively real grammatical entities for language users. For the stem space in Table 2.35, this is illustrated by a very common morphological change in substandard Spanish. For many speakers, the etymological form of the 2PL imperative is replaced by the form of the infinitive (*sed*>*ser*, *id*>*ir, tened*>*tener* etc.). As a result, the two cells (and only those two cells) become whole-word-syncretic and thus come to form an intra-paradigmatically identifiable morphological category 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 morphomes. Despite cases like the one just presented, the criteria used for stem-space identification and for morphome 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-morphome and PYTA respectively in morphomic literature. Be that as it may, in order to narrow down the object of study and to avoid a break with established terminology, the two concepts will be kept separate. Consequently, I reiterate that the requirement will be enforced throughout this book that a morphome be identifiable within a single paradigm by some overt form(ative) exclusive to it.

2.6 Cross-linguistic recurrence

One of the few points where linguists of quite different convictions (e.g. Koontz-Garboden 2016; Maiden 2016) seem to have surprisingly agreed so far is the claim (or maybe the theoretical stance) that morphomes 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 constitute 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 this general line of thought is understandable, there are some fundamental problems with it. The first is related to circularity. We cannot *claim* to have found out that morphomes 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 morphomehood (or our diagnostics thereof) dependent on typological uniqueness, this precludes any further empirical discoveries related to this, which 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 cognitive distinctions based on it.

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) is unique. Thus, if we require identity with respect to every detail and variable, all morphomes will, indeed, be typologically unique. However, as with other grammatical entities (consider the long-winded debate on comparative concepts and descriptive categories, see Haspelmath 2010; Newmeyer 2010), 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 morphome A and morphome 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 morphomes has usually been predicated of their paradigmatic distributions as a whole. Maiden (2018b: 167), for example, defines the N-morphome 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.'[11] He insists on the typological uniqueness of such a paradigmatic structure, and makes it clear elsewhere (2018b: 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 moods that a morphome spans (three in this case), and whether or to what extent a morphome is confined to particular

[11] Under closer scrutiny it becomes apparent that, in fact, his assessment of whether two morphomes are 'the same' or not is not driven so much by synchronic paradigmatic distributional concerns as by etymological (i.e. 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 that an N-morphome (root) is recognized as such when its form is regularly or analogically 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 colleagues') approach to 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 monograph. In typology, comparisons and assessments of 'sameness' and differences cannot and should not be done from an etymological perspective.

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 and from the database in Chapter 4 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 principled 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, by the current state of language documentation, or by the amount of linguistic diversity left in the Anthropocene, rather than by any inherent property of the patterns themselves.

2.7 Blocking

The theoretical notion of blocking might also be understood to have important ramifications in the definition and identification of morphomes. Blocking is a conflict-resolution principle often assumed to operate between mutually compatible morphemes or realizational rules (see e.g. Bonami and Stump 2016). It states that, in cases where two rules or morphs are in a subset–superset relation, the most specific one will take priority over the more general one.

Table 2.36 Past-tense forms of 'get' in Daga (Dagan, PNG) (Murane 1974: 63)

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

Consider the paradigm in Table 2.36. 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 all past cells is that other suffixes exist (*-aan*, *-en*, *-aton*, *-ayan*) that are more specific. The distribution of all forms, therefore, could be stated as the realization of morphosyntactic properties if we assume blocking. Things can get more complicated, however.

Table 2.37 Imperfective tense paradigm of Chaha (Semitic) 'open' (Völlmin 2017: 122)

	SG	PL
1	ə-kəft	nɨ-kəftɨnə
2M	tɨ-kəft	tɨ-kəfto
2F	tɨ-kəfc	tɨ-kəftəma
3M	yɨ-kəft	yɨ-kəfto
3F	tɨ-kəft	yɨ-kəftəma
Impersonal	yɨ-kəf^{w}cɨm	

In the paradigm in Table 2.37, the morphosyntactic distributions of the prefixes *yɨ-* and *tɨ-* are both unnatural. The two formatives crosscut, 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]), which would make blocking of an underspecified prefix *yɨ-* possible.

Regardless of the plausibility of this particular solution here, one can easily find cases in natural language where blocking is unmistakably not taking place. Observe the exponence patterns in Tables 2.38 and 2.39 (and also Janda and Sandoval 1984).

Table 2.38 Some Daai Chin (Sino-Tibetan) personal pronouns (So-Hartmann 2009: 140)

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

Table 2.39 Partial paradigms of two Khwarshi (Nakh-Daghestanian) nouns (Khalilova 2009: 66)

	'sibling'		'mother'	
	SG	PL	SG	PL
ABS	ɨs	ɨs-na-ba	ɨšu	ɨšu-bo
ERG	ɨs-t-i	ɨs-na-za	ɨše-t'-i	ɨše-t'-za
GEN1	ɨs-t-i-s	ɨs-na-za-s	ɨše-t'-i-s	ɨše-t'-za-s
LAT	ɨs-t-i-l	ɨs-na-za-l	ɨše-t'-ɨ-l	ɨše-t'-za-l

In Daai Chin personal pronouns, the plural formative *-e* appears in a subset of the cells where the non-singular formative *-nih* also appears. According to the blocking principle, this should not happen because *-e* should have prevented the appearance of *-nih*. To avoid this, it is always an analytical possibility to avoid segmentation (see Section 2.9) in these cases (i.e. to leave *-nihe* as an indecomposable plural suffix). Sometimes, this might seem an elegant solution, but at 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 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 difficult to find from a merely probabilistic/combinatorial 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 never get to see the more concrete exponent.

Despite its problems, blocking is a mechanism which is usually adopted, under one name or another (Superset principle, Elsewhere condition, Panini's principle, remnant syncretism, etc.), by every constructivist theory of morphology. The question to be asked, from the empiricist's perspective, is whether it constitutes a real cognitive principle employed by language users, or is instead, in the light of the above-mentioned ontological and empirical shortcomings,[12] 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, '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'.

[12] 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 suggest that an absence can be blocking the presence of an overt exponent.

Another morphological fact regarding Paninian blocking is that there are also many clear cases of formatives that are semantically compatible, and whose values are not in a subset–superset relation, but which still cannot appear together. Consider the incompatibility of dual subject *-k* and plural object *-dár*, and of durative *-tam* and masculine object *-rár* suffixes, in Nimboran (see Inkelas 1993). Those conflicts tend to be accounted for with reference to syntagmatic slots and 'position classes', where those affixes belonging to a same position class compete for a single slot and cannot both surface simultaneously.

Bauer et al. (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 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.' For her, then, blocking analyses are cognitively real(istic).

Given the deep uncertainties surrounding the status of 'elsewhere' forms and 'defaults', I will remain agnostic as for whether they constitute exponents different from the ones that cannot be captured by blocking. Because of the empirical focus of this book, 'surface' distribution will always be trusted over any supposedly underlying one. The same 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 and Sims 2010: 179), it is my firm conviction that a typological investigation should not rely on theoretical/formal notions of this kind.

2.8 Stem vs affix

It is fair to say that most of the research around morphomes 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 (e.g. blocking, discussed in the previous section) 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 and 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 and Nichols, 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.

Given that most of the suffixes in Table 2.40 are tense-neutral (e.g. trag-t vs trug-t), saying that the stem alternation pattern is triggered by the suffixes (Bickel and Nichols 2007: 186) does not seem to follow easily from the empirical data.

Table 2.40 German verb *tragen* 'carry'

	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

There is abundant cross-linguistic evidence that stem alternations can sometimes serve as the sole exponent of morphosyntactic distinctions. In a particularly striking case (Table 2.41), the verb 'give' in Iha changes its stem according to the person and number of the recipient.

Table 2.41 Verb 'give' in Iha (West Bomberai, New Guinea) (Donohue 2015: 413)

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

It is not difficult either to find cases of clearly segmentable affixes that fail to encode any consistent morphosemantic value. Consider the distribution of *-ni*, *-di*, and *-li* suffixes in Gourmanchéma presented in Table 2.25. These cases suggest that, unless it is programmatically incorporated as part of their definition, the distinction between stems and affixes has little to do with the presence or absence of grammatical meaning. In this book, therefore, stem or affixal status will not influence the assessment of morphomicity.

2.9 Segmentability

A property of prototypical formatives, and also of prototypical words, is that they are units which are relatively easy to discriminate/segment from surrounding elements. That is, in more technical terms, they are islands of syntagmatic predictability surrounded by peaks of unpredictability (see e.g. Mansfield 2021). A property of all the Spanish 1PL verb forms (*somos*, *fuimos*, *damos*, *amaremos*, etc.) is that their shared form is easily identifiable and segmentable by linguists and language users. 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 anything other than 1PL, since it appears always in that morphosyntactic context and never in other contexts. Its properties are thus not very different from grammatical words (e.g. a preposition like 'under' or a pronoun like 'you') which have abstract meaning. As argued by Pertsova (2007: 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 1PL *-mos*.

Deviations from this unproblematic case are not difficult to find, however. Problems with segmentability and mutually incompatible segmentations are well known (e.g. Bank and Trommer 2012; Blevins 2016: 26–8). Sometimes, the elements which can be identified on syntagmatic-transitional grounds alone are relatively clear, as in Wardaman in Table 2.42.

Table 2.42 Wardaman (Yangmanic, Australia) intransitive indicative prefixes (Merlan 1994: 125)

	SG	DU	PL
1EXCL	nga-	yi-rr	
1INCL		nga-yi-	nga-rr
2	yi-	nu-	
3	Ø-	wu-rr-	

Despite this (apparent?) segmentability, the morphosyntactic distribution of some of the resulting formatives (*nga-*, *yi-*, or *rr-*) is problematic, which by itself, according to some analyses (see e.g. the approach to segmentation in Trommer and 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, unclear.

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-morphome. According to the traditional analysis, forms like Spanish *vengo* 'come.1SG' or *tengo* 'have.1SG' are decomposable into the stems *veng-* and *teng-* and the 1SG suffix *-o*. O'Neill's (2015) segmentation proposal, 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 subjective in the absence of clear quantitative criteria, and largely irrelevant for the present discussion, as, in either case, we are left with a morphological element with an L-shaped distribution in the paradigm which we need to account for.

Despite the irrelevance of segmentation for morphomicity in many cases, formatives can sometimes depend on (debatable) segmentations. Those arising from very unorthodox ones are more exposed to being 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 when it concerns elements that are combinatorially treated consistently as whole objects, like Spanish *-mos*, compared to cases when a formal identity involves forms with an uncertain or a variable combinatorial status.

Table 2.43 Agreement prefixes in Xincan (Xincan, Guatemala) (Sachse 2010: 233)

	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-	Ø-	Ø-

In some agreement contexts in Xincan (Table 2.43), 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 find unnatural patterns of morphological identity practically everywhere (see Table 2.44).

Table 2.44 Two unorthodox segmentations in German and Spanish

	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

Thus, in German, on purely combinatorial grounds, /t/ is a formative (all by itself) in 3SG and 2PL but not (or not so certainly) in 2SG, where the suffix is usually taken to be /st/. Similarly, in Spanish, /s/ is a formative in 2SG but probably just part of a larger formative in the case of 1PL and 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 2.43 and 2.44. With this in mind, uncontroversial morphomes should be based upon forms which are easily discriminated (i.e. segmentable), syntagmatically, from the neighbouring phonological material. Thus, I will refrain throughout the present research from performing non-canonical segmentations like these, and will stick to the choices of the original descriptions.

2.10 Morphological zeroes

It is usually taken for granted that the distribution of formatives deserves analysis and explanation in morphology. The explanation offered may be different depending on whether or not such elements correlate 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 is common.

Consider the paradigm in Table 2.45. The morphosyntactic distribution of the form *hembua* (3, 1PL, and 2SG) is decidedly unnatural. 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 elsewhere. Therefore, the formal identity of the shaded cells in the paradigm of

Table 2.45 Orokaiva (Trans-New Guinea) far past indicative of *hembu* 'walk' (Baerman et. al. 2005: 26, after Healey et al. 1969)

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

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.

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 deep-rooted uncertainties surrounding zero in morphology (regarding both its status and its actual distribution in concrete cases), I remain agnostic in this book about its nature, and will refer only to overt-formative morphomes from this point on.

2.11 Economy

The economy of the analysis is a criterion that could also be plausibly used when assessing whether we are dealing with a morphome or not. Deciding between alternative (formal) analyses of a phenomenon is often difficult. In the simplest case, an analysis/formalization that covers 100% of the facts is preferable to one that does not. However, once two different analyses/formalizations cover the facts perfectly, it is difficult to decide which one is 'better' or more cognitively plausible. Discussion in these cases revolves usually around matters of 'elegance' and 'economy'. However, there is hardly any consensus as to how these notions should be understood and whether they favour one analysis or the other in specific cases.

In this section I will compare how different analyses and formal rules 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, I will assess how recourse to Paninian blocking and to autonomous morphological rules can impact the descriptive length of different systems. Consider first the inflectional patterns from Yagaria in Table 2.46.

Table 2.46 Allomorphy of Yagaria mood affixes (Stump 2015: 128, after Haiman 1980)

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

A total of eight other moods have been omitted from the paradigm here. These show the same patterns of syncretism as the moods displayed. They have been left out for the sake of brevity, and also because they involve the same formal alternations as the moods above. In addition, the glottal stop which appears in duals has been left out from the rest of the discussion because it does not make a difference between alternative 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[13].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 > -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
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
μINTER > -ve	μIND > -e	μSUB > -ma	μCOORD > -ga	μAPOD > -hine
λINTER > -pe	λIND > -ne	λSUB > -pa	λCOORD > -na	λAPOD > -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:

[13] Combinations of values like 'singular' and 'dual', 'dual' and 'plural', 'first' and 'second', or 'second' and 'third' 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) helps us reduce the number of rules needed but represents 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 (Table 2.47).

Table 2.47 Nivkh converb inflection (Gruzdeva 1998: 56; Nedjalkov and Otaina 2013: 40–42)

	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

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. Without any machinery whatsoever, the exponence mappings are as follows:

2/3SG > -r   1SG.NFUT > -t_1 PL.NFUT > t_2 1SG.FUT > -n_1 PL.FUT > -n_2

With Paninian blocking but without independently morphological rules:

Superset Principle

2/3SG.NFUT > -r_1 2/3SG.FUT > -r_2 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 diagonal syncretism in Table 2.48.

Leaving aside consonant gradation and the exponence of those cases that are not involved in the syncretism, we would need the following exponence rules in an analysis with no blocking and no autonomous morphology:

LOC.SG > -s   LOC.PL > -in_1 COM.SG > -in_2 COM.PL > -iguin

If we allowed blocking but not autonomous morphological entities:

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

And if we had to make use of autonomous morphology instead to capture the syncretism:

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

Table 2.48 North Saami (Uralic)
viessu 'house' (Hansson 2007)

	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

The relative economy (measured in number of mapping operations)[14] of the different analyses and formalizations depends, thus, on the degree of complexity (e.g., allomorphy) of the system. Figure 2.2 summarizes this.

It 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 Saami), an autonomous morphological analysis seems to be actually more uneconomical than the competing alternatives. This is the reason why a minimum requirement will be set for morphomic status in Chapter 4 that a pattern be instantiated with at least two different exponents.

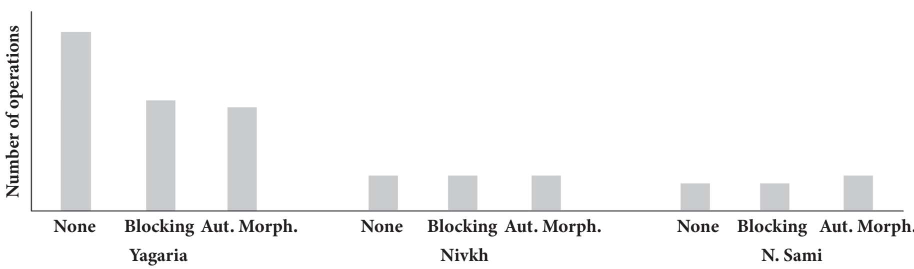

Figure 2.2 Comparison of the realizational economy of different analyses

[14] It is not evident by any means that this is the 'right' measure of realizational economy. One could think of alternative ones, e.g. the number of characters needed to represent the full set of rules.

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 and mapping operations 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 dark grey in Figure 2.3).

All this being said, it has to be recognized that there is no consensus in the discipline concerning what should count as more 'elegant' or more 'costly'. The operations that Figures 2.2 and Figure 2.3 count and lump together are of very different types, and we ignore 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 assume that all operations of the same kind should be equivalent. It has to be acknowledged, therefore, that we have absolutely no idea as to how/whether these considerations of formal economy and elegance of the analysis map onto language users' cognitive representations or onto actual psycholinguistic processing or production 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 have been discussed throughout this section. Judged by Figures 2.2 and 2.3, for example, there would be little reason to pursue an autonomous morphological analysis of the North Saami syncretism in Table 2.48, and yet it appears that in some dialects the pattern analysed here has spread to new contexts with different formatives (see Hansson 2007). This seems to suggest that language users did analyse the unnatural syncretism as systematic (i.e. morphomic) at some point. It remains to be understood (even imperfectly), therefore, how the factors that this section has dealt with guide the cognitive representations of inflectional patterns by language users. This is the reason why a typological investigation like the one in this book cannot rely on such factors for the identification of its object of inquiry.

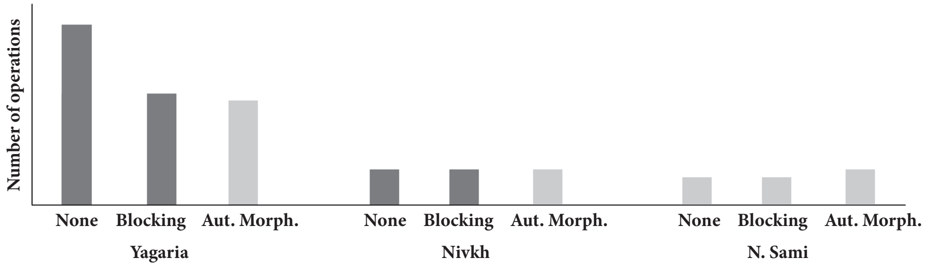

Figure 2.3 Comparison of the realizational economy and elegance of different analyses

2.12 Difficult cases

One of the facts that discussion around the morphome most urgently has to come to terms with is, as discussed in Section 2.2, 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 should be easier to tell, things that look morphosyntactically unmotivated at first sight may not always be straightforwardly so, as various degrees and sources of motivation are often possible. This section surveys a few problematic cases.

2.12.1 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. It is well known (e.g. Cysouw 2003; 2005) that especially 1PL and to a lesser extent 2PL are not straightforward plurals of 1SG and 2SG respectively. 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 usually treated as morphomic without further discussion (e.g. Baerman and Brown 2013; Stump 2015: 128).

Apart from the above-mentioned denotative affinity of 2SG and 1PL, there are other reasons to doubt that this might be the clearest 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.

Table 2.49 2SG/1PL morphological affinities in Papua

	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

As Table 2.49 shows, The 1PL/2SG syncretism is relatively common in Papuan languages. 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), and it can affect both agreement affixes (e.g. Ngkolmpu and Benabena) and pronouns (e.g. Suki and Yessan-Mayo) in genetically unrelated and geographically relatively distant languages.

Those cases where the 1PL shares exponence with the second-person as a whole are less clearly unnatural still. 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 infrequent (see Table 2.50).

Table 2.50 Some 1PL+2 morphological patterns

	***Darma ra* 'come' (Willis 2007: 350)**		***Mazatec* 'lay down' (Jamieson 1988: 106)**		**Aguaruna object agreement (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-	-Ø	-Ø

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, Otomanguean) and affixes (Aguaruna, Chicham) 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. Although they come close, the shaded cells of Table 2.50 are not reducible to the presence of the feature value 2. I will, consequently take these patterns as morphomic, 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 1INCL, 1EXCL, and 2PL all in different ways. English and other languages without clusivity conflate 1INCL and 1EXCL, and distinguish those from 2PL. However, the mirror-image of English also exists. A few languages do not have 1 as their core 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 1INCL and 2PL in an identical way and distinguish these from 1EXCL (see e.g. Sanuma (Yanomamic, Brazil) in Table 2.51). When some formative spans this addressee-centred plural complex and the 2SG, it may superficially appear that it has an unmotivated distribution (see Ojibwe (Algonquian)). However, there is, in these cases, a necessary and sufficient condition (reference to 2) that accounts for the distribution, which will thus be motivated and not morphomic.

Table 2.51 Some 1INCL=2PL paradigms (Cysouw 2003: 154–5)

<table>
<tr><th></th><th colspan="2">Sanuma non-emphatic pronouns</th><th colspan="2">Ojibwe intransitive prefixes</th></tr>
<tr><td></td><td>SG</td><td>PL</td><td>SG</td><td>PL</td></tr>
<tr><td>1EXCL</td><td>sa</td><td>samakö</td><td colspan="2">int-</td></tr>
<tr><td>1INCL</td><td>–</td><td rowspan="2">makö</td><td>–</td><td rowspan="2">kit-</td></tr>
<tr><td>2</td><td>wa</td><td>kit-</td></tr>
<tr><td>3</td><td colspan="2">Ø</td><td colspan="2">Ø</td></tr>
</table>

It has to be kept in mind that the use of the label 1INCL (as opposed to e.g. a label like 2INCL, 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. Cases like Sanuma *makö* or Ojibwe *kit-* should thus not be regarded as any more unnatural than English 'we' or Ojibwe *int-*.

2.12.2 Syntactically licensed morphomes

Traditionally, the term 'morphome' 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 also exist. Mielke (2008) represented a remarkable step in this direction in the domain of phonology. Less progress has been done in syntax, but it is safe to say that, also in that domain, unnatural classes can sometimes be the locus for particular operations or constructions. Take a look, for example, at the following sentences from Aguaruna (Overall 2007: 443–4):

10a)	ataʃu-**na**	yu-a-tata-ha-i	10b)	ataʃu	yu-a-tata-hi
	chicken-ACC	eat-HIAF-FUT-1SG-DECL		chicken	eat-HIAF-FUT-1PL
	'I will eat chicken'			'We will eat chicken'	

11a)

nĩ	ɨɨma-ta
3SG	carry.PFV.IMP

'You(SG) carry him!'

11b)

kutʃi	maa-ma-uhumɨ
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)

kutʃi-**na**	maa-aha-mɨ
pig-ACC	kill.HIAF-PL-RECPAST.3

'They killed a pig'

As illustrated by the sentences above, nouns or noun phrases in the object position in Aguaruna 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 entirely on the subject. This should, therefore, be described as a syntactic phenomenon. However, the set of subject values that 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. Although it is hard to assess without a targeted cross-linguistic exploration, cases like these might be relatively infrequent, but are by no means unique. Another comparable case comes from Marsalese (discussed in Corbett 2016: 82–3, from Cardinaletti and Giusti 2001):

13a)

Vaju	a	pigghiu	u	pani.
go.1SG	to	fetch.1SG	the.SG.M	bread

'I go to fetch bread.'

13b)

*Emu	a	pigghiamu	u	pani.
go.1PL	to	fetch.1PL	the.SG.M	bread

'We go to fetch bread.'

14a)

Vai	a	pigghi	u	pani.
go.2SG	to	fetch.2SG	the.SG.M	bread

'You(SG) go to fetch bread.'

14b)

*Iti	a	pigghiati	u	pani.
go.2PL	to	fetch.2PL	the.SG.M	bread

'You(PL) go to fetch bread.'

15a)

Va	a	pigghia	u	pani.
go.3SG	to	fetch.3PL	the.SG.M	bread

'(S)he goes to fetch bread.'

15b)

Vannu	a	pigghianu	u	pani.
go.3PL	to	fetch.3PL	the.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 this syntactic construction is available thus 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 paradigms in Table 2.52.

Table 2.52 Two (syntactically licensed?) morphomic patterns

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

If morphomes are defined as elements of form which are independent of other modules of grammar, the above morphological structures cannot possibly be considered morphomic. 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 *sensu stricto* independent from syntax, and cannot be said to be unmotivated in that sense. If we assume, as 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 externally motivated.

At the same time, it seems that excluding these elements from the ranks of morphomes would do violence to the whole enterprise. This is not how we usually think syntax ought to work. If anything, in cases like Marsalese, we would rather 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 morphomic pattern (the N-morphome) 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 an extramorphological correlate of this kind 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 book as bona fide morphomes, albeit conceding the problematic nature of these cases.

2.12.3 Gender or morphome?

I try throughout this monograph to define morphomes 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 morphome (or any other phenomenon really) is necessarily intertwined with our definitions of other phenomena and, in general, with 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 in the paradigm, and also on what we think other linguistic phenomena (e.g. gender) can be like.

Table 2.53 Gender–number affixes in Mian (Trans-New Guinea) (Fedden 2011: 163)

	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

Consider the agreement patterns in Table 2.53. Gender–number agreement inflection in Mian is clearly morphomic. The shaded affixes can appear, depending on the gender of the noun, in the singular, in the plural, in both values, and in neither. This seems therefore decidedly unnatural. However, there is an alternative analysis, which Fedden entertains, and discards as inferior to the analysis implied in Table 2.53. 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 2.53 would be simply the result of an over-articulated description of the language. If we have not identified correctly the relevant features and values involved, we cannot be surprised to find that the morphology operates at cross-purposes to the structure we have posited.

Consider Table 2.54. The distribution of the two allomorphs of the perfective positive appears unmotivated as laid out there. 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 instead 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 rest.

Table 2.54 Perfective positive suffix in Northern Akhvakh (Creissels 2008)

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

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 morphome in the language, as the exponence patterns displayed in Table 2.53 would be straightforwardly derived from the gender membership of the corresponding nouns. As gender (again as usually defined) is a purely syntactic feature, sensitivity to such a feature would never be labelled morphomic.[15]

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 on 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 intertwined nature of gender and number appears to be undesirable from a theoretical/logical perspective. Gender systems that are orthogonal to number and other features are preferred, and regarded as more 'canonical' cases of gender (Corbett and Fedden 2016). Because of this, cases like Mian (or like Romanian and German in Tables 2.55 and 2.56), in which the classification suggested by the forms deviates from orthogonality, are

[15] Note that, depending on our definition of 'morphomic' this is not at all unarguable. Gender membership is often (e.g. in French or German) arbitrary to a large extent and, apart from a few small semantic fields, relatively unpredictable on the basis of meaning. Gender membership, 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 morphome is also basically a list: a list of lexemes (in the case of inflection classes) or morphosyntactic contexts (in the case of metamorphomes) that only belong together because they share (some) inflectional properties.

most usually reported in terms of orthogonal features and values with abundant syncretism.

Table 2.55 Romanian definite articles (Gönczöl 2007: 30)

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

Table 2.56 German definite articles (partial paradigm)

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

Romanian shows how, to match our definition of gender, or of what gender can be like, values can be proposed in the absence of autonomous forms. Saying that, for some lexemes, the singular is masculine but the plural is feminine appears to be unacceptable[16] if we conceive of gender as a system of lexical classification. This was 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 do not usually say that *Auto*, for example, is no longer neuter in the plural. We say instead 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 morphome? Note that patterns similar to Mian, which offer alternative analyses, are not difficult to find.

The gender system of (unrelated) Burmeso in Table 2.57 seems strikingly similar to that of Mian. Three classes of nouns can be found in the language according to the forms they trigger in verbal agreement. It is the requirement of gender–number orthogonality that doubles the number of gender distinctions in the language. At

[16] The size of the class seems to make a big difference, however. The noun *arte* in Spanish, like the neuters in Romanian, behaves as masculine in the singular but as feminine in the plural, and yet linguists do not 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.

other times it is the interaction of gender with person that appears to lack the desired orthogonality.

Table 2.57 Conjugations in Burmeso (Donohue 2001: 100–102)

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-

As shown in Table 2.58, 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 and 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. The apparent 'deviation' from a canonical gender exponence can be greater.

Table 2.58 Subject agreement in Barasano (Tucanoan, Colombia) (Jones and Jones 1991: 73–4)

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

Table 2.59 Jarawara (Arawan, Brazil) possessor paradigm of 'arm' (Dixon 2004: 315)

	SG	PL
1	o-man-o	man-o
2	ti-man-o	man-o
3F	man-i	man-i
3M	man-o	man-i

According to Dixon's analysis, the 3PL pronoun in Jarawara controls feminine agreement (Table 2.59). This may be so, historically, because that pronoun might have grammaticalized from a noun meaning 'people', which may have been feminine originally. In addition, because of the agreement forms they trigger, Dixon conceives of the 1PL and 2PL pronouns as inherently masculine, regardless of the gender (M, F, or mixed) of their referents.

The agreement in some verbal paradigms in Omotic is similarly problematic. In Basketo (Table 2.60) and closely related Benchnon (Table 2.61), masculine singular and (most) plurals sometimes trigger the same agreement suffix, while first and second singular show the same form as 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 alleged origin of the forms, there seem few reasons to prefer such an analysis over one in terms of person–number agreement. For example, the pattern of syncretism of medial verbs displayed in Table 2.61 is contradicted by that found in final verbs (Table 2.62).

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 and 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 a person–number agreement system with syncretism, especially because such features are needed in the language anyway to account for the exponence patterns in other paradigms, such as the polar-question agreement suffixes (Rapold 2006: 218), which make the full set of distinctions (eight, as the pronouns).

Table 2.60 Basketo (Omotic) affirmative converb of 'know' (Hayward 1991: 536)

	SG	PL
1	ʔerer-a	ʔerer-i
2	ʔerer-a	ʔerer-i
3F	ʔerer-a	ʔerer-i
3M	ʔerer-i	ʔerer-i

Table 2.61 Benchnon (Omotic) medial verb agreement (Rapold 2006: 178)

<table>
<tr><th></th><th>SG</th><th>PL</th></tr>
<tr><td>1EXCL</td><td rowspan="2">-á</td><td>-á</td></tr>
<tr><td>1INCL</td><td>-ı̋</td></tr>
<tr><td>2</td><td>-á</td><td>-ı̋</td></tr>
<tr><td>3F</td><td>-á</td><td rowspan="2">-ı̋</td></tr>
<tr><td>3M</td><td>-ı̋</td></tr>
</table>

Table 2.62 Benchnon (Omotic) indicative final verb agreement (Rapold 2006: 179)

<table>
<tr><th></th><th>SG</th><th>PL</th></tr>
<tr><td>1EXCL</td><td rowspan="2">-ù</td><td>-ù</td></tr>
<tr><td>1INCL</td><td>-ù</td></tr>
<tr><td>2</td><td>-ù</td><td>-ènd</td></tr>
<tr><td>3F</td><td>-ù</td><td rowspan="2">-ènd</td></tr>
<tr><td>3M</td><td>-èn</td></tr>
</table>

The main point I try to convey, therefore, is that the orthogonality of features–values may not always appear to hold when one looks at the morphology in a paradigm. It is tempting to interpret the messiness of these patterns as a sign that the crucial feature or motivation for the exponents has been missed, as in the inverse system in Table 2.54. If we believe this is the case, meaningless features like gender could always be posited that would 'account for' any exponence pattern, even one like Daasanach in Table 2.63.

Table 2.63 Subject agreement of 'walk' in Daasanach (Cushitic) (Baerman et al. 2005: 106, after Tosco 2001)

<table>
<tr><th></th><th>SG</th><th>PL</th></tr>
<tr><td>1EXCL</td><td rowspan="2">seð</td><td>sieti</td></tr>
<tr><td>1INCL</td><td>seð</td></tr>
<tr><td>2</td><td>sieti</td><td>sieti</td></tr>
<tr><td>3F</td><td>sieti</td><td>seð</td></tr>
<tr><td>3M</td><td>seð</td><td>seð</td></tr>
</table>

In a way not entirely dissimilar to what we saw in Omotic, the two different forms upon which the agreement system is based in Daasanach apply to a heterogeneous list of morphosyntactic contexts. The form used in the masculine singular is also used in the 1SG, 1INCL, and 3PL. The form used in the feminine singular is the same that is used in 2 and 1EXCL. Presented in person–number terms, thus, this pattern appears to be as arbitrary as it can possibly get.

The alternative, as has been suggested in the literature for the previous patterns, would be to 'trust the forms' blindly and assume that 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 and Somali might argue against the latter analysis. The Daasanach paradigmatic arrangement illustrated in Table 2.63 appears to have originated from a full-fledged person–number agreement system in which sound change and phonological erosion have resulted in rampant syncretism (see Section 4.2.1.1 for more details). Structural reanalyses may occur, of course, so this origin is no guarantee that the Daasanach system should still be analysed synchronically in the same terms (i.e. with person and number features) as the agreement systems of Somali or Oromo. We can, however, conclude that the system has, at least, the same sound-change-triggered origin as some of the most prototypical morphomic patterns.

Lastly, and despite the efforts that here and elsewhere have been devoted to arguing for one of the two alternatives, I believe that analysing these patterns in terms of gender agreement or conceiving them instead as autonomous morphological syncretisms may not be 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 morphome have traditionally involved (e.g. Aronoff 1994; Round 2015). Cases like those presented throughout this section will therefore be considered morphomic here whenever they otherwise meet my definitional criteria for morphomehood.

2.13 What (else) can be morphomic?

In the cases that have been discussed in the current chapter, and in almost all the literature on the morphome, it is inflectional formatives which are discussed as the object of analysis. However, it is not only inflectional forms that may have unnatural distributions in the paradigm. Other morphological phenomena (e.g. syncretism, heteroclisis, defectiveness) can also apply differently in different parts of the paradigm and single out morphosyntactically unnatural sets of cells as their domain of application. At other times, derivational structures may also be thought of as paradigmatically organized and liable to display morphomic affinities. Orthogonal features and structures may also be found even outside paradigms, in which case unnatural patterns may arguably exist outside of them. This chapter explores the possibility of morphomic phenomena in less obvious domains.

2.13.1 Syncretism/feature sensitivity

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 here will thus involve whole-word syncretism.

Table 2.64 shows that syncretism is involved in two ways in 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*, *uɾcini*). On the other hand, the distinction between the tenses (i.e. present vs progressive) is only drawn within the morphome cells (i.e. *uɾo* vs *uɾca*), whereas the cells outside of it are underspecified for tense (i.e. *uɾciga vs uɾciga*). A different configuration can be found in Alpago Italian in Table 2.65.

Table 2.64 Partial paradigm of 'drink' in Mongo Daju (Dajuic, Chad) (Avilés 2008)

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

Table 2.65 Present-tense of 'sleep' in two Romance varieties of Italy

	Alpago (Zörner 1997)				**Standard Italian**			
	Indicative		**Subjunctive**		**Indicative**		**Subjunctive**	
	SG	**PL**	**SG**	**PL**	**SG**	**PL**	**SG**	**PL**
1	ˈdɔrme	dorˈmɔŋ	ˈdɔrme	dorˈmone	ˈdɔrmo	dorˈmjamo	ˈdɔrma	dorˈmjamo
2	ˈdɔrme	dorˈme	ˈdɔrme	dorˈmede	ˈdɔrmi	dorˈmite	ˈdɔrma	dorˈmjate
3	ˈdɔrme	ˈdɔrme	ˈdɔrme	ˈdɔrme	ˈdɔrme	ˈdɔrmono	ˈdɔrma	ˈdɔrmano

The cells constitutive of the N-morphome have become whole-word syncretic (/ˈdɔrme/) in this particular variety of Romance.[17] Thus, not only person and number but even the category of mood appears to be neutralized within the morphome in this paradigm. All distinctions continue to apply outside of the morphome 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 would be identified even in quite familiar places (see Table 2.66).

In both Balochi (Indo-European) and Standard German, the 2SG and PL person–number suffixes (an unnatural class) show syncretism between past and present. Similarly, various (un)natural classes of person–number values might be (in)sensitive to gender in Afro-Asiatic. In Kabyle (Berber), for example, gender agreement in the verb occurs in 3SG, 2PL, and 3PL (Naït-Zerrad 1994). In Mehri (see Table 4.87) and Arabic, by contrast, it is found in 2SG, 3SG, 3DU, 2PL, and 3PL,

[17] 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 diachronic emergence of whole-word syncretism (consider the typological parallel of Daasanach (Table 4.12) compared to Oromo and Somali (Table 4.13))

Table 2.66 Sensitivity of person–number agreement suffixes to tense

	Balochi (Axenov 2006: 164)				German			
	SG		PL		SG		PL	
	PRS	PAST	PRS	PAST	PRS	PAST	PRS	PAST
1	īn	un	an		e	Ø	en	
2	ay		it		st		t	
3	t	Ø	ant		t	Ø	en	

a different but similarly unnatural class. Syncretism and sensitivity to a particular feature like tense or gender can thus have morphomic distributions. Although Kabyle constitutes an exception, these patterns seem to be subject to a tendency to have more distinctions/allomorphs in more frequent values (see Table 2.70 for the frequency of different person–number cells), and in those that cannot be easily inferred from context (see Milizia 2015, Storme 2021).

2.13.2 Heteroclisis

Similarly 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.

Table 2.67 Pattern of heteroclisis of Czech *předseda* 'president' (Stump 2006: 290)

	'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
INS	ženou	ženami	předsedou	předsedy	filosofem	filosofy

This is the case of Czech *předseda* (Table 2.67), which behaves as a hard feminine noun in the NOM, GEN, ACC, VOC, and INS cases in the singular, and as a hard masculine elsewhere. This, could be therefore described as a morphomic pattern of heteroclisis.

The link between heteroclisis and more traditionally morphomic phenomena (e.g. stem alternations) is better known from the literature on Romance (see e.g. Maiden 2018b: 55, 220). Thus, particular morphomic stems (e.g. PYTA) may have a particular inflectional class (e.g. non-first conjugation) associated with them. When the inflectional class membership of the lexeme elsewhere differs, this results in heteroclisis. 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, that same pattern of heteroclisis is found in the absence of stem alternation (Table 2.68).

Table 2.68 Some inflectional forms in Spanish

	Conjugation I, 'love'	**'give'**	**Conjugation II, 'run'**
Infinitive	am-ar	d-ar	corr-er
2SG.PRS.IND	am-as	d-as	corr-es
2SG.PRS.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.SBJV	am-aras	d-ieras	corr-ieras

In the Spanish verb *dar* 'give', the unnatural paradigm subset known as PYTA is singled out by heteroclisis alone, instead of by stem allomorphy, thus constituting another example of morphomic heteroclisis. Similar cases are common in Romance. In Portuguese, for example, second-conjugation *v-er* 'see' is conjugated in the third conjugation in the same tenses (e.g. Pt. *v-er v-endo v-emos v-eria* vs *v-isse v-ira v-iste v-imos*).

2.13.3 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 indeed been shown to be the case (see e.g. Albright 2003 and Maiden and O'Neill 2010). In this section I will briefly present the issue in connection with the Spanish L-morphome.

In the paradigmatic domain of the L-morphome, near-suppletive stem alternations (e.g. *cab-er/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*, *salir/salgo*, *conocer/conozco*), alternation (or lack thereof) is just lexically stipulated. In verbs which are infrequent but which are of a phonological structure which never shows 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 include the velar augment (e.g. *engrandecer/engrandezco*, *palidecer/palidezco*), and so there is also no uncertainty for verbs belonging to this large (300+) class, despite their low token frequency.

The problem arises when the verb is not of this class, is infrequent, and is of a phonological structure which seems that it could maybe require an L-morphomic exponence. In some of those cases, normative grammar either prescribes one of the two possibilities (e.g. *mecer* 'rock' does not alternate according to Real Academia Española but *pacer* 'graze' and *asir* 'grab' do) or offers two or more correct alternatives (e.g. for *roer* 'gnaw', the forms *roo*, *roigo*, and *royo* are all accepted as the 1SG.PRS.IND, and for *yacer* the same applies to *yazgo, yazco*, and *yago*).

Despite the recommendations of prescriptive grammarians, the truth is that, whenever this uncertainty exists for a lexeme, speaker choices vary: nonstandard forms like *paza* (without the velar augment) or *mezca* (with the augment) are found alongside the prescribed variants *pazca* and *meza*. They constitute cases of overabundance which extend, as expected, to every cell within the L-morphome (see Table 2.69).

Table 2.69 L-morphome overabundance in two Spanish verbs, partial paradigms

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

In my opinion, however, the most accurate usage description would be that, because of the uncertainty they face in these paradigm cells, language users tend to avoid the forms altogether in those seldom-used verbs whose stem(s) are not sufficiently entrenched in the lexicon. It seems that—somewhat paradoxically, since they are definitionally opposite phenomena—the border between overabundance and defectiveness is fuzzy here.

As shown in Table 2.70, the frequency of forms within the L-morphome usually amounts to around 10% of the surveyed tenses. By contrast, in the case of those verbs[18] with L-morphome overabundance, those forms represent less than 1%. Overabundance and defectiveness therefore affect all the cells within the L-morphome in the same way, which confirms the deep morphological affinity of these forms in synchronic grammar, even in cases of non-canonical (i.e. no or multiple) inflectional exponence.

Table 2.70 Token frequency proportion in the two groups (percentages)

	L-morphome-overabundant verbs				**All verbs**			
	PRS IND	**PRS SBJV**	**IPF**	**PAST**	**PRS IND**	**PRS SBJV**	**IPF**	**PAST**
1SG	0.16	0	2.11	0.14	4.51	0.01	8.55	2.08
2SG	0.85	0	0.33	0.01	2	0.4	0.18	0.14
3SG	30.15	0.33	29.33	13.82	32.65	3.2	11.73	12.8
1PL	0.58	0	0.07	0.18	2.9	0.24	0.4	0.23
2PL	0.01	0	0	0	0.18	0.04	0.02	0.01
3PL	9.40	0.12	10.71	1.70	10.5	1.41	2.79	2.84

2.13.4 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 the existence of at least two orthogonal dimensions/features in a paradigm is necessary to identify unmistakable cases of morphomicity (i.e. affinities which are morphosyntactically unnatural regardless of any hypothetical feature structure one might posit—see Section 2.2). For this practical reason, the focus of this book 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 here before, the lexicon is full of cases where a resonance does not correspond straightforwardly to any shared semantics (e.g. *deceive*, *receive*, *conceive*). In many cases the formal similarities may be accidental and grammatically irrelevant. In other cases, however, there is evidence that those 'resounding' elements must constitute a grammatical unit, despite the lack of semantic content. Words with those bound stems, for example, can sometimes share unpredictable morphophonological processes in word formation (*deception*, *reception*, *conception*, etc.). There is psycholinguistic evidence (Giraudo et al. 2016) that these words prime one another beyond what

[18] The token frequencies of the verbs *mecer* 'rock', *asir* 'grab', *yacer* 'lie', and *roer* 'gnaw' have been surveyed in the corpus Corpes XXI as representatives of the group of L-morphome-overabundant verbs.

the shared form would account for, thus suggesting a deeper cognitive affinity of some sort. The concept of the morphome can also be useful, therefore, for lexical organization, and in derivation. Exploring, for example, the domain of terms related to ethnicity, Schalchli and Boyé (2018) find evidence (see Table 2.71) for systematic syncretisms like those usually described as morphomic.

Table 2.71 Some French terms related to ethnicity (Schalchli and Boyé 2018)

	Ethnicity	Area	Language	Ethnicity	Area	Language
Noun	*français*	*France*	*français*	*russe*	*Russie*	*russe*
ADJ	*français*	*français*	*français*	*russe*	*russe*	*russe*

The decision to focus on inflectional paradigms here is to be understood, therefore, as a way of narrowing down the object of study of the present book, and not as an advocation for morphomicity or paradigmatic structure, being exclusively inflectional phenomena.

2.13.5 Morphomicity in syntagmatics

Another domain where unnatural classes have received little attention concerns the syntagmatic order of sub-word elements. This has a prominent role in morphology and can also adopt natural and unnatural distributions in the paradigm. Consider Table 2.72.

Table 2.72 Two tenses of the verb ‘wash’ in Fula (Atlantic-Congo) (Arnott 1970: 191–2)

	Relative past passive		Subjunctive passive	
	SG	PL	SG	PL
1EXCL	lootaa-mi	min-lootaa	mi-lootee	min-lootee
1INCL	–	lootaa-ɗen	–	lootee-ɗen
2	loota-ɗa	loota-ɗon	loote-ɗaa	lootee-ɗon
3	ʻo-lootaa	ɓe-lootaa	ʻo-lootee	ɓe-lootee

Cumulative person–number affixes encode subject agreement in Fula unambiguously. While most frequently, and canonically, morphs indexing the same argument or feature would be expected to occur in the same syntagmatic slot (see Mansfield et al. 2020), this is not the situation in Fula (nor in many other languages: see Crysmann and Bonami 2016; Herce et al. forthcoming). Morphs for 1.EXCL and 3 appear as suffixes (light grey), those for 2 and 1.INCL are prefixal (dark grey), and 1SG *mi* can appear in either position in different TAMs. The syntagmatics of person–number markers in Fula can thus be described as morphomic, in that their syntagmatic position does not match any natural class.

More work needs to be done to explore the properties (i.e. their cross-linguistic recurrence, typology, diachronic resilience, learnability, their role in analogical change ...) of unnatural patterns in the less obvious domains presented in this section. Looking particularly at unnatural patterns in syntax and the lexicon would be interesting to see how these compare to traditional morphomes, and whether they can be considered part of the same broad phenomenon. This will be left for the future, however; this book will focus on the traditional domain of morphomic exponence: shared morphology in inflectional paradigms.

3
Morphomes 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.'

In linguistics, language change is also taken to be one of the main sources for true explanation. The case of morphomes is somewhat exceptional in that, here, diachrony has come to be almost embedded into the very definition of the phenomenon. Morphomes (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 little access 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 morphome 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 be stored in the lexicon separately, or constitute mere 'diachronic junk' with no synchronic grammatical import.

Although these discussions might make sense in finer-grained philological research, I believe they have no place in a broader typological endeavour like this book. It is not only impractical but also unreasonable to define or diagnose a synchronic grammatical phenomenon diachronically. Diachrony and morphomehood will thus be regarded as independent here, which will allow us to scrutinize and typologize the different ways in which morphomes may arise, change, and disappear from a language. This will be the purpose of the following section.

3.1 The emergence of morphomes

3.1.1 Sound change

The morphologization of sound changes and their paradigmatic effects is probably the first process that comes to mind when one thinks of the possible diachronic sources of morphomes. This is the ultimate[1] origin of most of the morphomes which have been discussed in the literature (e.g. the renown N-, and L-morphomes of Romance).

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

The Typological Diversity of Morphomes. Borja Herce, Oxford University Press. © Borja Herce (2023).
DOI: 10.1093/oso/9780192864598.003.0003

3.1.1.1 Morphological result of sound change

The label 'sound change', however, can refer to different processes of morphome emergence. Sometimes, as in the classical Romance morphomes N and L, sound changes, in conjunction with different phonological environments, generated alternations where there were formerly none. Consider, similarly, the cases in Tables 3.1, 3.2, and 3.3.

Table 3.1 The verb 'get tired' in two stages in Jabuti (Macro-Je) (Pires 1992: 45–6)

	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 3.2 The verb 'drive' in two stages in German (Braune and Reiffenstein 2004)

	Pre-Old High German		Modern German	
	SG	PL	SG	PL
1	far-u	far-em	fahr-e	fahr-en
2	*far-is	far-et	fähr-st	fahr-t
3	*far-it	far-ant	fähr-t	fahr-en

Table 3.3 Aorist past-tense of 'tie' in different stages of Greek (Holton et al. 2012)

	Ancient Greek		Modern Greek	
	SG	PL	SG	PL
1	'e-dēsa	e-'dēsamen	'e-desa	Ø-'desame
2	'e-dēsas	e-'dēsate	'e-deses	Ø-'desate
3	'e-dēse(n)	'e-dēsan	'e-dese	'e-desan

In Jabuti (Table 3.1), an originally non-alternating stem split into two different stems as a result of sound changes involving intervocalic voicing plus certain subsequent changes in point and mode of articulation. In German (Table 3.2), anticipatory distant vowel assimilation to a following /i/ (i.e. i-umlaut) created stem-vowel apophony from scratch. In Greek (Table 3.3), in turn, a past-tense prefix was deleted in unstressed pretonic contexts.

These three conditioned[2] sound changes show various additional differences in their details and subsequent development. For example, the phonological environment that gave rise to the alternation is still in place in Jabuti (and arguably in Greek) but has disappeared in German. Although the forms can be said to be completely morphologized in all cases (because the formal alternations are no longer synchronically productive phonological processes), only in the latter case (i.e. in German) can the new alternation potentially become informative and participate non-redundantly in the system of morphological contrasts in the paradigm. Note, in this respect, that the sound-change-triggered alternation /a/ vs /e/ has now become the only trait distinguishing 3SG and 2PL present in many German verbs like *fahren*.

Despite their differences, in Jabuti, Greek, and German, sound change has generated from scratch an alternation between two formerly identical forms. I will call this type of morphome origin the **morphological divergence** scenario. The research undertaken here for the compilation of the morphome database in Chapter 4 has demonstrated this to be the most common origin of morphomes cross-linguistically (see Ayoreo, Daasanach, French, Kele, Iraqw, or Saami for morphomes of comparable diachronic origin).

These cases where sound change creates morphomes by generating morphological variation or alternations from scratch (i.e. AA>AB) contrast to the opposite cases where sound change leads to a **morphological merger** (i.e. AB>AA) instead. In Livonian (Table 3.4), for example, comparative evidence suggests that morphological syncretism between 1SG and 3SG derived from a sound-change-generated conflation which was extended analogically.

Table 3.4 The verb 'kill' in two Finnic languages (Baerman 2007a)

	Estonian				Livonian			
	PRS		PAST		PRS		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

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

[2] 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. Although I know at present of no example of a morphome arising from an unconditioned sound change, this is entirely possible logically. If a phoneme's new pronunciation merges with that of a pre-existing one, this could 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 3.5 and 3.6 for comparable cases).

(1SG.PAST and 3SG.PAST) became identical in Livonian. This accidental formal conflation was analysed as systematic by language users and was subsequently extended to the present, where the two forms would not have become syncretic by regular sound change (see how 3SG *-b* must have spread in Livonian to 1SG). The accidental formal merger of formerly distinct forms as a consequence of sound change is, therefore, another possible source of unnatural syncretisms.

Another revealing example of this type of morphome emergence can be found in the history of Scandinavian. The infinitive and the 3PL present forms must have been 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 (see Table 3.5).

Table 3.5 Indicative mood inflection of 'drive' in two stages of Germanic (Zoëga 1910)

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

This arbitrary morphological identity, however, seems to have been actively preserved in diachrony and even to have extended occasionally to other forms. *Preterito presentia*, for example, because of their use of etymologically past forms in the present, should never have developed a syncretism of 3PL.PRS and INF (consider the paradigm of *eiga* in Table 3.6). However, probably 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.PRS *-u* suffix of these verbs into the infinitive. Thus, for example, *skulu* 'owe/have to' (also *munu* 'will') was not only the 3PL.PRS but also the INF form in Old Norse.

Table 3.6 Indicative inflection of two preterite-present verbs in Old Norse (Zoëga 1910)

	'own', infinitive: *eiga*				'owe', infinitive *skulu*			
	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	eigu	átti	áttu	skal	skulu	skyldi	skyldu

Other preterite-presents like *eiga* (see Table 3.6) or *vita* 'know' kept the 'mismatch' between an infinitive in *-a* and a 3PL.PRS in *-u* 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* (the infinitive form) 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 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 and verbs. In verbs that are used mostly in auxiliary modal contexts, for example, the infinitive form may have been too infrequent to provide an analogical model for analogy. More philological work would be needed to evaluate frequencies in historical corpora, and the developments in historical (e.g. Old Swedish) and modern (e.g. Elfdalian) varieties.

3.1.1.2 Paradigmatic locus of sound change

In an orthogonal contrast to its morphological results, sound-change-generated morphomic structures also differ with regard to another aspect. The sound change that gives rise to them can take place in different *loci* with respect to the resulting morphome. Change can target the paradigm cells constitutive of the morphome or can instead target their complement set. These two scenarios are not mutually exclusive since, sometimes, the sound changes that create a morphome may happen both in the morphome cells and in their complement set.

A well-known but particularly appropriate example of this last scenario is the L-morphome of Romance. Its emergence can be traced back to two independent sound changes. One involved the palatalization of velars before front vowels (see *nascer* in Table 3.7) and the other the palatalization of non-labial consonants before /j/ (see *medir*). Because front vowels and yods were in complementary distribution in the paradigm (e.g. 'do': *fak-jo*, *fak-is*, *fak-it*), 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.

Note that the shaded cells of *nascer* in Table 3.7 are those where palatalization (i.e. /naskes/>nas[ts]es) 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, a minority alternant within their

Table 3.7 Two verbs illustrative of the Romance L-morphome (Herce 2019a: 113)

	Old Spanish *nascer* 'be born'				Portuguese *medir* '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

paradigms, which is probably the reason why these cells, rather than their complement, are the ones which are taken to constitute a morphome. See the case of Svan in Section 4.2.2.13 for another morphome with a possibly similar diachronic origin. For morphomes created by sound change(s) in the morphome cells, see those of Chinantec (4.2.5.5) and Pite Saami (4.2.3.11), and for those created by sound change in the morphome's complement cells see e.g. Luxembourgish (4.2.3.9) and Wutung (4.2.4.20).

3.1.1.3 Zero as a source of morphomes via sound change

The arbitrary nature of the linguistic sign, promulgated most famously by Saussure, is one of the most celebrated axioms of linguistics. Although onomatopoeia, phonaesthemes, and other phenomena are known not to conform to this arbitrariness (see also Blasi et al. 2016), the core areas of grammar (e.g. the expression of morphosyntactic values in inflection by concrete 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 morphomes in general. If every form–meaning association is equally possible (e.g. 2PL=/i/, 2PL=/pu/, 2PL=/ar/, 2PL=Ø), 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 use frequency and length of expression. Put simply, more frequent words and meanings tend to be shorter. This is known as Zipf's (1935) 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 2021). Thus, 3 will tend to be shorter or unmarked compared to 2, SG shorter or unmarked compared to PL, etc. This means that zero will tend to appear preferably in certain values within the paradigm (see Table 3.8). Some of these more likely distributions are usually

Table 3.8 Some frequency-expected distributions of zero

	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	Ø						Ø		Ø	
2	Ø						Ø			
3	Ø		Ø	Ø	Ø		Ø	Ø	Ø	Ø

Note: See the frequencies provided in Table 2.70 for the approximate relative frequencies of the different person–number cells. In general: 3SG>3PL>1SG>1PL>2SG>2PL. That is, 'singular' is the most frequent number value and 'third' tends to be the most frequent person value. Because of this, inflectional patterns where SG, 3, 3SG, and SG+3 are zero/unmarked are not unexpected from a Zipfian perspective. The fifth pattern in Table 3.8 (1SG+3) is also not unexpected, since zero characterizes the 3 most frequent person–number combinations.

considered possible for the meaning side of lexical entries. Others (e.g. SG+3PL, 3+1SG) would count as morphosyntactically unnatural.

This is important because run-of-the-mill sound changes can and frequently do transform zero vs affixed configurations into morphomic A vs B configurations. These, maybe unlike zero,[3] need to be learned in some way, and can fulfil the criteria for morphomehood that I have set out in Section 3.2. One such case (Jabuti in Table 3.1) has already been presented here, and conforms to one of the paradigmatic distributions of zero assumed to be relatively more common due to Zipf's law. However, and because of the relative unpredictability of zero, all sorts of morphomic patterns are attested to derive from zero vs affixed.

All the formal alternations in Table 3.9 (i.e. /a/ 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 made Zipfian sense, since it characterized the most frequent number–case cell. In Wutung (Sko, New Guinea), the paradigmatic distribution of zero is more arbitrary. In English, the distribution of zero could well be said to be completely unexpected from a Zipfian perspective (it qualifies indeed as a typological *rarissimum*, see Plank

[3] 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 morphological objects on a par with overt affixes. Although morphologists often allow (or force) zero to participate in exponence rules in the same way as other morphemes (e.g. blocking other overt formatives, see Pertsova 2011), I believe 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 therefore may 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 morphomes in this book have been defined over overt formatives, and never over zero or whole-word syncretism by itself.

Table 3.9 Partial paradigms showing zero-derived alternations in various languages

	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:
INS	imenem	imenami	3	mua	mua	3	dʌz	du:

and 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_). Other times (e.g. in Russian), it is the affixed forms that are conservative since, in that position, the stem was 'protected' from changes that affected the unmarked cells: _imja_ < _jimę_ (Proto-Slavic) < _*inʔmen_ (Proto-Balto-Slavic) < _*h$_1$nómn̥_ (Proto-Indo-European) (Derksen 2007: 212).

As the above cases illustrate, overt morphomic alternations are often derived via sound change from former morphological zeroes. If the paradigmatic distribution of morphological zeroes is not random, which seems probable (consider Zipf 1935), this is likely to bias the properties of later morphomes, even of those that emerge via sound change. At the same time, cells that share simply a morphological zero can also be singled out as 'the same thing' by language users, which may give rise to a 'morphological niche' (Aronoff 2016) for the purposes of analogical change (see Section 3.1.3) or grammaticalization (see Section 3.1.5 and Bantawa in Section 4.2.2.2).[4] Although special reference to zero will not be made in those sections, it is something to be considered in other diachronic sources as well.

3.1.2 Semantic drift

Another, relatively well-known source of morphomes is the disintegration due to semantic drift of formerly natural classes. This is the origin of the renowned

[4] Bantawa and Athpariya show how particular formatives can intrude into those specific paradigm cells that are characterized by zero. Zero-marked cells, therefore, despite not sharing overt morphology (and thus not meeting the definitional requirements for morphomehood that I have set out here), can also sometimes provide a template for the distribution of incoming morphological elements. This suggests that they can have some morphomic properties under the right circumstances. It is a matter for future research to assess to what extent the properties of zero-based morphological affinities resemble those of overtly marked morphomes.

PYTA morphome 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 concerned aspect. Observe the Latin forms in Table 3.10.

Table 3.10 3SG forms of 'make/do' in various Latin tenses (Maiden 2011a)

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

As Table 3.10 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 (2011a) 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, in Table 3.11, the Spanish descendants of the above tenses, and their semantic content as reflected by their label.

Table 3.11 3SG forms of 'make/do' of various Spanish tenses

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

The set of tenses that could be classified as perfective in Latin (shaded in Table 3.10) 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 no common thread of meaning or function extending across this set of tenses in modern Spanish in contradistinction to other tenses. The inherited morphological affinity of the various former perfective tenses, however, has often been preserved, which makes morphological structures like this one morphomic.

For reasons related to feature–value orthogonality (to be presented in Section 4.1.1), morphomes like Spanish PYTA—i.e. so-called TAM morphomes

(Smith 2013) where the morphological allegiances relate to whole tenses—have not been included in the morphome database of Chapter 4, which makes it difficult to assess the relative prevalence of semantic drift in the creation of morphomes cross-linguistically. My overall impression is that this process might be comparatively rare as the force responsible for single-handedly creating morphomic structures, and it is certainly less common than sound change. Although it is not uncommon for semantically motivated forms to break free of their natural-class constraints,[5] this often happens in ways different from semantic drift (see Sections 3.1.3.1 and 3.1.5).

3.1.3 Analogy

Analogy is a term used so widely in linguistics, to mean so many different things, that it is impossible to explain it at any length within the confines of this short section (for more specialized treatments see e.g. Blevins and Blevins 2009 and Gaeta 2010). The word 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 their language's grammatical system. I take these changes to be copying errors in language transmission 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, deduce a grammatical system that differs slightly from that of their elders. It 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.

3.1.3.1 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 borrowed from a neighbouring cell (i.e. from a cell with which the form has a particularly close morphosyntactic relationship due to shared values). It may appear intuitively contradictory for morphosyntactically driven analogies to be able to result in morphomic patterns. However, they may do so when they involve the extension of some of the forms inside a natural class but not others.

[5] E.g. a realis/irrealis distinction becoming morphomic in Sye (Crowley 1998), or a past/non-past distinction becoming morphomic in Northern Talysh (Kaye 2013).

Consider the partial paradigms in Table 3.12, where 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 3.12 (see also Biak in Section 4.2.4.3), should probably be understood as manifesting the loss of number and tense distinctions in some (infrequent) contexts. The particular feature 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.

Table 3.12 Natural syncretisms resulting in unnatural morphological patterns

Slovene *človek* 'man' (Herrity 2000: 49; Baerman et al. 2005: 175)			**Gévaudan Occitan *cantar* 'sing' (Ronjat 1930; Camproux 1958)**		
	DU	**PL**		**PRET**	**IPF.SBJV**
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

Morphomes may originate by morphosyntactic analogy both from morphemic (i.e. natural-class distributed) formal elements, as in the cases above, and also from morphomic (i.e. unnatural class distributed) forms. Morphosyntactic analogical processes can therefore modify the paradigmatic extension of morphomic structures without bringing forms back to the realm of morphemes. Consider the change in Table 3.13.

Table 3.13 Possessive inflection of *muuka* 'head' in Wambisa (Peña 2016: 467)

	Pre-Wambisa		**Wambisa**	
	SG	**PL**	**SG**	**PL**
1	muuka-ru	muukĩ	muuka-ru	muukĩ
2	muukɨ-mɨ	*muukɨ-mɨ	muukɨ-mɨ	muukĩ
3	muukĩ	muukĩ	muukĩ	muukĩ

The tendency to level plural forms is morphosyntactically understandable and is documented in various different languages.[6] The formal levelling within the natural class ‘plural’, however, did not result in a natural morphological pattern in Wambisa because of the pre-existing syncretism of 3SG and 3PL. The morphome database in Chapter 4 suggests that developments of this kind are not uncommon. Although it might be difficult to go beyond impressionistic claims in this respect, it looks as if analogical changes operating on morphomic structures often seem oblivious to the status of those structures, and not generally aimed at bringing the forms in line with a natural class. See the morphomes of Nen (Section 4.2.4.14) and Italian and Servigliano (Section 4.2.3.8) for other morphomic structures that have been changed by morphosyntactically driven analogy but have stayed morphomic.

3.1.3.2 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 yet 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 case in Table 3.14 (also dealt with in Table 2.33). 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 pre-existing 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 pre-existing formal similarity of these two word forms has provided the motivation for the analogical change

[6] 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 *-aþ*, which later spread to the 1PL).

Table 3.14 A selection of word forms in different Spanish verbs (I)

	'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

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 affinity in the stem provides the motivation for the identity of affixes, seem not to be infrequent. See the diachronic insights on Yakkha in Herce (2021a) or the morphome of Girawa in Section 4.2.4.6) for other morphomic structures with a similar origin.

The locus of the formal similarity that provides a motivation for formally driven analogy, and the 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 analogical developments in Table 3.15, also in Spanish.

Table 3.15 A selection of word forms in different Spanish verbs (II)

	'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**

In some non-standard varieties, the stem of the gerund is replaced by the stem used in the so-called PYTA tenses (the 3PL past is provided in Table 3.22 as a representative of these cells). Thus, *poniendo* changes to *pusiendo* analogically, and *haciendo* changes to *hiciendo* (Pato and 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 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 (compare *est-a-r* vs *estuv-ie-ron* to regular *cant-a-r* vs *cant-a-ron*). Thus, the tonic suffix /je/ always selects the PYTA root, except in the gerund forms of some verbs like 'put' and 'make'. By extending the former perfective root to the gerund, these analogical changes remove this exception. Note, however, that in the process, a systematic morphological identity has been created between cells that have no particular morphosyntactic affinity.

Morphomes, 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, so the stems in one and the other were sometimes different (e.g. *krta-/čartanaiy* 'die'). However, the accidental formal resemblance of the infinitive and the past-tense forms provided the motivation for the systematic analogical extension of etymologically 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 in modern descendants like Persian (see Bonami and Samvelian 2009: 28) and Balochi (Axenov 2006). The formal alternations between PAST/INF and other word forms 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'; see Axenov 2006), so that the systematic nature of the morphomic affinity seems clear.

3.1.3.3 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 the syncretism in Table 3.16.

It seems clear that 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/motivated the analogical change. Thus, classification

Table 3.16 Weak masculine declension endings in Old English (Bazell 1960: 3)

	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

into the two types of analogy identified here is not to be understood as mutually exclusive.

3.1.4 Pattern interactions

Another way in which morphomes can emerge in a language is by means of the conflict or interaction between different patterns of allomorphy distribution.[7] These patterns can be morphomic or morphemic. For straightforward predictability relations to hold between pairs of cells in a paradigm, it is necessary for forms to be distributed in the same way across lexical items. This could even be thought of as the *raison d'être* of morphomic patterns. When two different patterns crosscut 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 the Romance L- and N-morphomes (Table 3.17).

Table 3.17 N- and L-morphomes and their paradigmatic distribution

	Spanish 'understand'		Spanish 'put'		Ansotano 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

The cross-cutting distributions of the N- and L-patterns in the paradigm give rise to four different areas in the paradigm (see the paradigm of 'come' in

[7] This section draws on the data and arguments in Herce (2019a).

Table 3.17) depending on which (or whether either) 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 morphome-like properties in that they afford formal predictions and may, because of this, provide a niche or template for other (incoming) forms.

Table 3.18 Some morphological patterns arising from morphome interactions

	Lags Romansh 'let, cause' (Maiden 2018b: 108)		Bolognese 'go' (Maiden 2012)		Felechosa Asturian 'bring' (Maiden 2012)	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	laʃel	laʃi	va:g	va:ga	trao	traa
2SG	lais	laʃies	vɛ	va:g	traes	traas
3SG	lai	laʃi	va	va:ga	trae	traa
1PL	ʃɛin	ʃejen	andain	andannja	traemos	triʃamos
2PL	ʃɛis	ʃɛjes	andɛ	andɛdi	traes	triʃaəs
3PL	lain	laʃien	van	va:gen	traen	traan

In the Lags Romansh paradigm in Table 3.18, for example, there is a stem alternant *lai-* which lacks the stem-final consonant /ʃ/ and has /i/ instead. This form is believed to have originated in the SG imperative and to have spread to these other cells analogically (see Maiden 2018b: 108). The SG imperative, 2/3SG indicative, and 3PL indicative constitute the set of cells that belong to the N-morphome but not to the L-morphome. It is the smallest morphomic niche to which forms originating in the SG imperative could possibly spread.

The Bolognese paradigm in Table 3.18 shows how the form /g/ characteristic of the L-morphome does not appear in the 1PL and 2PL subjunctive where it would be expected. It is relatively common for L-morphome roots to be expelled from these cells (see Maiden 2012), 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) introduced in 1PL and 2PL subjunctive. These are the cells that participate in the L- but not in the N-morphome.

The analogical processes described above illustrate how the different swathes of the paradigm that originate from cross-cutting formal elements (see Table 3.24) 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 morphomes apply can also be singled out, however, as the domain of allomorphy.

In Table 3.19, formal elements have spread to all the cells that participate in the N- and/or the L-morphome. Consider the Old French verb 'have'. 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

Table 3.19 Another morphomic pattern arising from morphome interactions

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

Ansotano Aragonese in Table 3.17. 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 thus acted 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'. Raising (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. Raising would thus have occurred, initially, in just those cells. In Spanish, however, as 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 are 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.

3.1.5 Grammaticalization

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

Although usually this is not explicitly discussed, not all linguists subscribe to the idea that morphomes must be typologically unique by definition. Stump (2015: 134), for example, discusses the case of a morphological affinity in Noon (Atlantic Congo), which he presents as a textbook example of a morphomic structure. This morphome 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 and Kuteva 2002: 236; Siewierska 2010) that 3PL is often a source for passive morphology, frequently via other intermediate functions like impersonal. The same morphological quirk is found in unrelated languages like Kven (Uralic) (Söderholm 2017).

Like 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. As similar examples of this particular affinity of 3PL=passive, one could offer other cross-linguistically recurrent changes like instrumental>ergative[8] (Palancar 2001) or 1SG.OBJ>antipassive (Bickel and Gaenszle 2015). The mere fact that these homonymies (e.g. ergative/instrumental) are most usually described as different cases/functions with homonymous exponents, rather than as a single case/macrofunction with various uses, suggests that this intuition is widely shared.

Morphological vestiges of grammaticalization processes can be relatively common cross-linguistically, like the ones mentioned above, or more idiosyncratic. In Lango (Nilotic), for example, there is a special morphological affinity between the infinitive and the progressive aspect forms. As Table 3.20 illustrates, the verbal system in Lango is based on three aspects (perfective, habitual, and progressive). In a way similar to the affinity between the infinitive and the past tenses in Balochi and other Iranian languages (see Section 3.1.3.2), the affinity of the infinitive and

Table 3.20 Partial paradigm of Lango 'stop sth', infinitive: *gìkkò* (Noonan 2011: 92)

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

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

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 Lango is due to the fact that the progressive originated, 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 conventionalization of that construction in Lango to express the progressive meaning and the later univerbation of that periphrastic construction into a synthetic tense are straightforward grammaticalization-related developments which, however, have left their mark in the synchronic paradigmatic organization of the language in the form of a morphological partial identity of infinitive and progressive. Notice, however, that similar processes have resulted in very different morphological affinities in other languages (e.g. infinitive and future/conditional in Romance), which suggests that these grammaticalization-derived paradigmatic structures are not less arbitrary/morphomic than those arising via formally driven analogy (e.g. between infinitive and past-tense in Persian, see Section 3.1.3.2), or via sound change (e.g. between the infinitive and 3PL present in Scandinavian, see Section 3.1.1).

Despite cases like Lango, because of the way syntax most usually behaves, the morphology that emerges from the accretion of formerly separate words tends to be relatively well-behaved in that it usually characterizes a natural class (e.g. a whole tense, or a set of related tenses). It is, however, definitely not the case that syntax is always only sensitive to natural classes (see Section 2.12.2), or that univerbation processes can only ever occur in natural classes. Consider the case of Athpariya in Table 3.21.

Table 3.21 Athpariya 'go', intransitive positive non-past (Ebert 1997: 163)

	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

As Schackow (2016: 230–31) explains, Athpariya *-yuk* goes back ultimately to a lexical verb *yuŋ*, which meant 'be' or 'stay'. This verb, and others in other Kiranti languages, must therefore have grammaticalized into the so-called tense markers we find synchronically, and, in the case of Athpariya, just in 2/3SG and 3PL. The fact that univerbation happened in these specific cells only must be related to the fact that those were the cells which lacked suffixes originally (see Bantawa (Doornenbal 2009: 391), or Puma (Sharma 2014: 424) for related languages still with zeroes in those cells).

3.1.6 Borrowing

The borrowing of morphological forms or patterns between languages is a common force in language change. Because of their very particular characteristics, however, morphomes (at least of the kind analysed here) seem to find themselves almost always at the worst end of the borrowability scale. In the analysis of which factors favour or hamper borrowability, the literature on language contact (e.g. Kossmann 2015; Matras 2015; Seifart 2015) 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 morphomes as defined here (i.e. they are grammatical, inflectional, complex-meaning structures), they are expected to 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 pointed out by Kossmann (2015: 260), '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 impossible in the case of morphomes which are, by definition, ill-suited for the transmission of meaning.

Probably for the aforementioned reasons, no incontrovertible examples of morphome borrowing have been found so far. There are, however, cases that come very close indeed, with respect both to matter and to 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.

As Table 3.22 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, nor in the paradigms of the verb 'have' in other Romance languages. In most cases, it is assumed that the presence of /g/ in this verb is due to analogical influence from other verbs (e.g. Sp. *decir/digo*) which would indeed have had a g-stem alternant in the L-morphome as a result of regular sound change. What is remarkable about the presence of this form in Sardinia is that, unlike in the rest of Romance, velars were not subject to the palatalizations that generated *decir/digo*-type alternations elsewhere. The

Table 3.22 Present-tense paradigm of Campidanese Sardinian *tɛnni* 'have' (Lepori 2001)

	IND	SBJV
1SG	ˈtɛngu/ˈtɛnju	ˈtɛnga/ˈtɛnja
2SG	ˈtɛnis	ˈtɛngas/ˈtɛnjas
3SG	ˈtɛnit	ˈtɛngat/ˈtɛnjat
1PL	tɛˈnɛus	tɛnˈgaus/tɛnˈjaus
2PL	tɛˈnɛis	tɛnˈgais/tɛnˈjais
3PL	ˈtɛnint	ˈtɛngant/ˈtɛnjant

formative /g/ as an exponent of the L-morphome must therefore be foreign to Sardinian, and must have been borrowed from another Romance language like Italian or Catalan.

This is undoubtedly a very interesting morphological development. However, it 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 Sardinian 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, also to the verb 'have', in other Sardinian varieties (e.g. *ˈtɛndʑɔ/ˈtɛnɛs* in Nuorese, see Pittau 1972). Only the formative /g/, thus, and not the L-morphome as such, can be said to have been borrowed into Campidanese Sardinian.[9]

Concerning the pattern-only borrowing of morphological categories, some striking cases exist of whole inflectional systems being restructured to match the categorial distinctions of another language. One of the most dramatic cases is found in Tariana (Table 3.23). As explained by Aikhenvald (2002: 102–4), 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 in the process, however, 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 (i.e. morphomic) to some extent.

[9] A similar situation holds with respect to some Campidanese Sardinian varieties (Loporcaro 2013: 152), which have acquired the *vādō* vs *andō* N-type suppletive stem alternation found in other Romance varieties. Although Sardinian did not develop N-morphomic stem-vowel alternations via sound change, stress also became morphological in that language, with rhizotonic forms following the N-pattern. This paradigmatic split in stress may have allowed/facilitated the borrowing of 'foreign' suppletive alternations with the same distribution.

Table 3.23 Morphological realization of some semantic functions in three Amazonian languages (adapted from Aikhenvald 2002: 102–4)

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

Finally, 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 pattern, not a paradigmatic one like those that this monograph deals with primarily.

3.1.7 Mixed origins

The previous sections have presented evidence of how morphomes can arise in a language in quite a few different ways: due to (i) sound changes (3.1.1), (ii) semantic drift (3.1.2), (iii) morphosyntactic or form-driven analogy (3.1.3), (iv) pattern interactions (3.1.4), (v) grammaticalization (3.1.5), and maybe even through (vi) language contact (3.1.6). I have so far attempted to present clear examples of morphomes that have emerged due to only one of these forces. The history of many morphomes, however, is a combination of several of the above-mentioned diachronic processes either simultaneously or at different stages. Consider, for example, the cases in Table 3.24.

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, changes must have singled out the 1SG (and 3PL?) indicative and the subjunctive forms of the present as those with a different stem from that found elsewhere. In Servigliano, however, morphosyntactically driven analogical processes involving the loss of mood distinctions in the first and second-person, and of number

Table 3.24 Present-tense paradigms of three Servigliano Italian verbs (Camilli 1929)

	pote 'can'		*di* 'say'		*ae* 'have'	
	IND	SBJV	IND	SBJV	IND	SBJV
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

distinctions in 3, have modified the original paradigmatic distribution of the inherited alternations. The morphome'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 be offered of morphomes having a complex diachronic origin. As for those in the present database (Chapter 4), the morphomes of Aragonese (Section 4.2.3.1) and Palantla Chinantec (4.2.5.6), for example, must have involved both (i) sound change and (iv) pattern interactions. As for morphomic structures discussed elsewhere, the Northern Talysh verbal morphomes 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 cases that I have assembled in Chapter 4, it seems that complex diachronic origins may well be the rule rather than the exception in morphome emergence.

3.2 Loss and change of morphomic structures

Earlier sections have dealt with the various ways in which morphomes may arise in a language. Even though these structures are usually taken to be quite stable in the literature,[10] it is obvious that, just like other grammatical traits, morphomes can disappear from a language. This section will present the different ways in which this may happen.

[10] The validity of these claims is not clear to me at this point. Even if we found that the average life expectancy of a morphome is 2,000 years, for example, it would still not be obvious 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 morphomes 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 replacement of different lexical items, or the life expectancy of other grammatical traits like ergative alignment, pro-drop, and clusivity (see e.g. Wichmann and Holman 2009).

3.2.1 Loss of productivity and gradual erosion

As soon as a class or category ceases to be productive and incorporate new members regularly, it can in some sense be said to be already on its way out from a language. In the absence of new recruits, and provided sufficient time, any class would eventually vanish due to the relentless trickle of 'desertions' that it would undoubtedly suffer. Note, however, that categories can remain largely unproductive for extremely long periods of time before they disappear completely.[11] During this time they may remain part of the grammar, 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 morphomes (Romance PYTA, and L- and N-morphomes) can be described as being at this stage to some extent. They are thus largely unproductive but nevertheless 'living'[12] entities in many Romance languages (e.g. Spanish). Some of these morphomes (L and N) have probably never been truly productive categories (in the sense that new lexemes did not display them by default at any stage). They may always have been losing members, therefore, ever since they first appeared in the language. Some other morphomes like PYTA, by contrast, were completely productive morphological categories at some point, as morphological distinctions were regularly made in Latin (e.g. adding a suffix /w/) to mark the perfective tenses. In languages like Friulian and Romansh (see Herce 2021b), PYTA has disappeared almost completely, largely due to this erosive effect of unproductivity. This should therefore be understood always as a prerequisite, and often also as a direct cause of morphome loss.

Because of the long periods of time over which unproductive categories can exist in a language, it is difficult to find and present examples of morphomes that disappeared exclusively due to the constant eroding effect that lack of productivity brings about. In lieu of an example where a formerly productive morphome 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' an unproductive morphomic pattern. These will I hope

[11] Consider e.g. 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.

[12] Nevins et al. (2015) attempt to show experimentally that the L-morphome is 'dead' in Romance and that it died largely because of this loss of productivity. There are, however, 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 morphomes), it can hardly be said to be 'dead', as it continues to provide a template for the distribution of some alternations. 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 *lluevo* and *llovemos* respectively. These cannot be memorized forms, since they never appear in natural speech. The forms are created online, analogically, on the basis of other verbs with the same formal alternations.

illuminate the reasons why particular lexical items may change their inflection by letting go of a morphomic alternation, generalizing a single form throughout their paradigm.

The Spanish N-morphome is a relatively robust morphomic pattern, appearing 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 understandably fail to learn that it was supposed to have two forms in the first place. When this happens, because of the smaller frequency of use (a ratio of around 1:3) of the N-morphome 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 in Spanish is *degollar* 'cut someone's throat'. Thus, the N-morphome verb *degollar/degüello* is being increasingly replaced by a non-alternating *degollar/degollo*. A similar, more widespread levelling (both diphthongization and lack thereof are prescriptively acceptable) is that of *asolar/asuelo* 'destroy' changing to *asolar/asolo*. Less frequently, it can 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' respectively. In the case of the first two, the related verbs *asolar* and *degollar* have become divorced from their source nouns.[13] In the case of the later, the connection to *mueble* remains evident to the Spanish language user, a fact which can steer the levelling into the preservation of this synchronic connection.

Apart from 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 2.4, the formerly alternating verb *levar/lievo* 'carry' split into two non-alternating verbs *llevar/llevo* and *levar/levo*, as a result of the sound change /lje/>/ʎe/, which transformed a typical N-morphomic alternation /e/ vs /je/ into

[13] In the case of *asolar/suelo*, the reason for 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 the synchronic connection must be due to the formal discrepancy /k/ vs /g/ produced by intervocalic voicing, which occurred only in the verb. These form- and meaning-driven break-ups of the synchronic derivational relation are very reminiscent of the lexeme splits presented in Section 2.1.2.

an exceptional one /l/ vs /ʎ/. This must have made it more difficult (although definitely not impossible, as witness its Romanian suppletive cognate) to identify e.g. *levar* and *ʎevo* as forms of the same lexeme, which motivated the split and the analogical filling-out of the missing forms.

Developments like these, and the diachronic morphological convergence that morphomes often display, speak against taking a morphome's applicability to novel forms (see n. 12 as the only way to assess whether a given morphological pattern is 'living' or 'dead'). Dichotomous taxonomies like this one in Nevins et al. (2015) are probably too coarse-grained, in any case, to capture a pattern's vitality in the grammar in any meaningful way.

3.2.2 Loss of morphosyntactic categories

Another, more abrupt way in which morphomes 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) can be lost seemingly in one fell swoop. When this happens, this will inevitably erase any (part of a) morphome that occurred inside the lost swathe of the paradigm.

In Pantesco Italian (Table 3.25), as well as in other southern Italian varieties, the present subjunctive has disappeared.[14] Without this tense, the earlier L-morphome stems (with classically L-morphomic exponences like /ts/, /k/, and /ɲ/) have become confined to a single cell in the paradigm, which can never be morphomic as defined here. Something similar can happen in the case of TAM morphomes like PYTA in Table 3.26.

Table 3.25 Present tense of some verbs in two Romance varieties

	Spanish *decir*		Pantesco (Loporcaro et al. 2018: 297–8)		
	IND	SBJV	pɔːtiri 'can'	'diːrɪ 'say'	'vɛːnɪrɪ 'come'
1SG	digo	diga	ˈpɔtːsu	ˈdiːkʊ	ˈvɛɲːʊ
2SG	dices	digas	ˈpɔ	ˈdiːʃɪ	ˈvɛːnɪ
3SG	dice	diga	ˈpɔ	ˈdiːʃɪ	ˈvɛːnɪ
1PL	decimos	digamos	puˈtɛːmʊ	diˈʃɛːmʊ	viˈnɛːmʊ
2PL	decís	digáis	puˈtiːtɪ	diˈʃiːtɪ	viˈniːtɪ
3PL	dicen	digan	ˈpɔnːʊ	ˈdiːʃɪnʊ	ˈvɛnːu

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 in Table 3.26). As one moves east

[14] See Servigliano Romance in Table 3.24 for an intermediate variety which has lost this tense (or has merged it with the indicative) only in the non-3 forms.

Table 3.26 3SG forms of 'make/do' of former perfective tenses

	Latin	Portuguese	Galician	Somiedo Asturian	French	Alpago Italian	Nuorese Sardinian
PAST.IND	fēcerat	fizera	fixera	fiˈʃera	–	–	–
PAST.SBJV	fēcisset	fizesse	fixese	–	fît	ˈfese	–
PRS.IND	fēcit	fez	fixo	ˈfiʃu	fit	–	–
PRS.SBJV	fēcerit	fizer	–	–	–	–	–
FUT.IND	fēcerit						

along the Romance dialect continuum, however, fewer of these tenses have been preserved: three are preserved in Galician, two in Somiedo Asturian (see Cano González 1981) and in French (although different ones), only one in Alpago Italian (see Zörner 1997), and none in Nuorese Sardinian (see Pittau 1972). In the last two varieties, the PYTA TAM morphome is (and can logically be) no more.

3.2.3 Sound change

Most of the processes identified in Section 4.1 as potential creators of morphomes can also participate in their disappearance or in their change into a different pattern. A force that may be involved in the demise or change of a morphomic 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 morphomic patterns.

The original distribution of the N-morphome (illustrated in Table 3.27 by Italian) has been disrupted[15] in various Italian varieties as a result of sound change. In Macerata, for example, an anticipatory assimilation of the stem vowel

Table 3.27 Present indicative of two cognate verbs in two Italian Romance varieties

	Macerata (Maiden et al. 2010)				Standard Italian			
	'sleep'		'feel'		'sleep'		'feel'	
	SG	PL	SG	PL	SG	PL	SG	PL
1	ˈdɔrmo	durˈmimo	ˈsɛndo	sinˈdimo	ˈdɔrmo	dorˈmjamo	ˈsɛnto	senˈtjamo
2	ˈdurmi	dorˈmete	ˈsindi	senˈdete	ˈdɔrmi	dorˈmite	ˈsɛnti	senˈtite
3	ˈdɔrme	ˈdɔrme	ˈsɛnde	ˈsɛnde	ˈdɔrme	ˈdɔrmono	ˈsɛnte	ˈsɛntono

[15] In line with the *modus operandi* in the rest of this book, morphomes here are defined over their paradigmatic distribution. Thus, the morphomes of Italian and Macerata above are considered different (albeit cognate) morphomes. The change in Macerata involves the disappearance of the SG+3PL morphome and the emergence of another. This is why it has been presented in this section (dealing with morphome disappearance). This way of thinking or talking about it is entirely a narrative convenience; one could just as easy have expressed it as the Italian-type morphome becoming a Macerata-type one.

to a following /i/ has modified the paradigmatic domain of occurrence of the classically N-morphomic open-mid vowels.

3.2.4 Analogy

As in the case of morpheme 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 already presented (Wambisa in Table 3.13 and Servigliano in Table 3.24) provide 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.

3.2.4.1 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 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 (e.g. /tuˈβemos/ vs /teˈnesemos/, /suˈpjemos/ vs /saˈpesemos/, /kiˈsjemos/ vs /keˈɾesemos/, /estuˈβjemos/ vs /esˈtasemos/ in Panticosa, see Nagore Lain 1986).

Another case of a Romance morphome retreating into a natural class can be found in Gallo-Romance, where the L-morphome has retreated from the 1SG indicative, thus becoming confined to the present subjunctive (see Table 3.28).

Table 3.28 Present-tense conjugation of three Seyne Occitan verbs (Quint 1998)

	'know'		'be worth'		'be able to'	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	ˈsabu	ˈsatʃe	ˈvalu	ˈvage	ˈpwɔ	ˈpuske
2SG	ˈsabes	ˈsatʃes	ˈvales	ˈvages	ˈpwɔs	ˈpuskes
3SG	ˈsabe	ˈsatʃe	ˈvɔw	ˈvage	ˈpwɔ	ˈpuske
1PL	saˈben	saˈtʃen	vaˈlen	vaˈgen	ˈpwen	pusˈken
2PL	saˈbe	saˈtʃe	vaˈle	vaˈge	ˈpwe	pusˈke
3PL	ˈsabun	ˈsatʃen	ˈvalũ	ˈvagen	ˈpwõ	ˈpusken

Cases like these are sometimes explained not so much as a direct fall back to morphosemantic distributional criteria but in alternative ways. For example, for Aragonese, Maiden (2018b) suggests a possible retreat of PYTA to rhizotonic cells initially (all of which must have occurred in the preterite), followed by a later analogical extension to the rest of the preterite cells.

In the case of the Occitan development in Table 3.28 Maiden 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 before the subjunctive present suffix *-a*. When they became word-final, some stem-final consonants devoiced in the 1SG present indicative, thus breaking surface stem identity with the present subjunctive.

It should be noted, however, that the same alignment of the L-morphome with the present subjunctive is sometimes found, less robustly, in other varieties too (e.g. Sardinian, Loporcaro 2012: 18–19), where this story cannot apply. My contention is that the changes in Aragonese, Occitan, and Sardinian, like probably most analogical changes, must be conditioned by a multiplicity of factors. There is little reason not to consider the alignment to morphosemantic values one of their motivations, maybe even the most important one. Beyond the morphomic literature on Romance, in fact, the alignment of formatives to natural classes has usually been considered relatively common (see e.g. Wurzel 1980).

Germanic offers some well-known examples of morphological forms changing an inherited unnatural distribution into a natural one in order to perform morphosemantic roles. Sometimes, as in Occitan above, there are confounding factors in the form of formatives with the target natural distribution. In this way, some changes into a natural class might also be partially explained as formally motivated analogies. Cases like those in Table 3.29, 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; 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.

Table 3.29 Declension of OHG 'lamb' (Wurzel 1980: 445–8) and OE 'foot' (Fertig 2016: 436)

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

Analogical changes like the ones in Table 3.29 demonstrate that alignment to morphosemantic values can be a force involved in the demise of morphosyntactically unnatural patterns. The reason why this is not observed frequently in Romance may be due to properties particular to them, such as their redundancy in the paradigm (i.e. they hardly ever perform whole-word discrimination roles) and their allomorphic diversity.

3.2.4.2 Change into a different unnatural class

As was shown in the case of morphome emergence, not all analogical processes result in more isomorphic form–function relations. Some of the cases presented in Section 4.1.3 illustrate 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, I will focus on a different case study: the analogical disintegration of Romance PYTA (i.e. its change into a different paradigmatic pattern).

As Table 3.30 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. Many varieties have thus clearly trimmed the inherited distribution of perfective root allomorphy to make rhizotony and the PYTA root (both purely morphological properties) paradigmatically coextensive (see Esher 2015; Maiden 2018a). These developments illustrate another possible motivation for the loss/change of morphomes in a language. The 'fall' of (the etymological distribution 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 trait, 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) also constitutes a simplifying development with respect to the predictability of one trait from the other.

Table 3.30 Remnants of PYTA root in various Romance varieties (Herce 2021b)

	Sicilian 'have' (Maiden et al. 2010)		**Italian 'cook'**		**Oscos Galician 'put' (Maiden 2018b: 76)**	
	PRET	IPF.SBJV	PRET	IPF.SBJV	PRET	IPF.SBJV
1SG	ˈappɪ	aˈvissɪ	ˈcossi	cuoˈcessi	ˈpuʃeŋ	poˈɲɛse
2SG	aˈviʃtɪ	aˈvissɪtʊ	cuoˈcesti	cuoˈcessi	poˈɲit͡ʃe	poˈɲɛses
3SG	ˈappɪ	aˈvissɪ	ˈcosse	cuoˈcesse	ˈpuʃo	poˈɲɛse
1PL	ˈappɪmʊ	aˈvissɪmʊ	cuoˈcemmo	cuoˈcessimo	poˈɲɛmos	poˈɲɛsemos
2PL	aˈviʃtɪvʊ	aˈvissɪvʊ	cuoˈceste	cuoˈceste	poˈɲɛstes	poˈɲɛseðes
3PL	ˈappɪrʊ	aˈvissɪrʊ	ˈcossero	cuoˈcessero	poˈɲɛroŋ	poˈɲɛseŋ

3.2.5 Mixed causes

Although logically different causes have been kept distinct in Sections 4.2.1–4.2.4, the story of the demise of most morphomes usually involves a combination of factors, rather than one motivation exclusively. Consider the pattern in Table 3.31.

Table 3.31 Present tense of two Gartempe Occitan verbs (Maiden 2018b: 288)

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

As Maiden (2018b) 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 (e.g. 'sing') are a result of regular sound changes. In the case of verbs with stem alternation (e.g. 'save'), whole-word syncretisms should not have resulted. However, the consolidation of the sound-change-triggered syncretisms at a deeper grammatical 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 (SG+3PL) and became confined to the 1SG=3SG cell. Sound change and analogy combined. This case is an example of morphome demise/change as a result of several different forces. Although the different motivations have been discussed separately in this section for convenience, in reality, most of the time it is a combination of factors that is responsible for a morphome's disappearance from a language.

3.3 Discussion

The emergence and disappearance of morphomic patterns in a language show important parallels. Largely the same forces have been identified as potential motivators of both morphome creation, morphome change, and morphome 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 then 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. Most 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 cases of form-derived morphological niches, i.e. of cases where form-derived templates take the upper hand over morphosyntactic or semantic ones.

Romance is well known for this in the morphome literature. In many varieties, formerly independent lexical items (e.g. Latin *ambulāre* and *vādere*) 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 Old High German paradigms in Table 3.32, and their descendants in Luxembourgish in Table 3.33.

Table 3.32 Present tense of four Old High German verbs (Braune and Reiffenstein 2004)

	faran 'drive'		*wësan* 'be'		*kweman* 'come'		*mahhōn* '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

In the history of Germanic, a vowel was sometimes fronted or raised before an /i/ in the next syllable (see Table 3.2). In the verbal paradigm, this happened in the 2SG and 3SG of some verbs (see e.g. *faran*), which gave rise to an alternation

Table 3.33 Present tense of the same four verbs in Luxembourgish (Schanen 2004)

	fueren 'drive'		*sinn* 'be'		*kommen* 'come'		*maachen* '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	méchs	maacht
3	fiert	fueren	ass	sinn	kënnt	kommen	mécht	maachen

pattern opposing 2SG/3SG to the other forms of the present. These sound change-created stem alternations, however, were subsequently used as a template for the distribution of other differences. They have acted, diachronically and in processes of analogical change, as 'islands' that favour internal homogeneity, with formal differences pushed 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 originally. 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 2PL to assimilate in the same way.[16] 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 any alternation whatsoever 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 (Schackow 2016). In this and in other East Kiranti languages (see Herce 2021a), verbs have two stems, one of which (usually longer) occurs before suffixes beginning with a vowel, while the other occurs before consonants. Consider the non-past-tense paradigms of intransitive (Table 3.34) and transitive (Table 3.35) verbs in Chintang, a closely related language, for an approximated illustration of the system ancestral to these East Kiranti languages.

Table 3.34 Paradigm of Chintang 'come level' non-past, intransitive (Paudyal 2013: 86)

	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

[16] 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.

Table 3.35 Chintang 'give' non-past, transitive, 3SG patient (Paudyal 2013: 294)

	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

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 somewhat 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 (Tables 3.36 and 3.37).

Table 3.36 Paradigm of Yakkha 'come' non-past, intransitive (Schackow 2016: 243)

	SG	DU	PL
1EXCL	am-meŋna	am-meŋciŋha	ab-iwaŋha
1INCL		am-meciya	ab-iwha
2	am-mekana	am-mecigha	ab-iwagha
3	am-meʔna	am-meʔciya	ŋ-am-mehaci

Table 3.37 Yakkha 'understand' non-past, transitive, 3SG patient (Schackow 2016: 244)

	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

As these tables 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: 230–31) explains, these suffixes go back ultimately to lexical verbs,[17] which grammaticalized into the tense markers we find synchronically. As Herce (2021a) explains, an utterly morphosyntactically unnatural stem alternation pattern has provided the niche for the incoming present-tense suffixes, which 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).

3.4 Conclusion

The present chapter has explored ways in which morphomes may arise, change, and disappear from a language, and the forces and reasons behind those processes. It has been found that sound changes (in various ways), semantic drift, analogical change (both morphosyntactically and formally motivated), pattern interactions, grammaticalization, and (maybe) language contact are all processes that can be involved in morphome emergence. Some of these (e.g. morphosyntax-driven analogy, and grammaticalization) might not have been expected, given the origins of the most thoroughly studied morphomes (i.e. the Romance N, L, and PYTA). The only possible conclusion regarding morphome diachrony is that basically any process that can produce a change in the paradigmatic distribution of some form(s) can be involved in processes of morphome emergence and loss.

The forces involved in morphome emergence, change, or loss seem at first sight not to differ from those at play in morpheme diachrony. However, although more quantitative research into this matter is needed, the particularities of morphomes seem to make certain diachronic origins more common (e.g. sound change) and others uncommon (e.g. borrowing). Of those morphomes in the database (see Chapter 4) whose history could be recovered (N=96), as many as 73 (76%) involved sound change,[18] another 19 (20%) involve analogical change (15 morphosyntactically driven analogy and 4 form-driven analogy), 8 (8%) semantic drift, 4 (4%) pattern interactions, and one case was found of grammaticalization. Often (in 8 cases, although this is likely to be an underestimation), more than one of these was involved in the history of the same morphome.

Although the criteria for morphomehood used in the database's compilation, as well as the state of linguistic documentation and knowledge of the languages' history, must influence the numbers of Figure 3.1, the prevalence of sound-change-generated morphomes seems clear, and should thus be regarded as this

[17] 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'.

[18] All of these except one were of the 'morphological divergence' type defined in Section 3.1.1.1. Most (65%) also involve sound change in the morphome-complement cells, rather than in the morphome cells. This may result from sound changes being more likely to be resisted ab initio when they affect a small number of cells only.

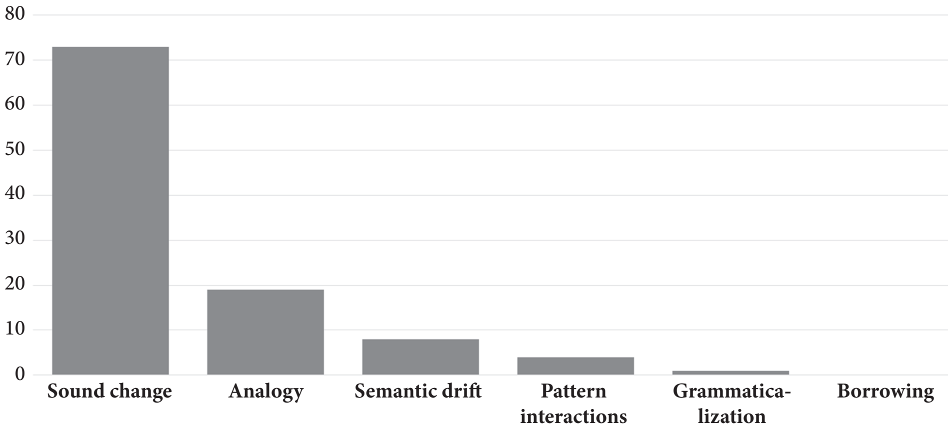

Figure 3.1 Diachronic origin of the morphomes in the morphome database of Chapter 4

section's most robust finding. This chapter has also contributed to our knowledge of morphome diachrony by calling attention to and typologizing the various ways in which sound changes create morphomes. On the basis of their domain of application, sound changes can happen either in the morphome or in its complement cells. On the basis of their result, sound changes can create morphomes by disrupting previous formal invariance (i.e. A~A > A~B), or by erasing a formal difference (i.e. A~B > A~A) between word forms that do not possess any particular morphosyntactic affinity. In addition to these types, as discussed in Section 3.1.1.1, it has been found to be quite common (a total of 18 (19%) such cases have been found here) for morphomes to emerge from zero vs affixed morphological configurations, and from longer-affix vs shorter-affix ones. The existence of trends regarding the paradigmatic distribution of zeros and short forms (consider Zipf's law) might lead to some cross-linguistic tendencies in these morphomes.

In the discussion in Section 3.3, and before in 3.1.3.2 and 3.1.4, some clear cases were presented of how, for reasons of paradigmatic predictability, morphological forms and alternations can provide a template for the organization of (novel) allomorphy. The reason why morphological categories, like functional categories, can behave like this is that language is an inherently productive system. On the basis of a limited input, language users need to infer/construct a watertight system. This means that 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 morphological and predictive regularities they observe in their input to infer and produce previously unencountered forms. This is the mechanism that allows morphomes (and morphemes), whether they originate from the morphologization of sound changes, from analogy, from semantic drift, or from grammaticalization, to sometimes become productive/active morphological categories that may, on a par with morphosyntactic values, participate in exponence rules and steer morphological change.

4
Morphomes in synchrony

The two most significant limitations of research on morphomes to date have been (i) the scarcity of data from sources other than the usual Romance suspects, and (ii) the difficulty of systematic comparison between different morphomes due to the absence of measurement uniformity (see Round and Corbett 2020). This chapter answers these challenges and presents the highlight of this book: a cross-linguistic morphome database.

The nature of the enterprise is such that, although dichotomous arbitrary choices regarding morphomicity were highlighted and avoided in previous chapters, they now become necessary. To ensure objectivity and cross-linguistic homogeneity regarding when concrete structures will be regarded as morphomes and included in this database, clear-cut criteria need to be in place to use the same yardstick with all examples. Section 4.1 presents these criteria, which will be grounded in the discussions of Chapter 2. Section 4.2 will present all morphomes in the database, in great qualitative detail. Section 4.3 will present the variables and measures of morphomic diversity, and the quantitative results regarding what synchronic morphomes tend to be like. Section 4.4 will present some preliminary statistics to obtain a deeper knowledge of these structures and to identify variable correlations and dependencies.

4.1 Criteria for inclusion in the morphome database

The common practice in morphomic literature has been to identify and discuss morphomes on a case-by-case basis, taking into account a wide range of unstructured and relatively subjective criteria. The most important of these have been (i) the failure to identify a semantic or morphosyntactic property exclusive to those cells, and (ii) some diachronic evidence that a particular set of cells has behaved in a unified way in analogical changes. Other criteria are seldom discussed, but I suspect that theoretical morphological notions including blocking or defaults, the generality of a pattern across the lexicon, and the degree of allomorphy involved must be, at least sometimes, lurking in the back of the mind of the morphologist when they try to assess whether or not a given pattern is a morphome.

It is evident that in the context of a broad typological investigation, such an approach is unsuitable. To quantify and classify morphomes cross-linguistically, clear and blindly applicable criteria are needed in order to overcome any personal

The Typological Diversity of Morphomes. Borja Herce, Oxford University Press. © Borja Herce (2023).
DOI: 10.1093/oso/9780192864598.003.0004

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 morphomes are not natural kinds (see Section 2.2), 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–morphome scale. I will therefore set purposefully high requirements for inclusion of a morphological structure in the present cross-linguistic morphome database.

4.1.1 Unmistakably unnatural paradigmatic distribution

In earlier sections it has been established that, of the various loosely connected meanings of the term 'morphome', this book is only going to be concerned with what Round (2015) called 'metamorphomes', i.e. with cells that, within the inflectional paradigm of a given lexeme, share exponents or morphological traits despite not constituting a natural class.

When assessing whether or not a set of cells constitutes a natural class, the assumed feature structure plays a crucial role. For someone who is maximally reticent to grant natural class status, the syncretism of any two or more values (e.g. dual and plural; dative, genitive, and ablative) may count as morphologically stipulated. Many (maybe most) morphologists will be more permissive in this respect, and will 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, as I advanced before, I want the threshold for 'naturalness' to be high here, I will go a step further and **allow any two or more values of a feature to form a natural class**.

As Table 4.1 illustrates, the morphological syncretisms within the non-singular are in Teanu (and in the other two languages of the Vanikoro island) at odds with any plausible semantic or morphosyntactic feature value or bundle of values. Syncretisms like 1EXCL/1INCL vs 2/3 would be straightforward. Syncretisms like 1INCL/2 vs 1EXCL/3 may also be derivable as the expression of +2 (i.e. addressee) vs a default. The syncretism in Table 4.1 is, therefore, the only two-way syncretism that appears to make no sense whatsoever morphosyntactically. Those values are not generally considered to be particularly close semantically and are not prone to syncretization cross-linguistically (Cysouw 2003: 156–7). However, because of the criterion

Table 4.1 Subject agreement prefixes in Teanu (Oceanic) (François 2009)

	Realis			Irrealis		
	SG	DU	PL	SG	DU	PL
1EXCL	ni-	ba(i)-	pi-	ne-	ba(i)-	pe-
1INCL	–	la(i)-	li-	–	la(i)-	le-
2	a-	ba(i)-	pi-	u-	ba(i)-	pe-
3	i-	la(i)-	li-	i-	la(i)-	le-

adopted above, this and other morphological affinities of cells which differ in just one value will not count as morphomic for the purposes of inclusion into the present database.

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 Table 4.2. In Icelandic, 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, and has whole-word syncretism of infinitive and 3PL present indicative. 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 finite 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 4.3.8) of a metamorphome consisting e.g. of 3PL.PRS and infinitive. This book will thus focus on those parts of the paradigm where orthogonality does hold, excluding those paradigm cells (e.g. nonfinite forms,

Table 4.2 Paradigm of *eiga* 'own' in Icelandic (Jörg 1989)

	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							

imperatives, 1PL inclusives) 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 a subset of the paradigm, then it will necessarily remain unmotivated in any larger superset.

A similar orthogonality challenge is presented by cases of so-called TAM morphomes (Smith 2013). Whereas some features, like person and number, or case and number, are generally well-behaved regarding their (logical) orthogonality, others like TAM are more troublesome. Thus, it is often difficult to find a perfect orthogonality of tense and aspect, aspect and mood, or tense and mood. The difficulty or impossibility of organizing these into orthogonal features and values has the consequence that establishing what counts as a natural class is difficult in this type of morphomes (consider, for example, the discussion around the 'non-canonical' morphome Fuèc, comprising the future and conditional tenses in Occitan, see Esher 2013).

Another example of a morphological pattern that does not qualify for inclusion in this database is stem alternation in Daai Chin (Sino-Tibetan). In around 20% of the verbs, one stem (arbitrarily labelled 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. Its complement, stem B, is present in (i) indicative intransitive verbs, (ii) interrogative (unless in the presence of narrow focus), (iii) imperative, and (iv) non-final clause chains. Each of the stems seems therefore to be 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 uncertain feature and value structure, it is impossible to ascertain this, let alone quantify it as I intend to do. Because of this, this type of morphomes will be excluded from the present cross-linguistic study, even if it includes some of the most famous ones in the literature like PYTA (as present e.g. in Spanish or Portuguese) or the Latin third stem.

The last type of paradigms that will be excluded from this database are those that, even in the presence of perfect orthogonality, involve features that are very closely related by virtue of having similar or identical values.

Consider the case of Komnzo in Table 4.3. Agent number and patient number are different features. 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 best-known morphomes in the literature like the N-morphome (SG and/or 3) or the L-morphome (1SG.PRS and/or PRS.SBJV). At the same time, 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 on what we allow rules of exponence to do. Apart from agent number and patient number, other combinations where this may be found

Table 4.3 Form of a number marking formative in Komnzo verbs (Döhler 2018: 218)

		Patient no.		
		SG	DU	PL
Agent no.	SG	-wr	-n	-wr
	DU	-n	-n	-n
	PL	-wr	-n	-wr

are agent person and patient person, possessor number and possessee number, etc. These paradigms will also be excluded here pre-emptively from the ranks of morphomes.

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 metamorphomes in this synchrony-oriented part of this book will be found in person–number and case–number inflection. It is hoped that the greater morphomicity and measurability achieved with these standards will outweigh the loss of variability and datapoints in general.

4.1.2 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 (i.e. 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: 'mollusc']) or instead code them as different meanings of a single entry (e.g. /mʌsl/1: 'strength-related thing'). 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 and Sims 2010) is the ability of a form to resolve syntactic feature conflicts (see Section 2.1.1.1). 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 (ii) the typologist hardly ever has access to the wealth of descriptive data that would be required, in every language, to have the necessary information on these morphosyntactic-conflict-resolution-triggering constructions. Other tests and diagnostic criteria, as already discussed in Section 2.1.1, are also unsuitable.

Undoubtedly for similar reasons similar, some of the linguists that have encountered 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 practice,[1] and will be adopted here too for inclusion of a morphological identity into the synchronic morphome 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 morphomes which are based on these formal exponences. 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 morphome.** That is to say, there must be minimally one segment which appears in every single one of the cells constitutive of the metamorphome and in no other paradigm cell outside of it. Consider again the whole-word syncretism in Table 4.4.

[1] 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'.

Table 4.4 Verb agreement in Udmurt (Uralic) (Csúcs 1988: 142)

	1st conjugation, *mi̯ni̯ni̯* 'go'				2nd conjugation, *daśani̯* 'prepare'			
	Present		**Future**		**Present**		**Future**	
	SG	**PL**	**SG**	**PL**	**SG**	**PL**	**SG**	**PL**
1	mi̯ni-śko	mi̯ni-śko-m	mi̯n-o	mi̯n-o-m	daśa-śko	daśa-śko-m	daśa-lo	daśa-lo-m
2	mi̯ni-śko-d	mi̯ni-śko-di̯	mi̯n-o-d	mi̯n-o-di̯	daśa-śko-d	daśa-śko-di̯	daśa-lo-d	daśa-lo-di̯
3	mi̯n-e	mi̯n-o	mi̯n-o-z	mi̯n-o-zi̯	daśa	daśa-lo	daśa-lo-z	daśa-lo-zi̯

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 future tense cells. Thus, a description of the inflectional exponence of Udmurt need not make any reference to the class 3PL.PRS+1SG.FUT as such. It is the cells 3PL.PRS+FUT that fulfil the requirements for morphomehood. The absence of a formative (in other words, a zero-morpheme) will not count as a formal affinity for the purposes of inclusion into the database, where only overt formatives will be considered.

In this same vein of trying to avoid reference to dubious objects and/or theoretical analyses in the identification of morphomes in this chapter, subtractive affixes will not be allowed to feature in synchronic morphology.

Consider the French paradigm in Table 4.5. In the inflection of *lire*, the segment /z/ appears at the end of the stem in the plural forms of the present indicative and in the present subjunctive and imperfect cells. In other verbs, this additional consonant can be different: /n/ (e.g. in *prendre* 'take' or *venir* 'come'), /s/ (e.g. in *connaître* 'know' or in regular second-conjugation verbs like *finir* 'finish'), /ɲ/ (in *atteindre* 'attain'), /j/ (in e.g. *broyer* 'crush'), /v/ (in *écrire* 'write' or *boire* 'drink'), and the form shared by these cells can also be longer, such as /ɔlv/ in weakly suppletive alternations like the one in *résoudre* (*ʁezɔlv-/ʁezu-*) 'solve'.

Table 4.5 Paradigm of French *lire* 'read'

	PRS.IND		PRES.SBJV		IPF		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

An analytical option could involve assigning these forms to the stem and positing an invariable underlying form for it (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 recapitulate the diachronic origin of these patterns, which are often the result of sound changes from Latin to French which deleted the last consonant(s) of the original stem in some environments.[2] I will not pronounce myself as for the virtues of these and similar analyses, but will simply reiterate my commitment to stick to the presence or absence of overt surface forms to diagnose morphomicity in this book.

4.1.3 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: 332–3)
>
> 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)

I will side with Maiden here in allowing largely no role to theoretical notions like defaults in the definition of what will count as a morphome in this typological investigation. This will be so, firstly, because I want to remain as close as possible to the empirical data, but secondly, because of the lack of consensus on how to identify defaults in the first place.

Despite this resolve here, the extant literature on metamorphomes has indeed focused overwhelmingly 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 Romance literature has thoroughly discussed the N-morphome, the L-morphome, and PYTA, but not the complements of these cells.

[2] The Latin ancestor of *écrire* 'write', for example, showed a stem-final /b/ everywhere through the paradigm (Lat. *scrīb-ō scrīb-is scrīb-it scrīb-imus scrīb-itis scrīb-unt*). Its French offshoot has lost this consonant, which became /v/, from some of these forms (Fr. *ekʁi ekʁi ekʁi ekʁiv-ɔ̃ ekʁiv-e ekʁiv*).

Table 4.6 Non-PYTA root in the Italian verb *cuocere* 'cook' (Maiden and Robustelli 2014: 226)

	PRS.IND	PRS.SBJV	IPF	PAST	IPF.SBJV	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

Consider the Italian paradigm in Table 4.6. The complement cells of many of the best-known morphomes would often qualify as morphomes in their own right according to the criteria that are usually discussed for morphome identification. In the concrete case of Table 4.14, 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 /t͡ʃ/) that are not found elsewhere. The same pattern of stem identity is also 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'). The shaded cells, in addition, show diachronic properties entirely comparable to the more traditional morphomes. In the verb *cuocere* above, for example, the stem uniformity of /kwɔʧ/ within this domain has been achieved via analogical changes that have levelled other formal alternations (*wɔ/o*, *ʧ/k*) that must have been formerly present within this set of cells as the regular product of sound change.[3]

The reason why complement sets like this are not usually discussed as morphomes is not entirely clear to me but must be, I suspect, related to notions like basic/default. Languages need lexemes, and lexemes need at least one phonological form to exist in a language. The form *cuoc-*, because it occurs in most of the cells, could be conceived of as merely the form of the lexeme. Thus, only the 'odd man out' (i.e. the stem *coss-*) would need to be really 'explained' somehow. These concerns may be partially understandable. Because of this, and to allow for some continuity with extant morphomic literature, a concession will be made to those morphologists worried by defaults by not including in this cross-linguistic

[3] The existence of analogical processes aimed at preserving or extending a particular pattern could also be thought of as a possible definitional requirement in the identification of morphomes. 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.

database any converse-type morphomes when this set of cells constitutes a clear majority within the paradigm (operationalized as over 70% of the cells). Thus, only when two complementary patterns are relatively balanced as for the number of cells that they span, and only if they fulfil the earlier two requirements presented in this section, will both morphological patterns be included here. This requirement automatically implies that converse-type morphomes of a single cell will never be included in the present database.

4.1.4 Some excluded morphomes

What these high standards for morphomehood are doing, obviously, is attempting to increase the 'cleanliness' of the data at the cost of reducing the number of datapoints in the sample. For 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 the morphome database but had to be ultimately excluded. Consider, for example, the morphological syncretisms in Binandere (Trans-New Guinea) in Table 4.7.

Table 4.7 Partial paradigm of *adu ari* 'fear' in Binandere (King 1927: 23)

	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

As Table 4.7 shows, 1SG and 1PL.INCL are always syncretic in the language. The same applies to 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 /ew/ respectively in the far past but /an/ 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 and Bobaljik

2013).[4] 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 morphome database. In the case of 1PL.EXCL+3PL, the conflation happens between different person values of a single number value 'plural'. This configuration does not qualify here as unmistakably unnatural (see the Teanu example in Table 4.9). 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.

Various other morphological structures have been excluded from the present database due to the problematic nature of the 1.INCL. Consider, for example, the ones in Table 4.8.

Table 4.8 Two morphomes that involve the 1PL.INCL

	Thulung 'drink' (Lahaussois 2002: 162)			*Ngiti* 'mother' (Kutsch Lojenga 1994)	
	SG	DU	PL	SG	PL
1EXCL	ɖu-u	ɖu-tsuku	ɖu-ku	**ɨyà**-du	**ɨyà**-kà
1INCL		ɖu-tsi	ɖuŋ-i		àlɛ-**tsá**-nà
2	ɖu-na	ɖu-tsi	ɖu-ni	**ɨyà**-nʉ	**ɨyà**-kʉ
3	ɖuŋ-y	ɖu-tsi	ɖu-mi	kà-**tsá**-nà	abádhí-**tsá**-nà

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 Ngiti (Sudanic), stem suppletion (stem in bold) and suffixation both follow the same unnatural pattern whereby 3 shares its form with 1PL.INCL. Despite their differences, the morphological affinities in both Thulung and Ngiti rely on the 1PL.INCL cell for morphomicity, as the exclusion of that cell would leave the patterns as morphosyntactically natural. This is the reason why they have been excluded from the present morphome database. Note, however, that morphomes 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 morphomes of Bantawa (Section 4.2.2.2),

[4] From a sample of 241 languages, 31 show an undifferentiated first-person (i.e. 1SG=1PL) and 22 show an inclusive/exclusive difference with no number distinctions (i.e. 1SG=1PL.EXCL).

and Kele (Section 4.2.4.7)). In these cases, the 1PL.INCL cell(s) will only be excluded in the assessment of the pattern's morphosyntactic coherence (see Section 4.3.8), but not for other descriptive measures.

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 the 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 the right to contradict the analysis in a source when this has a motivation clearly at odds with the goal of morphomic analysis.

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 the lines of Table 4.9, and seems to fall short of the orthogonality that characterizes person and number from a logical point of view, as some of the person categories posited for Kariña 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 that 'the first-person lacks a plural' (Mosonyi and 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 exploring is the relation between morphological and extramorphological structure. In line with the rest of this book, 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 semantic way, the paradigmatic distribution of verbal inflectional formatives in the language is shown in Table 4.10.

Table 4.9 Kariña pronominal system as described by Mosonyi and Mosonyi (2000: 407)

	SG	PL
1	aau	–
2	amooro	amoññaaro
3	mojko	mojkaaro
1+2	kümuooro	kümuoññaaro
1+3	na'na	–

Table 4.10 Partial paradigm of Kariña 'cultivate' (Mosonyi and Mosonyi 2000: 425)

	Present			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	voonaae	konoonaano		voonai	noonai	
1INCL	–	kotoonaano	kotoonaatu	–	kotoonai	kotoonatu
2	moonaae	moonaatu		moonai	moonatu	
3	konoonaano	konoonaatu		noonai	noonatu	

In Kariña, 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.[5] This may be a synchronic reflection of a nominal origin, which would explain its morphological affiliation with the 3SG. Despite this whole-word syncretism of 3SG and 1PL.EXCL, and despite the abundance of formatives in the paradigm, no set of cells qualifies for morphomehood here. 3SG and 1PL.EXCL never share any formative (let alone two as required here) to the exclusion of the rest of the paradigm. Most forms in Table 4.18 (e.g. *ko-*, *n-*, *-no*, *-tu*, *-i*) have a paradigmatic distribution which is unnatural but, crucially, unparalleled by other forms. Thus, no morphomes can be identified in Kariña with the demanding criteria adopted in this book.

The last class of structures that will be excluded from the present database involve complements and default forms (i.e. formatives that appear in a majority of cells). As mentioned before, those morphological identities that represent the complement cells of a more paradigmatically restricted morphome, or of a single cell, will not be included in the present morphome database. In the Gavião language (Table 4.11), for example, 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 what might 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). One form (the one shaded in Table 4.11, usually labelled 'long form' in the

Table 4.11 Formal alternations of some Gavião (Macro-Jê) verbs (Amado 2004)

	'eat'		'drink'		'save/keep'		'roast'	
	Final	NonFinal	Final	NonFinal	Final	NonFinal	Final	NonFinal
Past	kor	kor	kom	kom	tʃər	tʃər	tʃər	tʃər
NonPast	ko	kor	ko	kom	tʃə	tʃər	ka	tʃər

[5] E.g. 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').

literature) occurs in non-final positions in the sentence regardless of tense and, also in final position when past.

The mapping of this long form in Gavião to morphosyntactic/semantic properties is, therefore, unnatural as defined in this book. In addition, as Table 4.19 shows, the formal alternations involved are very varied.[6] However, because these stems could be understood to be the default, i.e. the complement of a single stored frequent cell,[7] 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.

4.2 A cross-linguistic database of morphomes

Morphomes, as previous sections have hopefully shown, are a very challenging object of analysis for typology. On the one hand, the phenomenon is only found, as defined or diagnosed in this book, in a relatively small proportion of natural languages (my rough estimate would put this at around 15% of grammatical descriptions). On the other hand, the very term 'morphome' is relatively recent, and even nowadays not widely known and used by field linguists. 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 morphomes within a reasonable period of time.

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 usually has to read through all the morphology and inflection-related sections of a grammar to find out whether the language in question has or lacks morphomes. The relative rarity of the phenomenon, obviously, means that one will usually have to read quite a few grammars to find one example which deserves to be included in this database according to the criteria that were presented in the Section 4.1.

Because the main problem with morphomes is the scarcity of data, language sampling is particularly tricky. A 'probability sample' (Bakker 2011) therefore seems inadequate for our present purposes. Because of this, the figure of around

[6] One often finds the addition of segments /r/ (most frequent), /n/, or /m/ in the long form. Vocalic changes also occur (e.g. *kwɨr/kwa* 'hit', *tʃəm/tʃa* 'bite'), as well as consonant changes at various locations within the word (e.g. *pus/puj* 'arrive', *jəmjõr/jəmⁿgõr* 'pay', *pemter/amte* 'dream'), all the way to suppletion (e.g. *tʃər/ka* 'roast').

[7] 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-*'mother', and between *imja* and *imen-* 'name' 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 2011: 148–9) and most likely descend via sound changes from an unremarkable zero/suffixed configuration (see Section 3.1.1.3).

15% that I mentioned is everything I will have to offer in that sense. It goes without saying that highly isolating or highly agglutinative languages will lack morphomes more frequently than the cross-linguistic average, whereas highly fusional, morphologically complex languages will constitute the best breeding ground for morphomes. 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 morphome has been included in this synchronic database as long as it fulfilled the criteria in Section 4.1. Only cognate morphomes have been excluded when these agreed on their paradigmatic configuration.[8]

Figures 4.1 and 4.2 show the geographical distribution of the languages in this database. It can be seen that, despite an understandable slight European bias resulting from more extensive documentation of these languages, the sample is by and large balanced geographically. Out a total of **79 languages**, 10 are from Africa, 15 from Asia, 14 from Europe, 17 from the Americas, and 23 from Australasia.

The genetic diversity of the sample is also considerable, with **37 highest-level stocks** represented. In terms of the distribution of individual languages across these, 11 languages are Indo-European, 7 Sino-Tibetan, 6 Trans-New-Guinean, 5 Austronesian, 5 Oto-manguean, 4 Uralic, 4 Nilotic, 3 Afro-Asiatic, 2 Nakh-Daghestanian, 2 Yam, 2 Koiarian, 2 Chicham, and the rest belong to different stocks.

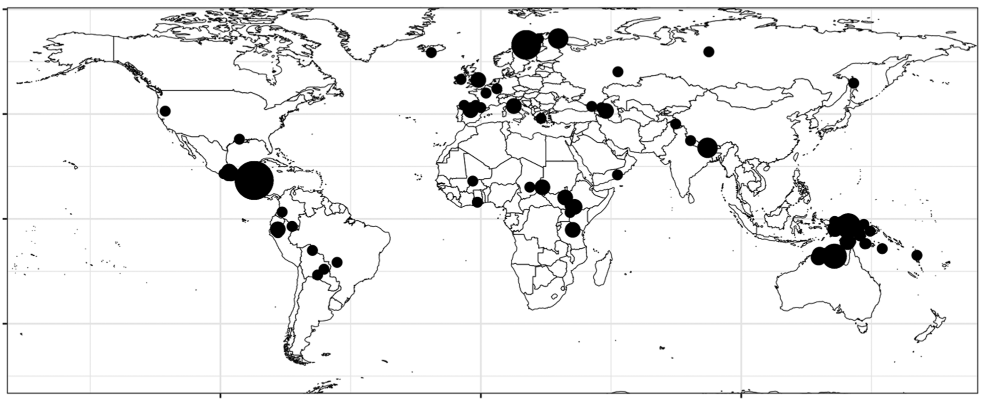

Figure 4.1 Geographical location of the languages in the database by number of morphomes

[8] E.g. because the Spanish, Portuguese, and Italian N-morphomes all have the same paradigmatic extension, only one of them (the Spanish one in this case) has been included in this database.

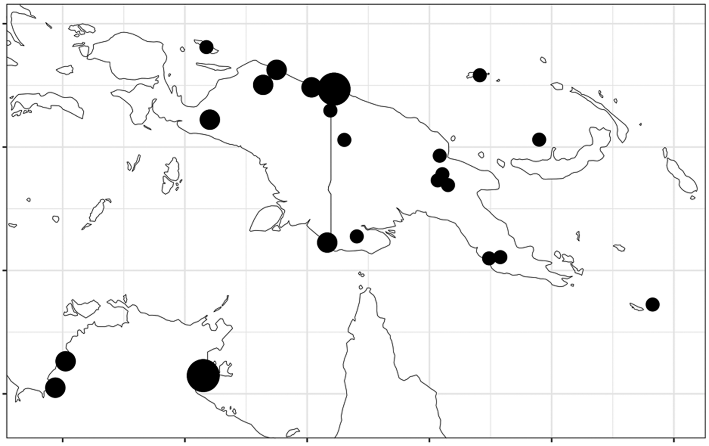

Figure 4.2 New Guinea zoom-in

In many of the languages (25, 32%), more than one structure qualified as morphomic as defined here. This percentage is substantially higher than the overall cross-linguistic prevalence of morphomes (estimated at around 15%), which means that these structures are unevenly distributed across the world's languages. Thus, having one morphome makes a language more likely to have a second, or a third. The multiple occurrences of the phenomenon in some languages brings the total to **120 morphomes in this database.**

The remainder of this (long) section will present a description of all these morphomes organized by geographical area and by language in alphabetical order.

4.2.1 Africa

4.2.1.1 Daasanach (Tosco 2001)

As briefly shown before (see Table 2.63) 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 cryptic labels suggest, the paradigmatic distribution of the two forms is chaotic from a morphosyntactic perspective (see Table 4.12). The actual formal alternations involved are also quite diverse.

As the paradigms in Table 4.12 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. There are many different ways in which Form A and Form B may differ in Daasanach. Apart from the ones in Table 4.20 we have pairs like *yes/ces*

Table 4.12 Partial paradigms of three verbs in Daasanach (Tosco 2001: 112, 140, 172)

	'open' perfect		'die' imperfect		'drink' perfect	
	SG	PL	SG	PL	SG	PL
1EXCL	furi	fuɗɗi	kufuma	kufunanna	ʃii	ʔiyyi
1INCL		furi		kufuma		ʃii
2	fuɗɗi	fuɗɗi	kufunanna	kufunanna	ʔiyyi	ʔiyyi
3F	fuɗɗi	furi	kufunanna	kufuma	ʔiyyi	ʃii
3M	furi	furi	kufuma	kufuma	ʃii	ʃii

'kill.PFV', *guurma/guuranna* 'migrate.IPFV', *leeði/leeti* 'fall down.PFV', yeeðe/ceete 'say/become.IPFV', etc. Both unnaturalness and systematicity, therefore, are high for this morphome.

This bizarre system originated from a relatively unremarkable person–number agreement system still present in more conservative Cushitic languages (see Table 4.13).

Table 4.13 Agreement affixes of Oromo and Somali (Cushitic)

	Oromo (Ali and 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-een	t-idh-i	t-idhaahd-een
3F	déem-t-e	déem-an	keen-t-ay	keen-een	t-idh-i	y-idhaahd-een
3M	déem-e	déem-an	keen-ay	keen-een	y-idh-i	y-idhaahd-een

Leaving the 1PL aside, where clusivity complicates the picture, the contexts that take Form B (e.g. *fuɗɗi*) in Daasanach are those that take consonantal affixes in more conservative Cushitic languages, while those that take Form A (e.g. *furi*) are those that take vocalic or zero affixes. The morphological alternations found synchronically in Daasanach, at both the right and the left edges of the stem, are for the most part readily interpretable as the result of run-of-the-mill sound changes that, through the history of the language, affected the original stem consonants differently in different phonological environments:

**yes* > **yes* > *yes* ('kill' Form A Perfect, Tosco 2001)
**t-yes* > **tyes* > *ces* ('kill' Form B Perfect, Tosco 2001)

**guuram-a* > **guurama* > *guurma* > *guurma*
('migrate' Form A Imperfect, Tosco 2001)
**guuram-t-a* > **guuramta* > *guuranta* > *guuranna*
('migrate' Form B Imperfect, Tosco 2001)

Analogy, however, must also have played some role in the emergence of the new system (see Sasse 1976). Thus, speakers of Daasanach, when faced with these stem alternations, appear to have responded by getting rid of any other morphology, and reorganized their person–number paradigm into one with only two arbitrarily distributed forms.[9]

Daasanach1: 1SG/3SG.M/3PL
Daasanach2: 1PL/2/3SG.F

4.2.1.2 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 other times, it is instantiated by an overt formative, which can have different phonological forms depending on tense or verb type.

The system in Table 4.14 (reminiscent of that in Ayoreo, see Section 4.2.5.3) 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 3.1.1.3 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 so (Cysouw 2003). The idiosyncrasy of this system lies 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.

Table 4.14 Some partial paradigms in Mongo Daju (Avilés 2008)

	'drink' present		'drink' progressive		'hide oneself' present	
	SG	PL	SG	PL	SG	PL
1EXCL	uɾ-o	uɾ-ciga	uɾ-ca	uɾ-ciga	nol-wa	nol-din-ciga
1INCL		uɾ-cina		uɾ-cina		nol-din-cina
2	uɾ-o	uɾ-cini	uɾ-ca	uɾ-cini	nol-wa	nol-din-cini
3	uɾ-o	uɾ-o	uɾ-ca	uɾ-ca	nol-wa	nol-wa

Note: The 1DU forms *ur-cik* and *nol-din-cik* have not been presented 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 in question.

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 *alaʃʃina*<**alas(e)-cina* '1PL.INCL.throw', Avilés 2008: 71–2). Analogical processes may have also played a role (e.g. in the case of the reflexive).

[9] See also Iraqw (Section 4.2.1.4) for an intermediate system, i.e. a system where a form A/form B stem alternation has emerged but where affixes still disambiguate most of the values that are collapsed (i.e. whole-word-syncretized) in Daasanach.

4.2.1.3 Fur (Waag 2010)

In Fur (Furan, Sudan) verbal inflection (see Table 4.15), there is a morphological affinity of the 3SG and the 3PL non-human, which are opposed to the rest of the paradigm.

Table 4.15 Partial paradigms of three Fur verbs (Waag 2010)

	'tie' imperfective		'hang' imperfective		'grind' imperfective	
	SG	PL	SG	PL	SG	PL
1	ʔirg-ɛl[a]	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	lɪg-ɛl	kalg-ɛl-ɪ	kɔɔn	kawnɛ
3.NHUM	rig-ɛl	rig-ɛl-ɪ	lɪg-ɛl	lɪg-ɛl-ɪ	kɔɔn	kɔɔnɛ

[a] The glottal stop occurs automatically as a subphonemic onset before a vowel (Jakobi 1990: 42; Waag 2010: 115) so the 1SG should probably be thought of as unprefixed.

Although a few exceptions exist where the two are identical (e.g. 3SG *rɪg-ɛl* vs 1SG *ʔa-rɪg-ɛl* 'lie in waiting', Waag 2010: 125), almost all verbs in Fur show stem alternation according to the pattern in Table 4.23. As these 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 morphological perspective. In 'tie', for example, we find consonant/vowel metathesis, in 'hang', vowel deletion/epenthesis, and in 'grind' weak suppletion involving both initial consonant and vowel apophony. The two sets of cells also differ frequently in their tone.

Fur 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, so forms like 3SG **irg-ɛl* would also be ill-formed. 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 (e.g. Beaton 1968), the formation of the 3SG in 'grind' will involve the insertion of /k/ as a prefix.

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 *paarɛl*) but have a different stem (1SG *ʔaarɛl* vs *ʔawrɛl* respectively) elsewhere.

Because of the great number of processes and forms involved and because of the 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 morphome, although it also seems related to the presence of prefixes in some person–number forms and the absence of prefixation from others (Section 3.1.1.3).

Fur1: HUM.SG/NHUM
Fur2: 1SG/2SG/PL

4.2.1.4 Iraqw (Mous 1992)

Verbs in the Iraqw (South Cushitic) language 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.

As illustrated in Table 4.16, the two sets of cells show morphological differences which can be very diverse (*a/eer*, *r/t*, *ay/g* in Table 4.16, 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 facts suggest that this pattern is robustly morphomic.

Table 4.16 Present paradigms of three Iraqw verbs (Mous 1992: 156–7)

	'leave' present		'follow' present		'eat' present	
	SG	PL	SG	PL	SG	PL
1	máw	mawáan	eehár	eeharáan	ʕáay	ʕaayáan
2	méer	méeraʔ	eehát	eehatáʔ	ʕág	ʕagáʔ
3M	máy	mayáʔ	eehar	eeharír[a]	ʕaay	ʕaayír
3F	méer	mayáʔ	eehát	eeharír	ʕág	ʕaayír

[a] There are two alternative forms for the 3PL in these verb and others. The two alternatives, however (*eehariyáʔ* and *eeharír* in this verb), always share the exponence (/r/ in this case) which is at stake here.

Most of the alternations we see today, however, can be traced back to regular sound changes. Following the common Afroasiatic pattern (still readily observable, e.g. in Table 4.21, 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 branch 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 **ʕaag-t* > **ʕaag-t* > **ʕag-t* > *ʕag*
eat.3SG.M **ʕaag-i* > **ʕaay-i* > **ʕaay-i* > *ʕaay*

This development is parallel to (but completely independent from) the emergence of the morphomic agreement system described before for the East Cushitic language Daasanach.[10] It seems, thus, that affixal configurations like the (accidental?, see Harbour 2008) Afroasiatic homophony of 2 and 3SG.F *-t* are particularly good breeding grounds for morphomes.

Iraqw1: 2/3SG.F
Iraqw2: 1/3SG.M/3PL

4.2.1.5 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 prefixes in Table 4.17. Passive prefixes seem to be derived from the active ones. Whereas the active and passive are the same in the third-person, first- and second-person passive forms are formed by adding segments to active forms. The actual forms being added, however, differ from one mood to another, and from 1PL to 2. In the subjunctive, for example, 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.

Table 4.17 Karamojong Conjugation 1, past, passive prefixes (Novelli 1985: 202)

	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-

4.2.1.6 Nuer (Reid 2019)

In the Nilotic language Nuer (and in the very closely related Reel, and in somewhat less closely related Dinka, see Reid 2010 and Andersen 1993), tone, vowel apophony, and vowel lengthening participate prominently in verbal stem alternation patterns in both inflection and derivation. In the domain of person–number

[10] Conservative languages in both East Cushitic (e.g. Oromo, see Ali and 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 1,000 km, thus making areal influences similarly unlikely.

Table 4.18 Inflectional paradigms of three transitive verbs (Reid 2019)

	kʌ̤́ʌ̤p 'catch'		*lúj̤* 'kill in secret'		*lé̤p* 'open.APPL'	
	SG	PL	SG	PL	SG	PL
1EXCL	ká̤a̤p-ʌ̤́	ká̤a̤p-kɔ̄	lṳ̄ɔj̤-ʌ̤́	lɔ̄ɔj̤-kɔ̄	lép-ʌ̤́	lép-kɔ̄
1INCL	–	ká̤a̤p-nɛ́	–	lɔ̄ɔj̤-nɛ́	–	lép-nɛ́
2	kʌ̤́ʌ̤p-í̤	ká̤a̤p-ɛ̄	líj-í̤	lɔ̄ɔj̤-ɛ̄	lé̤p-í̤	lép-ɛ̄
3	kʌ̤́ʌ̤p-ɛ́	ká̤a̤p-kɛ̄	líj-ɛ́	lɔ̄ɔj̤-kɛ̄	lé̤p-ɛ́	lép-kɛ̄

inflection, vowel length and tone have natural distributions and split the paradigm neatly into SG vs PL. Vowel apophony (the distinction between so-called vowel grades A and B), however, is morphomic.

In transitive verbs (see Table 4.18), the 1SG and the PL cells have a stem vowel different from the rest. The vowel in one stem is almost perfectly predictable from the vowel in the other, with the vowel in 1SG+PL most often being a diphthongized version (with a lower offglide) of the one in 2/3SG: /ɪ/>/ɪɛ/, /ɛ/>/ɛa/, etc. In the case of /e̤/ and /o̤/, these vowels lose their breathiness instead (i.e. become /e/ and /o/ respectively), and in the case of /ʌ̤/, the vowel is lowered to /a̤/. The vowel /a/ is the only one which does not change in the 1SG+PL, probably because it cannot be lowered further.

Intransitive verbs (also derived intransitives like antipassives) show a slightly different pattern regarding these stem vowel alternations (see Table 4.19). In these verbs, the modified stem vowel extends only through 1SG, 1PL, and 2PL, and unlike in transitive verbs, it is not present in the 3PL. The formal alternations involved, however, are the same.

Table 4.19 Inflectional paradigms of two intransitive verbs (Reid 2019)

	gɔ̤ɔ̤̄ɔ̤r 'write.AP'		*tʌ̤́t̪* 'mould.AP'	
	SG	PL	SG	PL
1EXCL	gɔ̤̄a̤a̤a̤r-ʌ̤́	gɔ̤̄a̤a̤a̤r-kɔ̄	tá̤t̪-ʌ̤́	tá̤t̪-kɔ̄
1INCL	–	gɔ̤̄a̤a̤a̤r-nɛ́	–	tá̤t̪-nɛ́
2	gɔ̤̄ɔ̤ɔ̤r-í̤	gɔ̤̄a̤a̤a̤r-ɛ̄	tʌ̤́t̪-í̤	tá̤t̪-ɛ̄
3	gɔ̤̄ɔ̤ɔ̤r-ɛ́	gɔ̤̄ɔ̤ɔ̤r-kɛ̄	tʌ̤́t̪-ɛ́	tʌ̤́t̪-kɛ̄

The fact that these inflectional diphthongizations are not found outside the Dinka–Nuer–Reel language family suggests that it is a relatively recent innovation. Although the details are not completely clear, the alternations must have emerged via sound change, triggered by the form of the following person suffixes. The forms of the singular-person markers in Western Nilotic languages (e.g. the vowels /a/, /i/, /ɛ/ for the 1SG, 2SG, and 3SG respectively in Anuak, see Reh 1996) largely agree in showing a low vowel in 1SG and a non-low vowel in the 2SG and 3SG. An anticipatory vowel assimilation to this low vowel (/ɪ/>/ɪɛ/, /ɛ/>/ɛa/ etc.) would explain the stem vowel alternations found in the singular in Nuer. The ones in the plural are more problematic due to the greater instability of those person suffixes in Western Nilotic (compare Nuer -kɔ̄ -nɛ́ -ɛ̄ -kɛ̄ to Dinka -kṳ́ -kṳ́ -ká̤ -ké̤ and Anuak -ɔ́ -wā -wū -gī). The general diachronic hypothesis is strengthened, however, by the observation that in Dinka (unlike in Nuer-Reel), only the 2PL (-ká̤) shows this lowered/diphthongized stem vowel.

Nuer1: 1SG+PL
Nuer2: 1SG+1PL+2PL

4.2.1.7 Turkana and Toposa (Dimmendaal 1991)

In Turkana (Nilotic) inflection, partial and whole-word syncretisms are widespread. There are two inflectional classes in the language, shown in Table 4.20. The prefixal syncretism of 1SG.PRS+1SG.PAST+3.PAST observed in class 1 is repeated in class 2 with the prefix *e-* (see Table 4.20), 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 earlier causative prefix *i-* in class 2 verbs, which became unproductive and 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).

Table 4.20 Partial paradigm of 'go' in Turkana (Dimmendaal 1991: 283–4)

	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)

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, Toposa in Table 4.21, however, shows an interesting deviation from this family-wide pattern in that the 1PL form does not have the expected *ki-* but takes a form that patterns as 3.

What happened in Toposa is that a formerly impersonal construction based on the third-person morphologically came to replace the original 1PL.[11] Because of the pre-existing patterns of syncretism, this did not result (only) in the identity of

[11] This constitutes a cross-linguistically recurrent development. Consider earlier discussion on Kariña (Table 4.17) and better-known cases like the contemporary uses of the impersonal in colloquial French, where the etymological 1PL (e.g. *nous allons*) is replaced by the impersonal (i.e. *on va*), a third-person morphologically.

Table 4.21 Person–number prefixes in two Turkana varieties (Dimmendaal 1991: 290)

	Turkana								Toposa							
	Class 1				Class 2				Class 1				Class 2			
	PRS		PAST		PRS		PAST		PRS		PAST		PRS		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-

3 and 1PL but spread the morphomic pattern in Turkana to the 1PL.PAST. This new pattern in Toposa is also morphomic and has been included in the database.

Turkana: 1SG/3.PAST
Toposa: 1SG.PRS/1.PAST/3.PAST

4.2.1.8 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.[12] Observe the partial paradigm of the verb *tɔ́* 'buy' in Table 4.22. 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 (*tɔ́* > *tɔ́ɔ́*) and the suffix *-yɛ́* both occur in the past affirmative and in the perfect negative. The latter morphological affinity, due to its allomorphy, qualifies for morphomehood here.

Table 4.22 Past and perfect forms of 'buy' in Twi (Stump 2015: 136)

	Before complement		Elsewhere	
	Affirmative	Negative	Affirmative	Negative
Past	tɔ̀-ɔ́	à-ǹ-tɔ́	tɔ̀-ɔ́-yɛ̀	à-ǹ-tɔ́
Perfect	à-tɔ́	ǹ-tɔ́-ɔ́	à-tɔ́	ǹ-tɔ́-ɔ́-yɛ̀

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 fewer distinctions in the negative than in the positive (4 vs 9 respectively

[12] 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.

according to Osam 1994: 103). The second is the incompatibility of the past (sometimes labelled 'completive') and negation in related languages (e.g. in Anufo, see Smye 2004: 88).

The diachrony I would like to propose is, thus, that the tense nowadays labelled 'past' (also sometimes 'remote past') in Twi must have formerly expressed completive aspect strictly and must have been semantically incompatible with negation at this stage. One can understand the logic of this: what has been completed cannot be expected not to have happened at all. At a later stage, the semantics of the tense must have drifted to include paste 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 (see Table 4.23).

Table 4.23 Proposed system of morphological oppositions in Pre-Twi 'buy'

	Before complement		Elsewhere	
	Affirmative	**Negative**	**Affirmative**	**Negative**
Past	*tɔ̀-ɔ̀	*à-ǹ-tɔ́	*tɔ̀-ɔ̀-yɛ̀	*à-ǹ-tɔ́
Perfect	*à-tɔ́	*à-ǹ-tɔ́	*à-tɔ́	*à-ǹ-tɔ́

The developments up to this point are not exceedingly surprising, and the system at this stage would have been the same as the one found in closely related Anufo (Smye 2004: 88) and in comparable TAM/negation morphology in Twi itself in other tenses.[13]

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.[14] It might make functional sense to try to (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 initially) that made changing the perfect negative 'preferable' to changing the past negative.

[13] E.g. 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-*.

[14] 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.

4.2.1.9 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 paradigms in Table 4.24. In the inflection of many tenses there is a morphological opposition of SG+2PL and 1PL+3PL. Both sets of cells, as Table 4.24 illustrates, may take exponents of their own. For the purposes of the present book, SG+2PL qualifies as a morphome.

Table 4.24 Partial paradigms of three Yorno-So Dogon verbs (Heath 2014: 209, 214, 223)

	'fall', imperfective		'hit', imperfect negative		'see', experiential perfect negative	
	SG	PL	SG	PL	SG	PL
1	númɔ̀-jɛ̀-m	númɔ̀y	lágà-lɛ̀-m	lágàỳnɛ̀	yè:tɛ̀-r-úm	yè:tɛ̀nɛ́
2	númɔ̀-jɛ̀-w	númɔ̀-jɛ̀-y	lágà-lɛ̀-w	lágà-lɛ̀-y	yè:tɛ̀-r-úw	yè:tɛ̀-r-íy
3	númɔ̀-jɛ̀	númɔ̀y	lágà-lɛ̀	lágàỳnɛ̀	yè:tɛ̀-ŕ	yè:tɛ̀nɛ́

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 in the family (e.g. Togo Kan, see Heath 2011) do not have it. What all Dogon languages do have is some sort or 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 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.

4.2.2 Asia

4.2.2.1 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 (Table 4.25) to transitive verbs (Table 4.26).

Table 4.25 Athpariya 'go', intransitive positive non-past (Ebert 1997: 163)

<table>
<tr><th></th><th>SG</th><th>DU</th><th>PL</th></tr>
<tr><td>1EXCL</td><td rowspan="2">khat-naʔa</td><td>khat-ciciŋa</td><td>khad-itiŋa</td></tr>
<tr><td>1INCL</td><td>khat-cici</td><td>khad-iti</td></tr>
<tr><td>2</td><td>a-khat-yuk</td><td>a-khat-cici</td><td>a-khad-iti</td></tr>
<tr><td>3</td><td>khat-yuk</td><td>khat-cici</td><td>u-khat-yuk</td></tr>
</table>

Table 4.26 Athpariya 'beat', transitive positive non-past, 3SG object (Ebert 1997: 180)

<table>
<tr><th></th><th>SG</th><th>DU</th><th>PL</th></tr>
<tr><td>1EXCL</td><td rowspan="2">lems-uŋtuŋ</td><td>lem-cucuŋa</td><td>lems-umtumma</td></tr>
<tr><td>1INCL</td><td>lem-cucu</td><td>lems-umtum</td></tr>
<tr><td>2</td><td>a-lems-utu</td><td>a-lem-cucu</td><td>a-lems-umtum</td></tr>
<tr><td>3</td><td>lems-utu</td><td>lem-cucu</td><td>o-lems-utu</td></tr>
</table>

Interestingly, this suffixal syncretism of 2SG, 3SG, and 3PL is also found, albeit with completely different exponents (*-no* and *-oko*), in the closely related language Chintang, which suggests that we are dealing with a stable morphomic affinity.

As Schackow (2016: 230–31) explains (see also Section 3.3), 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 must have lacked suffixes originally (zeroes can still be found there in other East Kiranti languages like Puma, Limbu, and Bantawa, see the following Section 4.2.2.2). Be that as it may, the set of values where these formatives appear synchronically does not constitute a natural class and counts, therefore, as a morphome for our present purposes.

4.2.2.2 Bantawa (Doornenbal 2009)

A trademark feature of Kiranti languages (see also Athpariya in Section 4.2.2.1) is that they display stem alternation in the verb. In East Kiranti, to which Bantawa belongs, stem alternation is correlated with the presence of consonant- or vowel-initial suffixes after the stem (Herce 2021a).

A stem alternant (*kon-*) appears in Table 4.27 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*. At other times, the prevocalic

Table 4.27 Paradigm of Bantawa 'walk' non-past (Doornenbal 2009: 391)

	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

stem is characterized by a segment which is absent from the preconsonantal stem. This can be *s*, *t*, *w*, *y*, and *ʔ*.

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 prevocalic stem alternant (e.g. *kol-*) appears in this tense. However, there is in this domain, interestingly, another form (the suffix *-a*, see Table 4.28) which has the same paradigmatic configuration as stem alternation in the present.

Table 4.28 Paradigm of Bantawa 'walk', past (Doornenbal 2009: 391)

	SG	DU	PL
1EXCL	kol-a-ŋ	kol-a-ca	kol-inka
1INCL		kol-a-ci	kol-in
2	ti-kol-a	ti-kol-a-ci	ti-kol-in
3	kol-a	kol-a-ci	mi-kol-a

Although such reasoning would have problems of and by itself (see Section 2.4), because of its coextensivity with coherent phonological environments (i.e. _V vs _C), one could argue that the stem alternation in Table 4.27 is phonologically conditioned and thus not morphomic. However, because 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 4.27 and 4.28) is what is found in the inflection of intransitive verbs. However, the exact same morphological contrasts are found, albeit with a different paradigmatic configuration, in transitive verbs (see Tables 4.29 and 4.30). 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 4.28–4.30, that an alternation between zero and *-a* indicates tense in those (darker shaded) paradigm cells where those forms appear while the

Table 4.29 Paradigm of Bantawa 'take' non-past (Doornenbal 2009: 397)

	SG	DU	PL
1EXCL	kʰatt-u-ŋ	kʰat-cuʔa	kʰatt-u-mka
1INCL		kʰat-cu	kʰatt-u-m
2	tɨ-kʰatt-u	tɨ-kʰat-cu	tɨ-kʰatt-u-m
3	kʰatt-u	ɨ-kʰat-cu	ɨ-kʰat

Table 4.30 Paradigm of Bantawa 'take' past (Doornenbal 2009: 398)

	SG	DU	PL
1EXCL	kʰatt-u-ŋ	kʰatt-a-cuʔa	kʰatt-u-mka
1INCL		kʰatt-a-cu	kʰatt-u-m
2	tɨ-kʰatt-u	tɨ-kʰatt-a-cu	tɨ-kʰatt-u-m
3	kʰatt-u	ɨ-kʰatt-a-cu	ɨ-kʰatt-a

rest (marked with *-i* in intransitives and with *-u* in transitives) do not make tense distinctions.[15]

Given the criteria that are being used in the present book, three different morphomes 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 Doornenbal (2009: 134), stem alternation is present in around 92% of the lexemes in Bantawa, although only 16.6% have (like *kon-* in Table 4.35) 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. *kʰatt-* vs *kʰat-*). This refers exclusively 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

[15] 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 morphomes (see e.g. the distinction between present and progressive in Daju in Table 2.64).

4.2.2.3 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.

While other (partial) syncretisms do not extend to every context, there is a set of cells (M.SG, X.SG, Y.SG, and Y.PL) for which there is particular systematicity (see Table 4.31). It is worth noting that class Y nouns are less compatible with pluralizability because they are often abstract or mass nouns (Yoshioda 2012: 33).

Table 4.31 Some Burushaski partial paradigms (Yoshioda 2012)

	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

4.2.2.4 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 4.32 illustrates, and also, with slightly different suffixes (*-de* instead of *-he*), in transitive verbs.

Table 4.32 Paradigm of Darma *ra* 'come' (Willis 2007: 350–56)

	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-ɟu	ra-ɟu

The formal affinity shaded in Table 4.32 is, therefore, morphomic. The situation in closely related languages is 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), various /n/-containing syncretisms exist.

As Table 4.33 shows, these may involve (i) all plural cells, (ii) PL+2SG, (iii) 1PL+2 (as in Darma), and 2SG+3SG. The diachronic evolution of these forms in West Himalayish is not clear at all to me (although see Saxena 1992). The 'mess'

Table 4.33 Rongpo verb *rha-pəŋ* 'come' in various tenses (Sharma 2001b: 226)

	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

observed in the distribution of formatives in related languages is probably derived from the loss of an earlier bi-argumental agreement system. It might be (although this is largely conjectural) that Darma has managed to impose some order on this mess by generalizing a single paradigmatic distribution of /n/ and by organizing the allomorphy of tense markers along the same lines as well.

4.2.2.5 Jerung (Opgenort 2005)

Jerung (Western Kiranti, Sino-Tibetan) has a morphologically determined pattern of stem alternation which involves the same (longer) stem in the SG and 3.NSG. As the paradigm in Table 4.34 illustrates, this pattern of stem alternation can involve both final consonant(s) and stem vowel. These alternations are confined to transitive verbs and most often involve 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).

Table 4.34 Paradigm of Jerung 'give', 3SG patient (Opgenort 2005: 330)

	SG	DU	PL
1EXCL	gɔk-ma	go-cum	go-kum
1INCL	-	go-cim	go-kim
2	gɔk-nim	go-cim	go-nimme
3	gɔk-t-im	gɔk-cim	gɔk-me

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. This might be the origin of the 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 (see 2DU vs 3DU).

4.2.2.6 Ket (Georg 2007)

In Ket (Yeniseian) inflectional morphology, the neuter plural is associated with the same morphology as the singular. Sometimes (see Table 4.35), this identity of neuter singular and plural leads to a syncretism with the feminine singular as well. This syncretism is morphosyntactically unnatural but is repeated across 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); it is just their agreement targets that fail to do so.

Table 4.35 Some inflectional formatives in Ket (Georg 2007: 104, 119, 268)

	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-

4.2.2.7 Khaling (Jacques et al. 2011; Jacques 2017)

The verbal inflectional morphology of Khaling (Kiranti, 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 labelled morphomic by the criteria I have set here.

Observe how, in Table 4.36, different forms may differ in the stem they use. Despite what might seem to be the case in that table, none of these alternations is a regular phonological rule of the language. Both are purely morphological, which is revealed by 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, nor 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, 2PL, and 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 allomorphic variation to classify this morphological affinity as systematic according to the criteria that have been set here.

Another morphological affinity in Khaling (one which affects a superset of the cells discussed in Table 4.36 and which is instantiated by similar forms) also deserves inclusion into the present database of morphomes.

Table 4.36 Two Khaling verbs, non-past, intransitive (Jacques et al. 2011: 1102, 1148)

	'have enough'			'look nice'		
	SG	DU	PL	SG	DU	PL
1EXCL	soɔ̂m-ŋʌ	səp-u	soɔp-kʌ	bū:-ŋʌ	biŋ-u	bʌŋ-kʌ
1INCL		səp-i	soɔp-ki		biŋ-i	bʌŋ-ki
2	ʔi-soɔp[a]	ʔi-səp-i	ʔi-soɔ̂m-ni	ʔi-bʌŋ	ʔi-biŋ-i	ʔi-bū:-ni
3	soɔp	səp-i	soɔ̂m-nu	bʌŋ	biŋ-i	bū:-nu

[a] 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 morphome database.

Table 4.37 Paradigm of Khaling 'sleep' (*ʔip-*) past, reflexive (Jacques 2017: 6)

	Present			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	ʔʌ̂m-si-ŋʌ	ʔip-si-ju	ʔʌp-si-kʌ	ʔʌ̂m-tʌsu	ʔip-sî-jtu	ʔʌp-si-ktʌkʌ
1INCL		ʔip-si-ji	ʔʌp-si-ki		ʔip-sî-jti	ʔʌp-si-ktiki
2	ʔi-ʔʌ̂m-si	ʔi-ʔip-si-ji	ʔi-ʔʌ̂m-si-ni	ʔi-ʔʌ̂m-tɛ-si	ʔi-ʔip-sî-jti	ʔi-ʔʌ̂m-tɛ-nnu
3	ʔʌ̂m-si	ʔip-si-ji	ʔʌ̂m-si-nu	ʔʌ̂m-tɛ-si	ʔip-sî-jti	ʔʌ̂m-tɛ-nnu

In Khaling (see Table 4.37), 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 tone. In the past, these cells are different from the rest in that the reflexive suffix *-si* does not appear immediately after the stem. Instead, the past suffix *-t(ɛ)* 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 to the mercy of analogical processes and later sound changes.

Khaling1: 1SG/2PL/3PL
Khaling2: SG/2PL/3PL

4.2.2.8 Khinalugh (Kibrik 1994)

According to their agreement morphology in the verb, Khinalugh 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 in III or IV is less systematic than in the other two genders). The agreement markers that reveal this gender division, however, are syncretic in morphomic ways.

As Table 4.38 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.

Table 4.38 Gender agreement morphology in Khinalugh (Kibrik 1994: 387)

	Set 1		Set 2		Set 3	
	SG	PL	SG	PL	SG	PL
I Male	Ø	b	j	v	h	f
II Female	z	b	z	v	s	f
III Animate	b	Ø	v	j	f	h
IV Inanimate	Ø	Ø	j	j	h	h

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, comparative evidence suggests that are very old and quite stable diachronically. The syncretism of I/II.PL+III.SG in Khinalugh, for example, is also found in other (relatively distantly related) Daghestanian languages like Tsakhur (Schulze 1997), Hunzib (van den Berg 1995), and Archi (Chumakina and Corbett 2015), and even has cognates in the Nakh branch.

The antiquity of these patterns does not preclude the occasional reconfiguration of these morphological gender–number morphomic classes. The other morphomic class of Khinalugh, for example, appears to have involved the fusion of two different exponents, since I.SG has a non-syncretic exponence in the related languages mentioned above. The merger of these two morphological classes into one in Khinalugh may have resulted from their exponents falling together in some of their allomorphs (maybe as zero in Set 1) and this identity being subsequently extended to the other allomorphs. This remains, however, speculative.

Khinalugh1: IPL/IIPL/IIISG
Khinalugh2: ISG/IIIPL/IVSG/IVPL

4.2.2.9 Mehri (Rubin 2010)

As in other Semitic languages, the verbal conjugation of Mehri is characterized by the heavy use of vowel apophony on a more or less invariable consonantal skeleton. There is, in the perfect, a syncretism of the third singular masculine and the third plural feminine.

Affixally (see Table 4.39), both cells are characterized merely by the absence of an affix, which would not qualify as a systematic formal identity here. The two forms, however, also behave alike in every verb concerning ablaut, sometimes, as in the verbs 'put on the fire' and 'break', sharing a stem vowel to the exclusion of every other paradigm cell.

Table 4.39 Perfect paradigms of two Mehri verbs (Rubin 2010: 91, 94)

	'put on the fire'			'break'		
	SG	DU	PL	SG	DU	PL
1	arákb-ək	arákb-əki	arákb-ən	ṯə́br-ək	ṯə́br-əki	ṯə́br-ən
2M	arákb-ək	arákb-əki	arákb-əkəm	ṯə́br-ək	ṯə́br-əki	ṯə́bər-kəm
2F	arákb-əš	arákb-əki	arákb-əkən	ṯə́br-əš	ṯə́br-əki	ṯə́bər-kən
3M	arōkəb	arkəb-ē	arákb-əm	ṯībər	ṯəbr-ō	ṯə́br-əm
3F	arkəb-ēt	arkəb-tē	arōkəb	ṯəbr-ūt	ṯə́bər-tō	ṯībər

4.2.2.10 Nivkh (Gruzdeva 1998; Nedjalkov and 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) can take two forms, and the values with which each occurs do not constitute a natural class.

As Table 4.40 illustrates, the first-person singular and the plural subjects occur with the same form. This suffix varies (/t/ vs /n/) according to tense, so the formal identity of 1SG+PL can be classified as 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 broader system in the language (Nedjalkov and Otaina 2013: 15–16). In synchrony, however, the alternations between the forms in Table 4.48 do not correlate to different phonological environments, as all of them simply follow the verb stem synchronically. The pattern is, therefore, morphomic.

Table 4.40 Nivkh (East Sakhalin) converb inflection (Nedjalkov and Otaina 2013: 40–42)

	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

4.2.2.11 Northern Akhvakh (Creissels 2008)

The perfective positive suffixes in Northern Akhvakh (Nakh-Daghestanian) are characterized by allomorphy along various orthogonal axes (see Table 4.41). Corresponding to a conjunt/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 distinction is understood to be related to the epistemological properties of speech act participants in particular speech acts and is thus not morphomic.

Table 4.41 Perfective positive paradigm of two verbs (Creissels 2008)

	'grasp'				'do'			
	Conjunct		Disjunct		Conjunct		Disjunct	
	SG	PL	SG	PL	SG	PL	SG	PL
M	w-ux̯-ada	ba-x̯-idi	w-ux̯-ari	ba-x̯-iri	gw-ēda	guj-idi	gw-ēri	guj-iri
F	j-ix̯-ada	ba-x̯-idi	j-ix̯-ari	ba-x̯-iri	gw-ēda	guj-idi	gw-ēri	guj-iri
N	b-ix̯-ada	r-ix̯-ada	b-ix̯-ari	r-ix̯-ari	gw-ēda	gw-ēda	gw-ēri	gw-ēri

Each of those morphemes, however, is in turn subject to various allomorphies. The gender and 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:/), yielding further allomorphy.

4.2.2.12 Sunwar (Borchers 2008)

Like other Western Kiranti languages (Sino-Tibetan), Sunwar shows morphological stem alternations in some of its verbs. In the case of the verb 'understand', as Table 4.42 shows, a stem augment *-g(a)* appears, in the negative past, in the SG, and in the third-person.[16] Other lexemes show this exact same paradigmatic configuration with stem extensions in /d/ or /ŋ/ instead. This distribution might be ancestrally related to the vowel-/consonant-initial suffixes that are associated with the use of different stems in Eastern Kiranti (see e.g. Bantawa in Section 4.2.2.2).

[16] Note that in other lexemes these stem augments occur in the singular forms exclusively. These cases, of course, do not classify as morphomic.

Table 4.42 Paradigm of Sunwar 'understand', negative past (Borchers 2008: 200)

	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

4.2.2.13 Svan (Tuite 1995)

In the Kartvelian language Svan (Georgia), 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.

Table 4.43 Aorist tense paradigm of three Svan verbs (Tuite 1998: 12; 1994: 323)

	'extinguish'		'cut'		'wreck'	
	SG	PL	SG	PL	SG	PL
1EXCL	o-dəg	o-dig-d	o-č'k'or	o-č'k'wer-d	žoxw-žwem	žoxw-žom-e-d
1INCL	–	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

As shown in Table 4.43, the morphological instantiations of this opposition are very diverse. Some verbs (e.g. 'cut' above) mark these cells by umlauting[17] the stem vowel. Some other verbs show umlauting of the 1SG and 2SG instead (see 'wreck') as well as suffixation on 3SG+PL. Yet other verbs (e.g. 'extinguish') show more ancient vowel apophonies[18] 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 must be necessarily complex (i.e. it must involve, like Romance L, separate events or sound changes). It may boil down

[17] This started (it is no longer a synchronic phonological rule) as the anticipatory assimilation of /a/, /o/, /u/ (and possibly /ə/) to a following front high vowel, which yielded /æ/, /œ/, /y/, and /i/ respectively. Note that the form /we/ shown in Table 4.51 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/).

[18] These are the alternations known as Ablaut in Kartvelian studies. These vowel apophonies (which are reminiscent of the Proto-Indo-European 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.

ultimately 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-tense?) suffix /i/ to be erased from 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 probably 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, with 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-hand side of Table 4.52). It seems, thus, that the paradigmatic distribution of a more ancient vowel alternation (Ablaut) might have been modified to fit that of a more recent and robust one (umlaut) (see Table 4.44). This may have been facilitated by the morphological and distributional similarity of the two patterns.

Table 4.44 Converging patterns of vowel apophony in Svan

	Old alternations				Umlaut-derived metaphonies								Restructuring			
	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

The morphological variation found in different Svan varieties confirms the productivity and diachronic resilience of this 1SG/2SG vs 3SG/PL split. The morphological means to distinguish 1SG+2SG from 3SG+PL differ from one variety to another (see Table 4.45). Looking at 1SG/2SG, we see a suffix *-sgw* in Becho, a suffix

Table 4.45 The verb 'prepare' in the imperfect tense in various Svan varieties (Tuite 1995: 30)

	Becho		Laxamul		Lashx	
	SG	PL	SG	PL	SG	PL
1EXCL	xwamar-a-**sgw**	xwamar-a-d	xwamar-**Ø**	xwamar-a-d	xwamār-**is**	amār-(d)ad
1INCL	–	lamar-a-d	–	lamar-a-d	–	amār-(d)ad
2	xamar-a-**sgw**	xamar-a-d	xamar-**Ø**	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

-*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 described in this section is still productive or has been so historically.

4.2.2.14 Thulung (Lahaussois 2002)

Stem alternations in Thulung (Sino-Tibetan) are numerous and often involve the addition of segments in particular paradigm cells. The pattern displayed in Table 4.46, for example, can also be instantiated with the forms -*k* (vs -Ø), -*p* (vs -*m*), and -*ɖ* (vs -*n*). This stronger/longer stem appears, thus, in the non-past, in the 1PL.INCL, and everywhere in the past except in the 1SG and 3PL.

Table 4.46 Paradigm of Thulung 'come (up)', intransitive (from Allen 1975: 204)

	Non-past			Past		
	SG	DU	PL	SG	DU	PL
1EXCL	ge-ŋu	ge-tsuku	ge-ku	ge-ŋroro	get-tsoko[a]	get-toko
1INCL		ge-tsi	geɖ-i		get-tsi	geɖ-ɖi
2	ge-na	ge-tsi	ge-ni	geɖ-na	get-tsi	geɖ-ni
3	ge	ge-tsi	ge-mi	geɖ-ɖa	get-tsi	ge-miri

[a] The alternation between /ɖ/ and /t/ is automatic (i.e. phonological).

In the case of transitive verbs, the distribution of these alternations is slightly different. As shown in Table 4.47, the long stem appears in a superset of the contexts where it did in intransitive verbs, extending to the 1SG and 3PL present and to the whole of the past.

Table 4.47 Paradigm of Thulung 'look', transitive, 3SG patient (Lahaussois 2002: 158)

	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[a]

[a] Lahaussois mentions the existence of variation in the 3PL, in both past and present regarding the stem used in those two cells. This, however, does prevent this pattern being unavoidably morphomic.

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 (-*ḍ*-) which subsequently disappeared in many contexts. Traces of this suffix can still be found by comparing 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.PRS/3SG.PRS/1SG.PAST/2SG.PAST/3SG.PAST/DU.PAST/PL.PAST

4.2.2.15 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 show an unnatural morphological affinity. The shaded cells in Table 4.48 share a suffix (or a stem extension) not found in the rest of the paradigm. This element takes slightly different forms in the two conjugations of the language, and therefore classifies as a morphome here.

Our knowledge of the Udmurt verb's history is incomplete, but the origin of this pattern can be largely recovered. This must have occurred in 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, unmistakably constitutes an innovation of Udmurt, probably motivated by the morphological disambiguation of present and future. It has been proposed that the suffix originally denoted a frequentative meaning (see Winkler 2001: 50).[19]

Table 4.48 Verb agreement in Udmurt (Csúcs 1988: 142)

	1st conjugation, *mi̯ni̯ni̯* 'go'				2nd conjugation, *daśani̯* 'prepare'			
	Present		**Future**		**Present**		**Future**	
	SG	**PL**	**SG**	**PL**	**SG**	**PL**	**SG**	**PL**
1	mi̯ni-śko	mi̯ni-śko-m	mi̯n-o	mi̯n-o-m	daśa-śko	daśa-śko-m	daśa-lo	daśa-lo-m
2	mi̯ni-śko-d	mi̯ni-śko-di̯	mi̯n-o-d	mi̯n-o-di̯	daśa-śko-d	daśa-śko-di̯	daśa-lo-d	daśa-lo-di̯
3	mi̯n-e	mi̯n-o	mi̯n-o-z	mi̯n-o-zi̯	daśa	daśa-lo	daśa-lo-z	daśa-lo-zi̯

[19] Note the similarity to the evolution of the inchoative suffix (with the same form -*sk*-) from Latin to some modern-day Romance languages (Meul 2010). There is, in a different part of the Udmurt paradigm, yet 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'l'a*- that is also a frequentative marker in the language (see Winkler 2001: 50).

A second step would have involved the emergence of the second conjugation from the first. It is usually assumed (see e.g. Frodl 2013: 21–2) that the /l/ which characterizes this verb class was originally part of the stem and appeared throughout the whole paradigm. Sound change would have then deleted the consonant in coda positions (e.g. 1SG.PRS **daśal-śko* > *daśa-śko*, 3SG.PRES *daśal* > *daśa*) while leaving intervocalic /l/ in place (e.g. in 3PL.PRS *daśal-o*).

4.2.3 Europe

4.2.3.1 Aragonese (Haensch 1958; Saura Rami 2003; Barcos 2007)

Local varieties of Aragonese differ as for the synchronic distribution in the verbal paradigm of the reflexes of the N-morphome (e.g. diphthongization). The most conservative of them (see Table 4.49) have those stem alternants in the cells where the alternation emerged in the first place. In these varieties, the N-morphome appears, as expected, in those cells that were rhizotonic in Latin, i.e. in the SG and the 3PL of the present-tense in both indicative and subjunctive, and in the 2SG imperative. These cells, in fact, continue to have stress on the root in varieties like the one of Ansotano.

Table 4.49 Ansotano Aragonese 'have', present (Barcos 2007)

	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

This paradigmatic configuration of diphthongization (i.e. /je/ vs /e/ as in Table 4.49, or /we/ vs /o/ in other verbs) is stable across verbs even in the presence of another stem alternation, the L-morphome, 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 morphome. In Alta Ribagorza Aragonese (see Table 4.50), diphthongization has preserved its inherited distribution in those verbs where only the N-morphome is found (e.g. in 'sleep' in Table 4.58), but has innovated a different distribution in those verbs where the L-morphome also occurs (e.g. in 'twist'). Notice that, in the latter verb, diphthongization has extended to the 1PL and 2PL cells of the present subjunctive.

These morphological changes in the paradigmatic configuration of the N-morphome must therefore be the result of interaction/interference with the L-morphome. 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 morphomes a subset of the other (see Herce 2019a).

Table 4.50 Two Alta Ribagorza Aragonese verbs (Haensch 1958)

	'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

To complete the picture of the N-morphome-related variation in the language, it must be mentioned that, in the most innovative varieties of Aragonese, the domain of the N-morphome has changed in all verbs, even in those without an overt L-morphome (see Table 4.51). Due to these changes in the 1/2PL.SBJV, the diphthongizations typical of the N-morphome no longer correlate to rhizotony in these varieties.

Table 4.51 Benasque Aragonese 'sleep', present (Saura Rami 2003)

	Indicative		**Subjunctive**	
	SG	PL	SG	PL
1	dwérmo	dormém	dwérme	dwermám
2	dwérmes	dorméts	dwérmas	dwermáts
3	dwérme	dwérmen	dwérma	dwérman

The pattern of diphthongization of Alta Ribagorza Aragonese 'twist' in Table 4.50 has been the only one included in the morphome database. Those alternations that have the same paradigmatic distribution as the L- or N-morphomes in Spanish have not been included in the database due to their cognacy with these.

4.2.3.2 Basque (personal knowledge)

The verbal inflection of Basque is mainly agglutinative, and relies for the vast majority of verbs on the use of auxiliaries that bear the A, S, and P 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 paradigms in Table 4.52. 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. The presence of these morphs in the 2SG has a straightforwardly diachronic explanation. 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

Table 4.52 Partial paradigms of three Basque verbs

	etorri 'come', present		*ibili* 'walk', past		*egon* '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

polite reference to a 2SG addressee. The earlier 2SG forms (e.g. *ha-tor* 'come.2SG') thus became reserved for familiar address. Unlike in English or French, however, a new 2PL pronoun and a new 2PL verbal form were innovated by adding pluralizers (*-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 like plural forms.

4.2.3.3 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 out here. The first is found in the paradigm of the English verb *be* (see Table 4.53), which shows an unnatural syncretism not found elsewhere in the language but systematic in that verb as defined here because it is repeated with different exponents.

Table 4.53 Paradigm of the English verb 'be'

	Present		Past	
	SG	PL	SG	PL
1	am	are	was	were
2	are	are	were	were
3	is	are	was	were

As is well known, the presence of the form *are* in the 2SG of *be* is due to the use of an earlier 2PL form (*you*) for the 2SG in the modern language.[20] Such a change was driven by the common strategy (see also Basque in Section 4.2.3.2) of signalling politeness by referring to singular addressees with a plural pronoun.

The second pattern that classifies as morphomic in English is found in three verbs which have a longer stem in the 1SG, 2SG, and PL of the present compared with other cells (see Table 4.54). The emergence of this particular pattern is related to the fact that those cells are the ones in which the verb stem is not followed by a

[20] 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.

suffix. This different phonological context has made it possible for sound changes to apply differently in different cells. These verbs (also 'be' before) are among the most frequent in the language, which has also made it possible for them to preserve these structures even in an ocean of invariance.

Table 4.54 Finite forms of three English verbs

	'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

English1: 2SG/PL
English2: 1SG/2SG/PL

4.2.3.4 French (Meul 2010; Esher 2015)

In French inflection, verbs vary in the extent to which they show what could be considered their 'full stem' throughout the paradigm. Consider the paradigm in Table 4.55. The stem /ʁezolv/ (vs /ʁezu/) appears in the plural forms of the present indicative and in all forms of the present subjunctive and the imperfect. As explained before (Table 4.55), the same situation obtains with other segments (/z/, /n/, /s/, /ɲ/, /j/, /v/) 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).

Table 4.55 Paradigm of French *résoudre* 'solve'

	PRS.IND		PRS.SBJV		IPF		FUT	
	SG	PL	SG	PL	SG	PL	SG	PL
1	ʁezu	ʁezolvɔ̃	ʁezolv	ʁezolvjɔ̃	ʁezolvɛ	ʁezolvjɔ̃	ʁezudʁe	ʁezudʁɔ̃
2	ʁezu	ʁezolve	ʁezolv	ʁezolvje	ʁezolvɛ	ʁezolvje	ʁezudʁe	ʁezudʁe
3	ʁezu	ʁezolv	ʁezolv	ʁezolv	ʁezolvɛ	ʁezolvɛ	ʁezudʁe	ʁezudʁɔ̃

4.2.3.5 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 paradigms in Table 4.56. This affix appears usually as /e/, which must have been originally its

Table 4.56 Aorist past-tense paradigm of two Greek verbs (Holton et al. 2012)

	'tie'		'know'	
	SG	PL	SG	PL
1	ˈe-desa	ˈdesame	ˈi-xera	ˈxerame
2	ˈe-deses	ˈdesate	ˈi-xeres	ˈxerate
3	ˈe-dese	ˈe-desan	ˈi-xere	ˈi-xeran

Note: The other past-tense, the imperfect, shows the same pattern.

only form. In just a few verbs, it has nowadays the form /i/ instead. In Ancient Greek (and also in other older Indo-European 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 however (e.g. in the case of 'know', 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 due to 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-morphome 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.

4.2.3.6 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 at the earliest stages were more or less correlated with semantic distinctions, but later sound changes and analogical changes have meant that parts of the paradigm share a form despite a lack of semantic or morphosyntactic common thread.

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 shared by these cells can vary. In verbs without stem alternation, no particular morphological affinity will be apparent between the shaded cells. In other cases (see *sjá* 'see' in Table 4.57, also *sjóða* 'boil', *auka* 'enlarge', and many others), only one segment /j/, or the stem vowel, or a diphthong is shared by the shaded cells. In yet other cases (see *eiga* 'own' in Table 4.2), the whole of the stem is exclusive to the mentioned paradigm cells.

Table 4.57 Paradigm of *sjá* 'see' in Icelandic (Jörg 1989)

	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

The diachrony of the Germanic verb is mostly well understood. The shaded forms in Table 4.57 derive from the Proto-Indo-European e-grade, which was found across the present. Due to later sound changes in Germanic (see Table 3.2), some of the singular PRS.IND cells (all of them in North Germanic) developed a different stem vowel. The rest of the former e-grade cells where therefore left behind as an unnatural class.

4.2.3.7 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 nouns in Table 4.58. As they illustrate, the forms involved may be diverse: palatalization of the last consonant of the stem (/bˠɑ:d̪ˠ/ vs /bˠɑ:dʲ/ 'boat'), sometimes along with a different stem vowel (/mˠɑk/ vs /mʲɪc/ 'son'), suffixation (/bʲeːsˠ/ vs /bʲeːsˠə/ 'habit'), and even suppletion occasionally (/bʲanˠ/ vs /mˠnˠɑː/ 'woman'). This morphological affinity is a very old Indo-European trait that goes back to suffixal identities that are

Table 4.58 Declension of some Irish nouns (Doyle 2001)

	'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

Note: 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.

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').

4.2.3.8 Italian and Servigliano (Camilli 1929; Maiden and Robustelli 2014)

Italian verbal inflection, as that of other conservative Romance varieties, is characterized by morphomic stem alternation patterns. Two Italian morphomes have been included in the present database, the Italian versions of the morphomes referred to as the U-morphome and PYTA in the Romance morphome literature.

Consider the former in Table 4.59. As is well known, these stem alternations originated as a result of the palatalization of some consonants before front vowels and yods (e.g. in *dire*, an older 2SG.IND *dī[k]is* > *dī[t͡ʃ]is*). The alternations are completely morphological in the modern language, and appear with a different paradigmatic configuration in close varieties.

Table 4.59 Present-tense paradigms of three Italian verbs (Maiden and Robustelli 2014)

	cogliere 'pick'		*dire* 'say'		*apparire* 'appear'	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	colgo	colga	di[k]o	di[k]a	appaio	appaia
2SG	cogli	colga	di[t͡ʃ]i	di[k]a	appari	appaia
3SG	coglie	colga	di[t͡ʃ]e	di[k]a	appare	appaia
1PL	cogliamo	cogliamo	di[t͡ʃ]amo	di[t͡ʃ]amo	appariamo	appariamo
2PL	cogliete	cogliate	dite	di[t͡ʃ]ate	apparite	appariate
3PL	colgono	colgano	di[k]ono	di[k]ano	appaiono	appaiano

In the variety of Italian spoken around Servigliano (see Table 4.60), the inherited morphomic distribution has been modified by subsequent morphosyntactically driven analogical changes in the language (e.g. the loss of the IND/SBJV distinction in 1 and 2). Because of the different extension of the L/U-morphome in this variety, this has been included as a separate one in this database.

Another morphological quirk of many Italian verbs is the presence, in three cells of the preterite tense (see Table 4.61), of a special stem not present in the rest of the paradigm. This alternation emerged from a semantically motivated one: a perfective stem opposed in Latin to an imperfective one. Those roots would have been associated originally with whole tenses, and still are in some contemporary Romance varieties like Portuguese. Italian, however, lost these roots in those cells which were arrhizotonic. The result is a person–number morphome that, like previous ones, is morphologically diverse (e.g. *fec-i fac-esti*, *conobb-i conosc-esti* 'know', *apparv-i appar-isti* 'appear', *nacqu-i nasc-esti* 'be born').

Table 4.60 Present-tense paradigms of three Servigliano Italian verbs (Camilli 1929)

	pote 'can'		*di* 'say'		*ae* 'have'	
	IND	SBJV	IND	SBJV	IND	SBJV
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 4.61 Past-tense paradigms of three Italian verbs (Maiden and Robustelli 2014)

	fare 'do'		*cuocere* 'cook'		*rompere* '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

Italian1: 1SG.IND/3PL.IND/SG.SUBJ/3PL.SUBJ
Servigliano: 1SG/3.SUBJ
Italian3: 1SG/3

4.2.3.9 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 some suffixes (see Table 3.2) raised the stem of many verbs, creating a pattern of stem alternation where the 2SG and 3SG cells are opposed to the rest (i.e. 1SG+PL). Other unrelated sound changes, e.g. closed-syllable shortening, other types of umlaut (see Albright 2010), gave (or would have given) rise to different patterns of stem vowel alternation. These, however, have often been made to conform to the (morphomic) pattern of stem alternation presented in Table 4.62.

Table 4.62 Three Luxembourgish verbs, present-tense (Schanen 2004)

	sinn 'be'		*kommen* 'come'		*maachen* '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

As explained in Tables 3.32 and 3.33 and the ensuing discussion, the morphological affinities displayed in Table 4.62 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 non-alternating 'make', etc. show that the morphological identity of the 1SG+PL present has been acting as a template for the distribution of allomorphy in the paradigm.

4.2.3.10 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 partial paradigms in Table 4.63. The syncretism in 'house' and other polysyllabic stems (i.e. with the formative /in/) 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 on the basis of the large class of polysyllabic nouns where the two cells were syncretic initially.

Table 4.63 Two partial paradigms in East Finnmark North Saami (Hansson 2007: 25, 28)

	'house'		'who'	
	SG	PL	SG	PL
LOC	viesu-s	viesu-in	gea-s	gea-inna
COM	viesu-in	viesu-iguin	gea-inna	gea-iguin

It is worth noting that the /i/s in these two *-in* suffixes could potentially lend themselves to different segmentations. One may feel justified in segmenting 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) 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 homophony, the two cells have led parallel lives in North Saami, and also elsewhere.

As illustrated in Table 4.64 (see also Table 4.66), in varieties with stem alternation, the stem in the two cells is usually the same, even when this means (as in COM.SG *kūs'k*-) deviating from a more natural distribution. Syncretism is thus maintained even in the presence of various non-linear inflectional operations (i.e. consonant gradation or vowel apophonies) that might have disrupted it, which is suggestive of a systematic morphological identity.

Table 4.64 Partial paradigm of Kildin Saami *kuess'k* 'aunt' (Rießler 2022)

	SG	PL
ILL	kuasska	kūs'ket'
LOC	kuas'kes't	**kūs'ken'**
COM	**kūs'ken'**	kūs'keguejm
ABE	kues'kxa	kūs'kexa

4.2.3.11 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 associated to particular phonological environments, 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 morphome status.

As Table 4.65 illustrates, the strong grade[21] of the stem, and also a different stem vowel (/wa/ [vs /o/] and /ɛ/ [vs /e/]) appear in nominative and illative singular, and in the essive, whose singular and plural forms are the same.

Table 4.65 Two nominal paradigms of Pite Saami (Wilbur 2014: 96, 101)

	luakkta 'bay'		*bärrgo* 'meat'	
	SG	PL	SG	PL
NOM	luakkta	luokta	bärrgo	biergo
GEN	luokta	luoktaj	biergo	biergoj
ACC	luoktav	luoktajd	biergov	biergojd
ILL	luakktaj	luoktajda	bärrgoj	biergojda
INESS	luoktan	luoktajn	biergon	biergojn
ELAT	luoktast	luoktajst	biergost	biergojst
COM	luoktajn	luoktaj	biergojn	biergoj
ABESS	luoktadak	luoktadaga	biergodak	biergodahta
ESS	luakktan	luakktan	bärrgon	bärrgon

Nominal declension can also show a different morphological pattern that involves vowel apophonies different from the ones that participated in the former alternation. In this case, we are dealing with vowel raisings which include the following: /e/>/i/, /o/>/u/, /a:/>/ɛ/, /a:/>/i/, /ɔ/>/u/, /a/>/ɛ/, /a/>/e/, and /a/>/i/.

[21] Strong grade in Pite Saami usually involves gemination with respect to the weak grade (as in Table 4.71) but can also involve adding a segment /t/, /p/ or /k/ (e.g. /va:jmo/ 'heart.NOM.PL' vs /va:jpmo/ 'heart.NOM.SG').

Consider the paradigms in Table 4.66. In the inflection of *guolle* and *vágge*, a high-vowel stem appears in various cases in the plural and 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 what the paradigms in Table 4.66 might suggest, it is not possible synchronically to identify a phonological context where these forms occur, nor to consistently derive one vowel from the other (Wilbur 2014: 79).

Table 4.66 Two nominal paradigms of Pite Saami (Wilbur 2014: 101)

	guolle 'fish'		*vágge* 'valley'	
	SG	PL	SG	PL
NOM	guolle	guole	vágge	vágge
GEN	guole	gulij	vágge	väggij
ACC	guolev	gulijd	vággev	väggijd
ILL	guolláj	gulijda	vággáj	väggijda
INESS	guolen	gulijn	vággen	väggijn
ELAT	guolest	gulijst	vággest	väggijst
COM	gulijn	gulij	väggijn	väggij
ABESS	guoledak	guoledaga	vággedak	vággedaga
ESS	guollen	guollen	vággen	vággen

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 paradigm in Table 4.67.

Table 4.67 Pite Saami *viessot* 'live' (Wilbur 2014: 172)

	PRS			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

As in nouns, the strong grade also may occur along with stem vowel apophony (/wa/ [vs /o/] and /ɛ/ [vs /e/]). The one shown in Table 4.73 is the distribution of the strong grade in all Pite Saami verbs that show gradation. Vowel raising, however, shows a different picture, as there are two classes of verbs according to where raising occurs in the paradigm.

In the first of these classes (Table 4.68), vowel raising applies to 1DU.PRS, 3PL.PRS, 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

Table 4.68 Pite Saami *bassat* 'wash' (Wilbur 2014: 174)

	PRS			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

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 morphome interactions (Herce 2019a), but does not extend to the other verbal class (Table 4.69). Here, vowel raising applies to a superset of the cells where it applies 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 Saami, see Kahn and 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*.

Table 4.69 Pite Saami *basset* 'fry' (Wilbur 2014: 174)

	PRS			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

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.PRS/DU.PRS/PL.PRS/1SG.PAST/2SG.PAST/3PL.PAST
Pite Saami4: 1DU.PRS/3PL.PRS/1SG.PAST/2SG.PAST/3PL.PAST
Pite Saami5: 1DU.PRS/3PL.PRS/PAST

4.2.3.12 Skolt Saami (Feist 2015)

Skolt Saami's stem alternations are the same as those in Pite Saami. In the verb, however, there are a few relevant differences. One is the loss of the dual. Since a 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. /pː/ vs /p/, /tː/ vs /t/) have become qualitative (e.g. /pː/ vs /v/, /tː/ vs /ð/) in Skolt Saami.

In the paradigm in Table 4.70, the weak grade (/ɣ/) appears in 1SG and 2SG present and in 3SG, 1PL, and 2PL past. The strong grade (/gː/) appears in the rest of

Table 4.70 Inflectional paradigm of *njorggad* 'whistle' (Feist 2015: 204, 210)

	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-ik̆	njoor[ɣ]-id
3	njorgg	njorgg-a	njoor[ɣ]-i	njurgg-e

the paradigm. The paradigmatic distribution of the two forms is, therefore, unnatural. In addition to this, as Table 4.71 shows, the paradigmatic distribution of vowel raising is different in Skolt and Pite Saami: in Skolt Saami, it appears exclusively in the past. In some (e.g. *njorggad*), vowel raising appears in 1SG, 2SG, and 3PL of that tense. This is morphosyntactically unnatural, and contrasts with the distribution of raising in Pite Saami (see Table 4.68), where it also occurred in two cells in the present.

Table 4.71 Inflectional paradigm of *njorggad* 'whistle' (Feist 2015: 204, 210)

	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-ik̆	njoor[ɣ]-id
3	njorgg	njorgg-a	njoor[ɣ]-i	njurgg-e

In other Skolt Saami verbs, in the same way as in Pite Saami (see Table 4.69), raising extends to all the past cells (Feist 2015: 209). Due to its coextensiveness with the value 'past', this alternation has become semantically motivated in this class of verbs and does not classify as morphomic here. It does constitute an interesting example, however, of a morphomic stem alternation pattern becoming morphemic (see also Section 3.2.4.1).

Skolt Saami1: 1SG.PRS/2SG.PRS/3SG.PAST/1PL.PAST/2PL.PAST
Skolt Saami2: 3SG.PRS/PL.PRS/1SG.PAST/2SG.PAST/3PL.PAST
Skolt Saami3: 1SG/2SG/3PL

4.2.3.13 Spanish and Asturian (personal knowledge; Bybee 1985)

Romance languages are well known for being the family where morphomic stem alternation patterns have been most thoroughly studied (see e.g. Maiden 2018b). Spanish will be taken here as a representative of two of the most frequently discussed ones: the N-morphome and the L-morphome. The former is illustrated by the paradigm in Table 4.72. A diphthong (i.e. /je/ vs /e/) appears in *perder*, in the present, in the singular and 3PL cells. In other verbs (e.g. *poder* 'be able to'), the alternations /we/ vs /o/, and /we/ vs /u/ (in *jugar* 'play') have the 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 (and see also Aragonese in Section 4.2.3.1, where the domains of stress and dipthongization do not always coincide.).

Table 4.72 Present-tense paradigm of Spanish *perder* 'lose'

	Present indicative		Present subjunctive	
	SG	PL	SG	PL
1	pierdo	perdemos	pierda	perdamos
2	pierdes	perdéis	pierdas	perdáis
3	pierde	pierden	pierda	pierdan

Consider now the L-morphome in Table 4.73. As the paradigm of *caer* 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 (*[k]ep-o* vs *[k]ab-es* 'fit') the alternation is weakly suppletive.

Table 4.73 Present-tense of Spanish *caer* 'fall'

	Present indicative		Present subjunctive	
	SG	PL	SG	PL
1	caigo	caemos	caiga	caigamos
2	caes	caéis	caigas	caigáis
3	cae	caen	caiga	caigan

Other Romance varieties closely related to Spanish have similar paradigmatic alternations. An interesting one, cognate with the one in Table 4.72 but with a different paradigmatic configuration, is present in western Asturian (see Table 4.74). Diphthongization occurs in this variety, in some 35 verbs (see e.g. *murder*), in the 2SG, 3SG, and 3PL of the present indicative. Some of these (e.g. *ferber* 'boil') have another diphthong (i.e. /je/) with the same paradigmatic distribution, which makes this pattern morphomic as defined here. The diachronic origin of this alternants is to be found in the interaction between the two morphomes that have been described for Spanish in this section. The shaded cells in Table 4.74 are those that participate in the N-morphome allomorphy but not in the L-morphome one (see Herce 2019a for more details).

Spanish1: SG/3PL
Spanish2: 1SG.IND/SBJV
Asturian: 2SG/3SG/3PL

Table 4.74 Present-tense paradigm of *murder* 'bite' in western Asturias (Bybee 1985: 73)

	Present indicative		Present subjunctive	
	SG	PL	SG	PL
1	ˈmordo	murˈdemos	ˈmorda	ˈmordamos
2	ˈmwerdes	murˈdeis	ˈmordas	ˈmordais
3	ˈmwerde	ˈmwerden	ˈmorda	ˈmordan

4.2.4 Australasia

4.2.4.1 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.

The pattern of syncretism in Table 4.75 (1SG+PL) is found in various different suffixes, although the actual alternating segments 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 a zero morph. 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.[22] 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.

Table 4.75 Allomorphy of some Barai suffixes (Olson 1973: 48, 53,56)

	Past sequence 1		Past sequence 2[a]		Future sequence		Delayed past sequence	
	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

[a] The form /jo/ is found in verbs that end in a stressed front vowel and /βo/ is found elsewhere. Note that orthographic 'v' represents /β/.

[22] It is interesting to note that, in related Koiari (see Section 4.2.4.8), /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 therefore seems interesting, but is unclear to me at the moment.

4.2.4.2 Benabena (Young 1964)

For the purposes of stem alternation, SG+1 subject values constitute a morphological class in Benabena (Trans-New Guinea) and, to a lesser extent, in related Gorokan languages. Similar morphological alternations to those in Table 4.76 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. This and other formatives are most likely grammaticalized from verbs. Other forms of the verb in Benabena are often based on those in Table 4.76, except the future tense, which does not show the morphomic affinities described here and shows, for the verbs 'hit' and 'go', the stems *ha-* and *bi-* respectively.

Table 4.76 Two verbs in Benabena, past-tense (Young 1964: 50)

	'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

4.2.4.3 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.

Table 4.77 Biak verb 'die' (van den Heuvel 2006: 157)

	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

Consider the paradigm in Table 4.77. 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 in that, at the end of an intonational unit, they require an epenthetic vowel, as illustrated in Table 4.78.

As discussed by van den Heuvel (2006: 66), all those forms in *k(o)-* can be traced back to Proto-Austronesian **telu* 'three' (*/t/>/k/ is regular in Biak). This etymology, along with the comparison to closely related languages (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, the use of these forms spread to denote

Table 4.78 Biak verb 'eat' (van den Heuvel 2006: 159)

	SG	DU	PC/TR[a]	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

[a] Van den Heuvel is not consistent in the glossing of the forms. Sometimes he labels them 'trial' and at other times 'paucal'. It is thus not clear to me what the precise value is of the forms. This, however, does not affect the present analysis.

larger numbers too. The result is a morphological affinity of the shaded cells that is no longer semantically justified.

4.2.4.4 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.

Consider the agreement prefixes in Table 4.79. Excluding the prefixes *s-* and *t-*, for which a coherent meaning (animate plural) can indeed be identified, the distribution of the other prefixes does seem not make much sense morphosyntactically. Depending on the noun, all the prefixes may co-occur with the plural but not the singular, with the singular but not the plural, with both singular and plural, and with neither number value. It may also be 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 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 the two agreement classes appears to be completely arbitrary; however, because of the existence of two conjugations, we can see that these morphomic classes are systematic.

Table 4.79 Genders and conjugations in Burmeso (Donohue 2001: 100–102)

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-

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, is reminiscent of many others, such as those found in Khinalugh (see Section 4.2.2.8), Mian (see Section 4.2.4.12) 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 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 might have originated from a more unremarkable two- or three-gender system that somehow 'went wrong' when lexeme-number orthogonality of some nouns (e.g. *singularia and pluralia tantum*) was compromised.

Burmeso1: II.SG/III.SG/V.PL/VI
Burmeso2: I.SG/III.PL/IV/V.SG

4.2.4.5 Ekari (Drabbe 1952; Doble 1987)

In the Ekari (Trans-New Guinea) language, future tense suffixes display an allomorphic variation whose paradigmatic distribution is morphosyntactically unnatural (see Table 4.80). Allomorphic variation satisfies the criteria set for morphomicity here.

Table 4.80 Partial paradigm of 'go' (Drabbe 1952: 49–50; Doble 1987: 89)

	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

Because the languages most closely related to Ekari are not sufficiently described, it is difficult to make educated guesses about the diachronic origin of these alternations. As Table 4.80 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 an origin related to sound change.

Ekari1: 2SG/3SG.M/1DU/1PL
Ekari2: 1SG/2DU/2PL/3DU/3PL/3SG.F

4.2.4.6 Girawa (Gasaway and 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

Table 4.81 Partial paradigms of two Girawa verbs (Gasaway and Sims 1977)

	'eat' present			'hit' present, 2SG subject[a]		
	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

[a] 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 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 and Sims (1977) to confirm that 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

and in (subject and object) agreement suffixes. As Table 4.81 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 edge 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' in Table 4.81 do they reach (near-)suppletion (Gasaway and Sims 1977: 30).

Object suffixes, when they occur, immediately follow the verb stem. Their overall form (*-i* vs *-wa* vs -Ø) agrees with the morphomic patterns of stem alternation, which makes it plausible to argue that the different phonological profiles of these suffixes may have been responsible for the emergence of stem alternations. This receives support from comparative evidence from other Madang languages (see e.g. Amele in Table 4.82) which seem to lack stem alternations but do have object agreement suffixes with the same pattern. 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/) exists in Amele between 1/2SG and 1PL subject suffixes as well.[23]

Although the diachronic details are uncertain, it seems 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, which might have contributed to the emergence of 1/2SG identity

[23] This pattern is also found in Kosena (Section 4.2.4.9), and a very similar one is found in Yagaria (see Section 4.2.4.22). Both are Trans-New Guinea 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.

in other domains, for example facilitating the spread of the suffix /m/ to the 1DU in Girawa (compare Tables 4.81 and 4.82).

Table 4.82 Partial paradigms of two Amele verbs (Roberts 1987: 279)

	'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

4.2.4.7 Kele (Ross 2001)

In some Oceanic languages like Kele (see also Vurës in Section 4.2.4.18), nouns in their possessive inflection are subject to stem alternations. As the paradigms in Table 4.83 illustrate, 3SG and all the non-singular 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 stems 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 alignment.

Table 4.83 Possessor paradigm of two Kele nouns (Ross 2001: 133)

	dop '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

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).

4.2.4.8 Koiari (Dutton 1996; 2003)

Koiari (Koiarian, New Guinea) tense–aspect suffixes sometimes have a different form depending on the person and number of the subject. Frequently, only two forms are distinguished, whose paradigmatic distribution does not correlate with any value. As illustrated in Table 4.84, one allomorph appears in 2SG+PL, an unnatural class.

Table 4.84 Some Koiari TAM morphology (Dutton 1996: 23, Dutton 2003: 346, 351)

	'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

In Koita, the closest relative to Koiari, some of these forms (e.g. imperfect *-ima* vs *-a*, see Dutton 1975: 338) correspond to a natural SG vs PL distinction. It is unclear to me how the Koiari system may have emerged. In other languages where we find a PL+2SG morphomic pattern (e.g. Basque and English), this emerged when a plural pronoun started to be used to refer politely to a SG addressee. In Koiari, however, this does not seem to have happened, since Koita and Koiari have identical pronouns with the same values (2SG *a* vs 2PL *ya*). The history of this pattern is therefore unclear.

4.2.4.9 Kosena (Marks 1974)

In the grammar of Kosena (Trans-New Guinea, also known as Awiyaana) and in related Usarufa, there are various and complex morphophonological rules operating across morpheme boundaries. Paradigms often show unmotivated morphological allegiances. For example, the 1SG and the 1PL are usually syncretic to the exclusion of 1DU, and this syncretism often extends to the 2SG.

This morphological affinity in mood suffixes (see Table 4.85) is similar to the one found in related languages like Yagaria (see Section 4.2.4.22) and most probably shares an identical diachronic origin. Consider also the similarity of this pattern to the one in Amele (Table 4.82). Overall, it seems like an inconsequential morphological affinity (a shared /n/) of 2SG and 1PL caused these values to partake in the same sound changes, thus giving rise to the 2SG/1PL morphomic affinity we observe in Yagaria. Later changes, however, seem to have progressively extended the domain of this morphome (> 2SG/1PL/1SG in Kosena, >2SG/1PL/1SG/1DU in Girawa) in related Trans-New Guinea languages.

Table 4.85 Paradigm of various inflectional suffixes in Kosena (Marks 1974)

	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

4.2.4.10 Maranunggu (Tryon 1970)

In Maranunggu (Western Daly) and its closest related languages, Manda and Ami (Tryon 1974), several verb classes exist (18 in Maranggu), often associated with certain semantic correlates. Verbs of the same class (e.g. *wat* 'walk', *kalkal* 'climb', *tratrayme* 'look for', *tyapat* 'swim', *witlyuk* 'enter', *wurka* 'work', and *tat* 'rest' belong to class I) are characterized by requiring the same auxiliary verb. This is the locus for the expression of person–number agreement, and future vs non-future tense.

Table 4.86 shows the forms of this auxiliary in classes I and XII. The segmentation of the forms of these auxiliaries is extremely challenging, and irregularities abound. However, the forms that the auxiliary takes in 1PL.EXCL non-future and 2 future are very closely associated in Maranunggu, by virtue of having shared formatives in many verb classes, as well as by a high degree of mutual formal predictability. As in the shaded forms in Table 4.86, suffixing *-n* to the 2SG future very often derives the 1PL non-future, and suffixing *-ra* to that same form derives the 2PL future.

Table 4.86 Paradigms of Class I and Class XII auxiliaries in Maranunggu (Tryon 1970: 18, 23)

<table>
<tr><td></td><td colspan="4">Class I</td><td colspan="4">Class XII</td></tr>
<tr><td></td><td colspan="2">Non-future</td><td colspan="2">Future</td><td colspan="2">Non-future</td><td colspan="2">Future</td></tr>
<tr><td></td><td>SG</td><td>PL</td><td>SG</td><td>PL</td><td>SG</td><td>PL</td><td>SG</td><td>PL</td></tr>
<tr><td>1EXCL</td><td rowspan="2">kangani</td><td>warin</td><td rowspan="2">ngawani</td><td>ngarani</td><td rowspan="2">kangatan</td><td>yetan</td><td rowspan="2">ngawatan</td><td>ngaratan</td></tr>
<tr><td>1INCL</td><td>karrkani</td><td>ngarrkani</td><td>karrkatan</td><td>ngarrkatan</td></tr>
<tr><td>2</td><td>kanani</td><td>karani</td><td>wari</td><td>warira</td><td>kanatan</td><td>karatan</td><td>yeta</td><td>yetara</td></tr>
<tr><td>3</td><td>kana</td><td>kuninya</td><td>kawani</td><td>purani</td><td>katan</td><td>kutinya</td><td>kawatan</td><td>puratan</td></tr>
</table>

The diachronic emergence of this idiosyncratic morphological affinity is difficult to recover because in Maranunggu, Manda, and Ami (which form one of the two major branches of Western Daly) this configuration is already in place, whereas in the rest of the Western Daly languages it is nowhere to be found. It might be relevant to point out, however, that sometimes the morphology of 1PL.NF+2.FUT from Maranunggu and its closest relatives seems cognate with that in 2SG.FUT in other Western Daly languages. Marithiyel, for example, has the prefix *wari-* with this value (Tryon 1974: 79) in a class of verbs described as containing mostly motion verbs, the same as Maranunggu Class I. As should also be apparent from the forms in Table 4.86, the 2SG.FUT is usually the shortest one, so this morphomic affinity may be ultimately due to its unprefixed status compared to other forms (i.e. 1SG.FUT **ngawarini* > *ngawani*).

Another morphological affinity found in many Maranunggu auxiliaries is that the morphology common to the cells in Table 4.86 often extends to all future singular cells. In Class 1, for example, the form *wa* appeared in 1PL.EXCL.NF, SG.FUT, and 2PL.FUT. Similar paradigmatic affinities can be found in other auxiliaries (see

Table 4.87). This set of cells classifies, as a second, different morphome in Maranunggu, since, the same as the previous one, this morphological affinity is also instantiated with different forms in different paradigms.

Table 4.87 Paradigms of the auxiliaries of Class IV and Class XIV (Tryon 1970: 18, 23)

<table>
<tr><th></th><th colspan="4">Class IV</th><th colspan="4">Class XIV</th></tr>
<tr><th></th><th colspan="2">Non-future</th><th colspan="2">Future</th><th colspan="2">Non-future</th><th colspan="2">Future</th></tr>
<tr><th></th><th>SG</th><th>PL</th><th>SG</th><th>PL</th><th>SG</th><th>PL</th><th>SG</th><th>PL</th></tr>
<tr><td>1EXCL</td><td rowspan="2">kangiya</td><td>yun</td><td rowspan="2">ngayu</td><td>ngiriya</td><td rowspan="2">kengi</td><td>yen</td><td rowspan="2">ngiye</td><td>ngeri</td></tr>
<tr><td>1INCL</td><td>kirrkiya</td><td>ngirrkiya</td><td>kirrki</td><td>ngerrki</td></tr>
<tr><td>2</td><td>kaniya</td><td>kariya</td><td>yungu</td><td>yungura</td><td>keni</td><td>keri</td><td>ye</td><td>yeri</td></tr>
<tr><td>3</td><td>kaya</td><td>kuyinya</td><td>kayu</td><td>piriya</td><td>kanga</td><td>kinya</td><td>kiye</td><td>piri</td></tr>
</table>

Maranunggu1: 1PL.NF, 2F
Maranunggu2: 1PL.NF, SG.F, 2PL.F

4.2.4.11 Menggwa Dla (de Sousa 2006)

In Menggwa Dla (Senagi, Indonesia), also known as Dera, a few verbs display a stem alternation pattern that is phonologically and morphosyntactically unmotivated. As Table 4.88 illustrates, the 3SG and 2/3PL.M cells show a stem alternant different from that found in the rest of the paradigm.

Table 4.88 Menggwa Dla 'stand' past (de Sousa 2006: 541)

	SG	DU	PL
1	numb-ahahwa	numb-ehyahwa	numb-efahwa
2M	numb-afahwa	numb-afahwa	**nuŋg**-umahwa
2F	numb-afahwa	numb-efyahwa	numb-eihwa
3M	**nuŋg**-uhwa	numb-afahwa	**nuŋg**-umahwa
3F	**nuŋg**-wahwa	numb-efyahwa	numb-eihwa

Notice that these cells are also characterized by suffixes which begin with a high back vowel. Although this differential phonological environment (i.e. front vs back vowel) may have been the origin of this pattern, the alternation is not phonologically derivable synchronically, because /g/ and /b/ are fully fledged phonemes that can both appear in all phonological environments (cf. ***yaŋgifi*** /jagiɸi/ [jaŋgiβi] 'wake (someone) up', ***ambuha*** /abuxa/ [ʔambuɣa] 'cockatoo').

The pattern is clearly morphological in nature and also systematic, since the forms involved can also be suppletive. As the paradigm in Table 4.89 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'

Table 4.89 Menggwa Dla 'think/call' present (de Sousa 2006: 541)

	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

suffixes in Table 4.88. Other suppletive alternations with this same distribution include *eh-* (vs *s-* 'talk') and *ap-* (vs *e-* 'sleep').

4.2.4.12 Mian (Fedden 2011)

Gender agreement in Mian (Trans-New-Guinea) is similar to that in other languages already presented here like Khinalugh and Burmeso. The same agreement affixes are required by a class of nouns in the singular, by another in the plural, and by another in both singular and plural. Feminine singulars, neuter1 plurals, and neuter2 nouns behave all as a single unit in terms of agreement. As Table 4.90 shows, the agreement formatives take on a different form in different grammatical roles, so this pattern is systematic.

Table 4.90 Gender–number agreement affixes in Mian (Fedden 2011: 163)

	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

Although what we know about the history of the language is not enough, there are plausible ways in which these systems can emerge diachronically. In a typological parallel Fedden (2011: 168–9) 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.

4.2.4.13 Murrinh-Patha (Mansfield n.d.; Walsh 1976; Nordlinger 2015)

Murrinh-Patha (Southern Daly, Australia) verbal inflection is extraordinarily complex. For one thing, 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 value immediately lower to the value they actually express (see Table 4.91). 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.

Table 4.91 Perfect paradigm of 'sit' (Walsh 1976: 327)

	1EXCL[a]		2		3	
	Sibling	**Non-sibling**	**Sibling**	**Non-sibling**	**Sibling**	**Non-sibling**
SG	ŋem		t̪im		**di**m[b]	
DU	ŋa**ɾi**mka	ŋem-ŋinda[c]	ni**ɾi**mka	t̪im-ŋinda	**ka**ɾimka	**di**m-ŋinda
PC	ŋa**ɾi**m	ŋa**ɾi**mka-ŋime	ni**ɾi**m	ni**ɾi**mka-ŋime	**ka**ɾim	**ka**ɾimka-ŋime
PL	ŋa**ɾi**m		ni**ɾi**m		**ka**ɾim	

[a] The inclusive forms are not represented in this paradigm because they are not sensitive to the same number distinctions as other forms.
[b] 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.
[c] 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.

Almost every person–number exponent in the language adopts a paradigmatic configuration that is unnatural. Some forms (e.g. the /di/ in bold in Table 4.91, but also forms like /t̪i/ and /ŋe/) appear, within a given person, in the singular and the dual non-sibling. Other forms (e.g. the shaded /ka/ but also /ŋa/ and /ni/) appear, within a given person, in the opposite set of cells, i.e. in the dual sibling and in the paucal and plural. Other forms (e.g. /ɾi/) 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 /ɾi/ also appears at 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). The association of concrete exponents with particular morphomes is therefore also not always straightforward. Because of this 'misplaced' number morphology in non-sibling forms, only tense forms are semantically well-behaved in Murrinh-Patha. Thus, the perfect marker *-m* in Table 4.97 is opposed to zero in the future and to *-ni/-ne* in the imperfect).

The morphomic categories described here are instantiated by many different forms, which depend on the person, tense, or verb/conjugation. SG+DU.NSIB can be instantiated by person-specific forms like 1 /ŋe/, /ŋa/, 2 /t̪i/, /d/, /n/, 3 /di/, /w/, /j/, etc. or by person-indifferent forms like /ɾi/, /ɾ/, /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 /ɾi/, /ɾa/, /je/, /ɻ/, /ɳ/, /nn/, /ll/, and /ɖɖ/.

Murrinh-Patha1: SG/DU.NSIB
Murrinh-Patha2: DU.SIB/PC/PL

4.2.4.14 Ngkolmpu and Nen (Evans 2015; 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 prefixes[24] are particularly interesting. As Table 4.92 illustrates, their form changes according to person and number. Two of the three forms distinguished, however, are not aligned to a 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.

Table 4.92 Three tense subparadigms of the copula in Ngkolmpu (Carroll 2016: 245)

	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

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. As Table 4.93 illustrates, 2SG and 1PL, and 3SG and 2/3PL are not always syncretic beyond Ngkolmpu. Although it is, at present, not entirely clear which

[24] '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.

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 (see Komnzo) which merged in Ngkolmpu because of a sound change (/θ/>/s/).

Table 4.93 Undergoer prefixes in Nen and Komnzo

	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-	g̅-	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-

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 syncretism of that cell with the 2SG. This newly acquired morphological affinity of PL+3SG in Nen should also be regarded as morphomic, however, according to the present criteria.

Ngkolmpu1: 2SG/1PL
Ngkolmpu2: 3SG/2PL/3PL
Nen: 3SG/PL

4.2.4.15 Nimboran (Anceaux 1965; Inkelas 1993)

The Nimboran language (Nimboranic, New Guinea) is well known for its baroque verbal complex. The most interesting feature regarding morphomes is its stem alternation, which appears to correlate only imperfectly with the marking of number. Three stems are distinguished, whose distribution also matches that of certain suffixes (see Table 4.94).

Table 4.94 Nimboran 'draw', unspecified object, momentary, future (Anceaux 1965: 186)

	SG	DU	PL
1EXCL	ŋgedúo-d-u	ŋgedóu-ke-d-ú	ŋgedói-d-i-u
1INCL	–	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
3M	ŋgedúo-d-um	ŋgedóu-ke-d-úm	

What Anceaux labels 'singular stem' occurs also in 1+2 (i.e. in the 1DU inclusive). The so-called dual stem, and the suffix *-ke*, in turn, occur in 1DU.EXCL, 2DU, and 3DU, but also in 1PL.INCL and 2PL. The 'plural' stem, in turn, can occur with 1EXCL and 3 but, crucially, not with 1INCL or 2. These last facts are crucial for regarding this system as unmistakably morphomic since, although it resembles 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 *ŋgedóukedé* (instead of expected **ŋgedóidie*) makes the so-called dual stem morphomic as defined in this book.

Stem alternations are formally diverse (e.g. *suáŋ*[SG] *sáoŋ*[DU] *saóiŋ*[PL] 'water', *ŋgeduá*[SG] *ŋgedáu*[DU] *ŋgedói*[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 Table 4.95 shows, the earlier dual stem occurs now in the singular, and the earlier plural stem has spread to the dual. It any case, the synchronic distribution of the so-called dual[25] and plural stems in Nimboran is synchronically morphomic.

Table 4.95 Verb 'draw' in Nimboran, durative, present (Anceaux 1965: 236)

	SG	DU	PL
1EXCL	ŋgedóu-t-emné-y	ŋgedói-t-i-emné-y	
1INCL	-	ŋgedóu-t-emené-m	ŋgedói-t-i-emné-m
2	ŋgedóu-t-emné-i	ŋgedói-t-i-emné-i	
3M	ŋgedóu-t-emné-m	ŋgedói-t-i-emné-m	
3N	ŋgedóu-t-emné-m	ŋgedói-t-i-emnyé-m	

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 in related languages with a somewhat different distribution in the paradigm. In Kemtuk (van der Wilden 1976: 73–4), for example, the dual suffix *-ke* that we see in Table 4.94, 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.Momentary/2PL.Momentary/SG.Durative
Nimboran2: 1PL.Momentary/3PL.Momentary/NSG.Durative

[25] The dual stem is sometimes regarded as a default in the literature. This theoretical status may be derived from the greater morphological and distributional diversity of this stem compared to the others.

4.2.4.16 Sobei (Sterner 1987)

In the Oceanic language Sobei, around 20 verbs show a stem vowel apophony in their person–number infection in the present-tense. As Table 4.96 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 always with the forms /o/ (vs /a/), /i/ (vs /a/), and /i/ (vs /ei/). The forms and paradigmatic distributions involved mean that both parts of the paradigm qualify for morphomicity.

Table 4.96 Partial paradigm of two Sobei verbs (Sterner 1987: 41, 43)

	'slide'				'come'			
	Realis		**Irrealis**		**Realis**		**Irrealis**	
	SG	PL	SG	PL	SG	PL	SG	PL
1EXCL	i-tosis	me-tosis	ye-tasis	'e'-tasis	yo-mi	mi-mi	i-ma	'a'-ma
1INCL	–	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

Sobei1: 1SG/2SG/PL
Sobei2: 3SG.R/I

4.2.4.17 Vitu (van den Berg and Bachet 2006)

TAM particles in Vitu (Oceanic) change form according depending on the person–number of the subject. Consider the particles in Table 4.97. As van den Berg and 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'.

Table 4.97 Forms of some TAM particles of Vitu (van den Berg and Bachet 2006: 97)

	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

4.2.4.18 Vurës (Malau 2016)

In some Oceanic languages (see also Kele in Section 4.2.4.7), nouns have stem alternations in their possessive paradigms. The alternation in 'hair' (Table 4.98) is also instantiated by various other vowel pairs, more especifically *i* (vs *ē*), *iē* (vs *ia*), *ö* (vs *o*), and *ë* (vs *a*).

Table 4.98 Possessive paradigm for 'hair' (Malau 2016: 275)

	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

As explained for Kele before, these vowel apophonies must have 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 the latter language, 1EXCL and 2 (all numbers) share a stem different from the one found in 3 and 1INCL.

4.2.4.19 Wubuy (Heath 1984)

Wubuy (a.k.a. Nunggubuyu) is a language from the Gunwinyguan family of northern Australia. It is characterized by extremely complex verbal morphology which seldom maps into morphosemantic natural classes. Most relevant is the domain of its two sets of person–number indexing prefixes (see Table 4.99) across different tense and polarity values.

Table 4.99 Some subject agreement prefixes from the A and the B set (Heath 1984: 348)

	A set			B set		
	SG	DU	PL	SG	DU	PL
1	ŋa-	ni:ni-	nuru-	ŋan-	na:ni-	na:mbu-
2	nun-	ni:ni-	nuru-	ba-	nimbini-	numburu-
3	ni-	wini-	wuru-	(w)ani-	(w)ambini-	(w)amburu-

The formatives of the A and the B sets are always different, and they are so in various ways (compare *ŋa-* vs *ŋan-*, *ni:ni-* vs *na:ni-*, *nuru-* vs *na:mbu-*, etc.). Despite the heterogeneous nature of the surface formal differences between the A and the B sets, the latter is formed from the former, according to Heath (1984), by the addition of a formative *wan-. This affix would have been linearized in different places depending on the accompanying affixes, and would have then undergone complex morphophonological changes (e.g. 1SGb ŋan- < *ŋa-wan-, 3DUb wambini- < *wan-wini) to render the alternations opaque. Note, in any case, that such a formative does not explain the whole diversity of exponents, for example the suppletive 2SG *nun-/ba-*, the 1>2SG forms (not shown in Table 4.99)

Table 4.100 Use of A and B agreement forms across TAM and negation (Heath 1984: 339)

	Potential past	Past actual	Present	Future	Evitative
Punctual	B	A	A	B	A
Continuous		A		B	
Negative	B	B	B	A	A

ŋunu-/(w)a-, or the contrast, only in set B. between 1nsg and 2nsg. Be that as it may, the interesting fact for the purposes of morphomicity is the paradigmatic distribution of the A and B prefixes in relation to TAM and polarity.

Within a particular tense, as Table 4.100 shows, set A and set B prefixes can appear in positive forms, in negative forms, in both forms, and in neither, which constitutes a clearly morphomic pattern. Some aspects of these two sets' distribution (e.g. their extension in the potential, past, and present) would seem to follow from a realis (set A) vs irrealis (set B) distinction; but, although this might well be the origin of this morphology, the presence synchronically of the A-set forms in the negative future and the evitative cannot be explained synchronically.

The two pronominal agreement prefix sets described so far are insufficient to distinguish all the tense and negation combinations available in the language. These emerge from the intersection of A and B prefixes with suffixes which make a larger number of distinctions. Their distribution is, however, equally troublesome morphosyntactically (see Table 4.101). The allomorphs that instantiate these suffixal distinctions depend on the verb (1 can be characterized by *-ɲ*, *-iɲ*, *-aŋ*, *-ŋ*, *-ŋaɲ*, and *-riɲ*; 2 can be zero, *-ŋi*, *-i:ni*, *-ni*, *-ndi*, *-j*, *-ŋa:*, and *-raŋi*; 3 can be *-ɲ*, *-ŋ*, *-aŋ*, *-jaŋ*, *-ŋaŋ*, *-iɲ*, and *-raŋ*; 4 can be *-na*, *-ni*, *-i:na*, *-nji:*, *-a:na*, *-ra*, *-ŋana*, *-mana*, *-u:*, and *-ji:*; and 5 can be zero, *-i*, *-u:*, *-ji*, *-wi*, *-ŋi*, *-ri*, and *-ni*). Although all the suffixes are restricted to either past or non-past contexts, the rest of their distribution is otherwise erratic.

Table 4.101 Distribution of suffixal distinctions over TAMs and negation (Heath 1984: 338)

	Potential past	Past actual	Present	Future	Evitative
Punctual	2	1	4	3	5
Continuous		2		4	
Negative	2	2	3	5	5

Wubuy1: neg.past/past.cont/pot.past.punc (Set 2 suffixes)
Wubuy2: neg.prs/fut.punc.pos (Set 3 suffixes)
Wubuy3: prs.punc/prs.co/fut.cont (Set 4 suffixes)
Wubuy4: evit.pos/evit.neg/fut.neg (Set 5 suffixes)

4.2.4.20 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 almost 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.

As Table 4.102 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/ in Table 4.102),[26] 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 4.102), but they are often also compounds of either two inflecting roots (e.g. 'follow') or an inflecting root and an invariable root (e.g. *qang-qwur*, *me-qwur*, *nyi-qwur* ... 'lie down').

Table 4.102 Three Wutung verbs (Marmion 2010: 305–6)

	'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

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 itself, 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 into an unsegmentable form (see Table 4.103).

The reason why 1SG and 2PL are almost always syncretic, as Table 4.103 suggests, is simply because those two forms had a zero prefix that left the original stem-initial consonant 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,

[26] The digraph 'ng' indicates nasalization of a previous vowel and 'q' represents /ʔ/.

Table 4.103 Wutung free pronouns, proto-prefixes, and their phonological outcomes with different stem initials (Marmion 2010)

	Pronouns		Proto-prefixes		+pṼ		+lV		+qṼ	
	SG	PL	SG	PL	SG	PL	SG	PL	SG	PL
1	nie	ne(tu)	*Ø-	*n-	p	n	l	d	q	n/hn
2	me	e(tu)	*m-	*Ø-	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

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 maintain and reinforce inherited paradigmatic affinities, like the 1SG/2PL one, and other (less robust) morphological alliances that, because of the reasons explained in Table 4.104, tend to constitute supersets of the 1SG/2PL set.

Table 4.104 Three Wutung verbs (I) (Marmion 2010: 303, 305, 311)

	'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

Sometimes, as Table 4.104 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/, /a/ above).

Other patterns constitute still larger supersets. In the paradigms in Table 4.105, 3PL is added to the previous cells as the domain 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.

Table 4.105 Three Wutung verbs (II) (Marmion 2010: 303, 305, 311)

	'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

One last pattern that is relatively recurrent[27] in Wutung involves 3SG.F and 3PL. Table 4.106 shows that these cells can be whole-word syncretic and also share various segments (i.e. /ɲ/, /i/, and /ĩ/ in Table 4.106) not present elsewhere in the paradigm.

Table 4.106 Three Wutung verbs (III) (Marmion 2010: 321, 326)

	'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

Wutung1: 1SG/2PL
Wutung2: 1SG/2PL/3SG.M
Wutung3: 1SG/2PL/3SG.M/3PL
Wutung4: 3SG.F/3PL

4.2.4.21 Yagaria (Haiman 1980; Stump 2015)

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 2SG and 1PL, which share their exponence to the exclusion of the rest of the person–number values.

Consider the paradigm in Table 4.107. As presented in Section 2.11, the morphological contrast between a *-p* in the 2SG/1PL and a *-v* in the rest of the paradigm

[27] 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.

is repeated in other moods with different exponents, for example in the indicative (*-n* vs *-Ø*), 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 segments are always these five.

Table 4.107 *hu* 'do', interrogative mood (Stump 2015: 128)

	SG	DU	PL
1	hu-ve	hu-ˈ-ve	hu-pe
2	ha-pe	ha-ˈ-ve	ha-ve
3	hi-ve	ha-ˈ-ve	ha-ve

The diachronic explanation for these alternations, advanced by Foley (1986: 251), relies on the subject suffixes he reconstructed for Proto-Gorokan. These subject suffixes (see Table 4.108) would have been followed by invariable particles marking illocutionary force (e.g. interrogative *pe*). Later sound changes would have generated morphological alternations in those particles>suffixes depending on whether they followed a nasal(-final subject suffix) 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 nasal. Similar sound changes would have given rise to the rest of the synchronically attested morphological alternations (except for *-n* vs *-Ø*, which would just continue the initial situation (see Table 4.109), albeit with a resegmentation of the final nasal).

Table 4.108 Proto-Gorokan subject suffixes (Foley 1986: 74)

	SG	DU	PL
1	-u	-us	-uN
2	-a:N	-a:s	-a:
3	-i	-a:s	-a:

Table 4.109 *ormi* 'come down' indicative mood (Haiman 1980: 121)

	SG	DU	PL
1	ormu-e	ormu-ˈ-e	ormu-ne
2	ormi-ne	ermi-ˈ-e	ermi-e
3	ormi-e	ermi-ˈ-e	ermi-e

4.2.4.22 Yele (Henderson 1995)

In the Papuan insular isolate Yele, certain (third-person object) number agreement formatives have an idiosyncratic allomorphy determined by both TAM and the person and number of the subject.

Table 4.110 The morphology of object number in Yele (Henderson 1995: 39)

	ma 'eat' near past punctiliar, PL object			Remote past 3PL			Remote past 3SG		
	SG	DU	PL	SG	DU	PL	SG	DU	PL
1	nî ma **té**	nyi ma **té**	nmî ma **té**	too	too	too	ngê	ngê	ngê
2	nyi ma **té**	dpî ma **t:oo**	nmyi ma **t:oo**	too	tumo	tumo	ngê	ngópu	ngópu
3	Ø ma **té**	Ø ma **t:oo**	Ø ma **t:oo**	too	tumo	tumo	ngê	ngópu	ngópu

Consider the forms in Table 4.110. The verbal inflectional morphology of Yele (see e.g. *ma* 'eat' above) is characterized by cumulative but phonologically autonomous morphs. The ones before the root (e.g. *nî*, *nyi*) change according to TAM, and subject person and number. The morphs after the lexical root (the ones that concern us here, in bold in Table 4.110) indicate the number of a third-person object. They take different forms, however, also depending on TAM and the person–number of the subject. In the exact same way as the genetically unrelated morphomic allomorphy of Benabena verbs (see Table 4.76), one allomorph (e.g. *té*, *too*, *ngê*) is used with SG and/or 1st person subjects, and a different one (e.g. *t:oo*, *tumo*, *ngópu*) elsewhere.

4.2.5 America

4.2.5.1 Achumawi (De Angulo and Freeland 1930)

The Achumawi language (Palaihnihan, California) is characterized by complex stem alternation patterns. De Angulo and Freeland (1930) explain that most verbs distinguish three different stems, which they refer to as the 'normal', 'amplified', and 'collapsed' stems. As their names suggest, the amplified and the collapsed stems usually involve an addition and a substraction respectively of phonological material with respect to the normal stem. The different stems are not aligned to TAM or person–number distinctions. The paradigmatic domain of the normal and collapsed stems also varies from one verb to another, while the amplified stem, which appears in the indicative, subordinate, and optative moods, remains distributionally stable across these three moods and across verbs.

As Table 4.111 shows, the amplified stem *ă: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* and the collapsed stem *ú:n* appear elsewhere in the paradigm. The forms that may be present in the amplified stem but absent elsewhere are very diverse: 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 and 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 allomorphic robustness of the morpheme is, therefore, considerable.

Table 4.111 Partial paradigm of Achumawi 'come' (De Angulo and Freeland 1930: 110)

	Indicative			Volitional		
	SG	DU	PL	SG	DU	PL
1	s-**ă:n**-á	h-**ă:n**-á	h-**únn**-î:-má	l-**ú:n**-à	lh-**ú:n**-à	lh-**ú:n**-í:-dzà
2	k-**ă:n**-á	gèdz-**ă:n**-á	gèdz-**únn**-î:-má	t-**únn**-ô	dz-**únn**-í	dz-**únn**-ô
3	y-**ă:n**-á	éiy-**ă:n**-á	y-**ă:n**-íú	tsìl-**ú:n**-à	tsìnd-**ú:n**-à	tsìnd-**ú:n**-í:-dzà

Note: Cumulative forms (1>2, 3>2 etc.) have been ignored.

4.2.5.2 Aguaruna (Overall 2007)

In the possessive inflection of Aguaruna nouns (also in related Achuar, see Fast and Fast 1981: 60), the third-person and the first-person plural behave as a single morphological class and are always syncretic.

Consider the nouns in Table 4.112. Aguaruna has two main classes of nouns according to the morphological expression of the possessor (the same classes are found in related Chicham languages, see Table 4.114). Small irregularities occur in some nouns (see Table 4.113), due to sound changes or haplologies, and when this happens, the whole-word syncretism of 3+1PL is always preserved.

Table 4.112 Possessive inflection in Aguaruna (Overall 2007: 200–202)

	numpa 'blood'		*susu* 'beard'	
	SG	PL	SG	PL
1	numpa-hu	numpɨ̃	susu-hu	susu-hĩ
2	numpɨ-mɨ	numpɨ-mɨ	susu-humɨ	susu-humɨ
3	numpɨ̃	numpɨ̃	susu-hĩ	susu-hĩ

It might be interesting, in contextualizing this morphomic pattern, to mention that in other Chicham languages (e.g. in Wambisa (Peña 2016: 467) and in

Table 4.113 Possessive inflection of three irregular Aguaruna nouns (Overall 2007: 200–202)

	yatsu 'brother (of a female)'		*yawaã* 'dog'		*uwɨha* 'hand'	
	SG	PL	SG	PL	SG	PL
1	yatsu-hu	yatʃĩ	yawaã-hu	yawayĩ	uwɨ-hu	uwɨhĩ
2	yatsu-mɨ	yatsu-mɨ	yawai-mɨ	yawai-mɨ	uwɨ-humɨ	uwɨ-humɨ
3	yatʃĩ	yatʃĩ	yawayĩ	yawayĩ	uwɨhĩ	uwɨhĩ

Table 4.114 Possessive inflection of two Wambisa nouns (Peña 2016: 467)

	muuka 'head'		*nauantu* 'daughter'	
	SG	PL	SG	PL
1	muuka-ru	muukɨ̃	nauantu-ru	nauantu-rĩ
2	muuki-mɨ	muukɨ̃	nauantu-rumɨ	nauantu-rĩ
3	muukɨ̃	muukɨ̃	nauantu-rĩ	nauantu-rĩ

Shuar (Saad 2014: 49)) the cognate pattern of syncretism includes the 2PL. Notice in Table 4.114 that, besides this difference in the 2PL, the inflectional forms in Wambisa are completely parallel to the ones in Aguaruna in Table 4.112. It might be interesting to speculate here about which of the two patterns may represent the original distribution. Both an extension of a 2SG form to the 2PL and a levelling of the plural forms might seem plausible diachronic developments; however, it seems somewhat more likely that the Aguaruna syncretism (i.e. 1PL+3) represents the original one. This is supported by the presence of this syncretism in both of the deepest-level branches of Chicham (as currently understood), and by the fact that the 2SG and 2PL pronouns both have the formative *-mɨ* across the family.

4.2.5.3 Ayoreo (Ciucci 2016; Ciucci and Bertinetto 2017)

In the inflectional exponence 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. Most often (see 'fill up' in Table 4.115), a syllable appears to be deleted from the 1PL and 2PL forms (i.e. the suffixed ones). These are referred to as 'mobile syllables' in the literature, and 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 and Bertinetto 2017: 34,35).

As Table 4.115 also shows, the allomorph selection in the 1PL and 2PL suffixes correlates to whether there is a syllabic augment or not. As explained by Ciucci and Bertinetto (2017), this allomorphy is a by-product of the diachronic origin of the system. The stems and the suffixes must have been originally invariant (i.e. 1PL **-ko* and 2PL **-jo*). At some stage, word-internal elisions must have taken place

Table 4.115 Paradigm of three Ayoreo verbs (Ciucci 2016: 105–7)

	'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-tɕo	ba-tiogara	waka-tio-tɕo
3	pota	pota	tɕĩ-rate	tɕĩ-rate	tiogara	tiogara

in the suffixed forms whereas the rest 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 analysed as part of the suffix. The changes that gave rise to the system would thus be something like this: 1PL *ɲĩ-rate-ko > *ɲĩ-rat-ko > ɲĩ-ra-ko vs 2PL *wakã-rate-jo > *wakã-rat-jo > wakã-ra-tɕo. The resulting allomorphy in the suffixes must have been reanalysed by language users as a cue for the stem-final syllable deletion and thus spread to other verbs to become almost coextensive to it.[28]

4.2.5.4 Bororo (Crowell 1983)

Verbs (also other parts of speech like adpositions) are subject in Bororo (Bororoan, Brazil) to stem alternations involving the voicing of the stem onset. Alternations of /k/ and /g/ (see Table 4.116), /t/ and /d/, and /p/ and /b/ are found in many verbs with the same distribution as in 'go' above.

Table 4.116 Paradigm of the Bororo verb *kodu* 'go' (Crowell 1983: 17)

	Future		Non-future	
	SG	PL	SG	PL
1.EXCL	i-kodu-mode	xe-godu-mode	i-kodu-re	xe-godu-re
1.INCL	–	pa-godu-mode	–	pa-godu-re
2	a-kodu-mode	ta-godu-mode	a-kodu-re	ta-godu-re
3	kodu-mode	e-kodu-mode	kodu-re	e-kodu-re

With person–number prefixes of the form CV- the voiced allomorph is found. This alternation must have originated as a sound change, reminding of consonant gradation in Finnic, that made segments voiced in this environment. It should be clarified, however, that this is no longer an automatic phonological rule. The preposition *ki* 'up', for example, like other prepositions in the language, takes on

[28] Some mobile-syllable-related allomorphy remains in the suffixes. E.g. 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 *-tɕo*). Isolated cases also exist where two syllables are elided (see 'deserve' in Table 4.115), and of the use of suffixes *-ko* and *-tɕo* in the absence of stem elisions (e.g. 1SG *ji-garu*, 1PL *ji-garu-ko* 'to tie, to fasten', Ciucci and Bertinetto 2017: 34, 35).

the same person–number prefixes of verbs to express their notional complement. In combination with CV- prefixes, however, and unlike in other prepositions and verbs, its form remains unaltered (i.e. *pa-ki*, *pa-gi).

4.2.5.5 Chinantec, Lealao (Rupp 1989; 1996; Feist and Palancar 2015)

Chinantecan languages, and the Oto-manguean phylum more generally, are renowned for their prominent use of stem alternations in verbal inflection. Lealao Chinantec, for example, is representative of the kinds of alternations one may find. Inflectional affixes distinguish a total of seven person–number values (all three-persons and two number combinations plus a 1PL.INCL). The segmental and suprasegmental alternations in the stems, however, show less formal diversity and only distinguish four person–number combinations (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, alternations within a single verb's paradigm are usually unnatural.

As the paradigms in Table 4.117 illustrate, a stem alternant characterized by palatalizations and vowel raisings occurs in the 1PL of the irrealis, and in the 1PL and 2 of the completive. In other verbs, this stem appears in a superset of these cells. In addition to those in Table 4.117, the same stem appears in the third-person across all aspects, as well as in the 1PL incompletive (see Table 4.118).

Table 4.117 Stem alternants in two Lealao Chinantec verbs (I) (Feist and Palancar 2015)

	'grab'						'listen'					
	Incompletive		Irrealis		Completive		Incompletive		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 4.118 Stem alternants in two Lealao Chinantec verbs (II) (Feist and Palancar 2015)

	'pay'						'open'					
	Incompletive		Irrealis		Completive		Incompletive		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

These stem alternations are also present, with a similar paradigmatic domain, 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: 21–2). They most likely go back to a single segment /j/ which was infixed, as in the verb 'open' in Table 4.118, between the stem onset and the stem vowel. Similar formatives (i.e. inflectional infixed yods) are not uncommon in Mesoamerica (e.g. in Tol, see Holt 1999, and in distantly related Northern Pame, see Berthiaume 2004). The morphological diversity of alternations (including, analogically, cases of suppletion) would have emerged in Chinantec from this single formative /j/ by means of later sound changes (e.g. palatalizations and/or vowel fusions and raisings).

Chinantec, L1: 1PL.Irrealis/1PL.Completive/2.Completive
Chinantec, L2: 1PL/2.Completive/3

4.2.5.6 Chinantec, Palantla (Merrifield 1968; Feist and Palancar 2015)

The overall morphological system described for Lealao Chinantec in the previous section is by and large valid for Palantla too. The paradigmatic distribution of the inherited stem alternations is also very similar in the two varieties. The first of the morphomes (Table 4.117) differs from the pattern found in Palantla (Table 4.119) only in a single cell in the paradigm. This morpheme extends to the 1PL progressive/incompletive in Palantla whereas it did not do so in Lealao.

Table 4.119 Stem alternation in two Palantla Chinantec verbs (Merrifield 1968: 41)

	'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	hi̢	hi̢	hi̢	hi̢	hi̢	hi̢
2	la	la	la	la	lye	lye	hi̢	hi̢	hi̢	hi̢	hi̢	hi̢
3	la	la	la	la	la	la	hi̢	hi̢	hi̢	hi̢	hi̢	hi̢

Although Rensch (1989: 21–2) presented the one in Palantla as the original domain of the alternation, comparison with other Chinantecan varieties suggests that it might be Lealao which presents the original paradigmatic distribution, as 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 distribution of this alternation, the small change in Palantla would seem to be aimed at making the pattern of stem alternation more similar to the language's other morphome,[29] which has an identical distribution to the one in Lealao.

It might be interesting to note that, in both Lealao and Palantla Chinantec, one morphome constitutes a subset of the other. Something similar has been found throughout this section in the morphomes of Khaling, Saami, and Wutung (see

[29] In Palantla, the paradigmatic distribution of the larger morphome can be stated as: 'smaller morphome'+3, which could be taken to be a simpler description than the relationship between the two morphomes in Lealao: 'larger morphome'='smaller morphome'+3+1PL.Progressive.

Sections 4.2.2.7, 4.2.3.11, and 4.2.4.20). This seems to be a trademark feature of the architecture of many morphome systems. As discussed by Maiden (2018b: 13–14),[30] 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 a palatal stem (e.g. *nia*) in the 3sg (see Table 4.118) allows speakers to infer the use of the same stem in the 1pl. Note that the predictability does not hold in the opposite direction: use of the stem in the 1pl does not reveal whether the same form will be used in the 3sg (notice the difference between the paradigms of 'listen' and 'open' in this respect).

4.2.5.7 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 to 1sg+3. The alternations, displayed in Table 4.120, go back to an originally non-alternating paradigm. Van der Voort (2007: 150) argues that words like these probably had /t͡ʃ/ as their original stem-initial consonant in all the forms, as this sound is found in the closely related language Arikapu. In Jabuti, however, in some intervocalic environments, this phoneme changed to /r/ (maybe through some intermediate stage ʒ). Later on, the /r/ before nasal vowels changed in turn to /n/, thus creating the diversity of alternations found in Jabuti synchronically. It is important to note that /h/, /r/, and /n/ (and also /t͡ʃ/ for that matter) are not allophones in Jabuti synchronically but different phonemes synchronically (Pires 1992: 24–8).

Table 4.120 Paradigms of two Jabuti verbs (Pires 1992: 45–6)

	'get tired'		'fall'	
	SG	PL	SG	PL
1	habä	hi-rabä	hõkü	hi-nõkü
2	a-rabä	a-rabä	a-nõkü	a-nõkü
3	habä	habä	hõkü	hõkü

4.2.5.8 Koasati (Kimball 1985)

In the verbal person–number inflection in Koasati (also in the most closely related Muskogean languages like Alabama, see Lupardus 1982: 140), one can identify a clear morphological affinity between 2sg, 1pl, and 2pl in most conjugations. As the paradigms in Table 4.121 show, these values are marked in the same syntagmatic position within the word (e.g. compare *hófna-l* 'smell.1sg' to *ho<lí>fn*

[30] 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 in this database, however, it seems that this rarity cannot be extrapolated to morphomes as a whole.

'smell.1PL', Kimball 1985: 70), and sometimes share formal exponents as well (e.g. *-ká* and *h-k* in Table 4.121).

Table 4.121 Person–number inflection in two Koasati verbs (Kimball 1985: 76, 80–81)

	míkkon 'be a chief'		*cákkin* '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

4.2.5.9 Maijiki (Velie and Velie 1981)

The verbal morphology of Maijiki (Tucanoan) 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 changing allegiance of the 2SG. As Table 4.122 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 book.

Table 4.122 Preterite paradigm of the verb 'go' in Maijiki (Velie and Velie 1981: 124–5)

	Declarative		Interrogative	
	SG	PL	SG	PL
1	sa-hɨ	sa-hɨ	sa-te	sa-te
2M	sa-hɨ	sa-hɨ	sa-ki	sa-te
2F	sa-hɨ	sa-hɨ	sa-ko	sa-te
3M	sa-ki	sa-hɨ	sa-ki	sa-te
3F	sa-ko	sa-hɨ	sa-ko	sa-te

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 and Criswell 1993), for example, the forms appear in SG.M and SG.F contexts. In closely related Secoya (Johnson and Levinsohn 1990) and Siona (Wheeler 1970), the forms appear in the 3SG.M and 3SG.F only, 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 be the original one. The similarity of the suffixes to the 3SG.M and 3SG.F pronouns (e.g. *kũ̃* vs *ko* in Tucano) suggests that their incorporation and grammaticalization as gender–number markers probably constitutes the ultimate source of the formatives.

It is at present unclear to me what the motivation might be for the innovation in Maijiki that caused the emergence of the morphomic arrangement in Table 4.122. Language contact, however, might constitute a promising avenue for explanation here. 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). It might represent, thus, a Tucanoan attempt to replicate these foreign structures.

4.2.5.10 Mazatec, Chiquihuitlan (Jamieson 1988; Feist and Palancar 2015)

It is common for Mazatec languages (Oto-manguean) to display a morphological affinity of 1SG and 3 (in both stems and agreement suffixes), and of the converse set of cells 1PL and 2 to a smaller extent (only stems).

Table 4.123 Chiquihuitlán Mazatec verbs, positive, neutral aspect (Feist and Palancar 2015)

	'remember'		'forbid'		'scratch'	
	SG	PL	SG	PL	SG	PL
1EXCL	basɛ	časin	tsičoʔo	ničoʔin	hentsun	čhentsin
1INCL	–	časen	–	ničoʔon	–	čhentsun
2	čase	časun	ničoʔe	ničoʔun	čhentsin	čhentsun
3	basɛ	basɛ	tsičoʔo	tsičoʔo	hentsun	hentsun

Table 4.123 shows that 1SG+3 share a stem opposed to the one in 2+1PL. These morphologically diverse alternations originate from a system of auxiliaries, many of which already showed these unnatural morphological affinities, that simply 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

4.2.5.11 Me'phaa (Suárez 1983)

As in other Otomanguean languages, verbal inflection in Malinaltepec Me'phaa is complex. A tense prefix occurs first. As in the present-tense in Table 4.124, tense prefixes tend to have an /a/-containing allomorph in the singular and an /o/- or /u/-containing allomorph in the plural. Next, in many verbs but not all, comes a

2SG prefix with many different allomorphs.[31] After this comes the verb stem, which may or may not show alternations, and, in many verbs but 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). Finally, person clitics can be suffixed for disambiguation to the whole complex described so far.

Table 4.124 Some inflectional forms of 'play' in Me'phaa (Suárez 1983: 122)

	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

The morphological trait that is most relevant here is that there are several irregular verbs in the language that display forms in 2SG+PL which are not present in the 1SG and 3SG. As Table 4.125 shows, the forms involved are 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).

Table 4.125 Some inflectional forms in Malinaltepec Me'phaa (Suárez 1983: 155, 158, 160)

	'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

The concrete changes by which these stem alternations emerged are not entirely clear, but may have involved the effects in the stem of both (i) the 2SG agreement prefix present in a great number of verbs (see Table 4.130 and n.31) and (ii) the back vowel of the tense prefixes found with plural subjects (see the present in Table 4.124). 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: 159–60). 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

[31] The allomorphs are: *ta-*, *t-*, *tha-*, *ra-*, *tra-*, *štr-*, *šta*, all characterized by an alveolar stop which has sometimes become /r/ as a result of sound change.

to some or most of these is that the 2SG and the PL alternants are often characterized by alveolars opposed to velars in the rest of the paradigm. This morphology, which probably resulted from regular sound changes, may have been the reason for the occasional merger of some 2SG and PL stems into a single form.

4.2.5.12 Páez (Jung 1989)

In the Colombian isolate Páez, the 2SG feminine and the 2PL are always syncretic (see Table 4.126). This is so in every single TAM and across various exponents. Despite the morphological diversity, one can spot a segment sequence common to all of the morphs 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 affixes, all involve adding segments to an invariable sequence: ([*iʔ*]*k*)*we*.

Table 4.126 Two suffix sets in Páez (Jung 1989: 124)

	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

4.2.5.13 Tapieté (Gonzalez 2005)

The marking of person–number in the verb in Tapieté (Tupian, Bolivia) follows an active–inactive system. The forms in Table 4.127 appear in active intransitive verbs and in transitives.[32] Here, the 1PL.EXCL shares the same prefix as 3. This has different allomorphs in different verbs, so the prefixal syncretism is systematic. Judging by the cognates of the suffix *-ha* in other Tupi-Guarani languages, González (2005: 145–6) argues that the morphology in 1PL.EXCL 'may be a recycling of the agentive nominalization of the verbal root'. Its new function may have been acquired via the impersonal, which has the same form in the language (see Toposa and Kariña for comparable diachronic developments).

[32] In Tapieté and Tupi-Guarani, a hierarchy 1>2>3 determines which argument (agent or patient) is indexed on the verb. Thus, if one of the arguments is 1, this will be the indexed one. If there is no first-person argument, agreement will be with the second-person argument if there is one. Only in in 3>3 contexts will the agreement be with 3. In addition to this, 1>2 contexts have cumulative marking. See Jensen (1990) for more details.

Table 4.127 Person–number forms of three Tapieté verbs (González 2005: 143–145)

	'sleep'		'die'		'bring' (3SG object)	
	SG	PL	SG	PL	SG	PL
1.EXCL	a-che	o-che-ha	a-mano	∅-mano-ha	a-ru	we-ru-ha
1.INCL	–	ya-che	–	ya-mano	–	ya-ru
2	ndɨ-che	pɨ-che	ndɨ-mano	pɨ-mano	nde-ru	pe-ru
3	o-che	o-che	∅-mano	∅-mano	we-ru	we-ru

4.2.5.14 Tol (Dennis 1992; Holt 1999)

Person–number agreement inflection in Tol (Jicaquean, Honduras) is characterized by complex segmental alternations in stems.[33] As Table 4.128 shows, Class 1 verbs in Tol (mostly transitive verbs) show a morphological affinity of SG and 1PL. In these values, in both the past and 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, its vowel may also differ from SG+1PL to 2PL/3PL.

Table 4.128 Past-tense inflection of some class 1 Tol verbs (Holt 1999: 23)

	sɨpɨ 'hit'		*nuku* 'see'		*pake* 'write'	
	SG	PL	SG	PL	SG	PL
1	syɨp^h	syɨpɨk^h	t^{h}inyuk^h	t^{h}inyukuk^h	tepyak^h	tepyakek^h
2	syɨp^h	sɨpɨ	t^{h}inyuk^h	t^{h}unuku	tepyak^h	tapake
3	syɨpa	sɨp^h	t^{h}inyuka	t^{h}unuk^h	tepyaka	tapak^h

In addition to this alternation, a different stem consonant allomorphy can also be found in some verbs of Class 1. These verbs in Tol (see Table 4.129) show a morphological alternation in the right edge of the stem, with one stem (e.g. *hok*^{h}) appearing in unsuffixed paradigm cells and another one (*hoʔ[o]*) elsewhere. The alternations are very diverse morphologically (parallelly to *hok*^{h}*/hoʔ* [see Table 4.125] we have *tat*^{h}*/taʔ* 'have', *k*^{h}*ol/k*^{h}*ok*^{h} 'grind', *sok*^{h}*/sok*^{h}*t* 'untie', *la/lah* 'eat', *ʔinan/ʔiʔn* 'kill', etc., see Dennis 1992: 54–5). Although the differential phonological environment (i.e. suffixed vs unsuffixed) was probably responsible for their emergence, there is little hope for a phonological derivation of these alternations in synchrony, given their morphological diversity.

As the present-tense paradigm of 'cut' shows, 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 mostly consonants at the right edge of the

[33] 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

Table 4.129 Inflectional paradigm of verb 'cut', class 1 (Dennis 1992: 21, 33)

	Present[a]		Future	
	SG	PL	SG	PL
1	hyokʰ	hyoʔ-o-kʰ	mo-hokʰ	mo-hoʔ-o-kʰ
2	hyokʰ	hoʔ-o	mo-hoʔ-o-n	mo-hoʔ-o
3	hyoʔ-o	ha-hokʰ	mo-hoʔ-o-s	mo-hoʔ-o-kʰ

[a] The past-tense behaves as the present for the purposes of this 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).

stem) are fully compatible and participate actively in the system of morphological distinctions in the language.

The other big class of verbs in Tol (Class 2, mostly intransitive) shows a completely 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 the 2PL instead. As Table 4.130 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.

Table 4.130 Inflectional paradigm of *ʔasʔɨ* 'bathe', class 2 (Holt 1999: 29)

	Present		Past	
	SG	PL	SG	PL
1	ʔosʔis	ʔyasʔikekʰ	tʰaʔasʔis	tʰeʔyasʔikekʰ
2	ʔosʔim	ʔosʔike	tʰaʔasʔim	tʰaʔasʔike
3	ʔosʔɨ	ʔyasʔiŋ	tʰaʔasʔɨ	tʰeʔyasʔiŋ

Like vowel apophonies, stem-right-edge alternations also show a very different pattern in Class 2. The stem alternations illustrated in Table 4.131 are also

Table 4.131 Partial paradigm of verb 'drink, Class 2 (Dennis 1992: 65, 74)

	Present[a]		Future	
	SG	PL	SG	PL
1	miʔ-ɨ-s	myis-ikekʰ	ka miʔ-ɨ-s	ka mɨs-ikekʰ
2	mɨs	mɨs-ike	ka miʔ-ɨ-m	ka mɨs-ike
3	miʔ-ɨ	myiʔ-ɨ-n	ka miʔ-ɨ-m	ka miʔ-ɨ-n

[a] The past-tense again behaves as the present for the purposes of this morphological alternation.

former sound changes in the language; I am sceptical of the validity of this analysis in synchrony, however, and I will only deal with surface forms here.

morphologically diverse, although less so than those of Class 1. Alongside *mɨʔ/mis* we find *pʰak/pʰaʔ* 'hear' and *peʔ/pec* 'defecate'.

Last but not least, the irregular verb 'go' shows yet another stem alternation pattern whereby the 3pl form shares morphology with sg forms across all tenses (see Table 4.132).

Table 4.132 Paradigm of Tol 'go' (Holt 1999: 30)

	Present		Past		Future	
	sg	pl	sg	pl	sg	pl
1	hum	leke	tʰum	tleke	mis	nlaka
2	hay	lowa	tʰay	tlowa	mim	nlawu
3	hama	hil	tʰemey	tʰil	mim	mal

Tol1: sg/1pl
Tol2: 3sg.prs/1pl.prs/2pl.prs/2sg.fut/3sg.fut/pl.fut
Tol3: 1sg.prs/2sg.prs/3pl.prs/1sg.fut
Tol4: sg/2pl
Tol5: 1sg.prs/3.prs/sg.fut/3pl.fut
Tol6: 2sg.prs/1pl.prs/2pl.prs/1pl.fut/2pl.fut
Tol7: sg/3pl

4.2.5.15 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 and are always syncretic. This falling-together of an unnatural class of cells with different formatives (see Table 4.133) constitutes a morphome according to our definition (see the Section 4.2.5.2 on Aguaruna for diachronic insights on this pattern).

Table 4.133 Possessive inflection of two Wambisa nouns (Peña 2016: 467)

	muuka 'head'		*nauantu* 'daughter'	
	sg	pl	sg	pl
1	muuka-ru	muukɨ̃	nauantu-ru	nauantu-rĩ
2	muukɨ-mɨ	muukɨ̃	nauantu-rumɨ	nauantu-rĩ
3	muukɨ̃	muukɨ̃	nauantu-rĩ	nauantu-rĩ

Another area of the Wambisa grammar where a morphological affinity is observed within an unnatural set of cells is the different-subject morphology of the verb, where the 1SG and the third-person are syncretic. As shown in Table 4.134, these cells are characterized by shared forms, which changes from simultaneous to sequential DS.[34]

Table 4.134 Different subject inflection in the Wambisa verb *puhu* 'live' (Peña 2016: 808)

	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

The same morphomic affinity holds in the related Chicham languages Achuar (Fast and 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 (see Table 4.135).

Table 4.135 Different subject inflection in Aguaruna *antu* 'hear' (Overall 2007: 398–9)

	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-mɨ-nĩ	anta-ku-humɨ-nĩ	antu-ka-mɨ-nĩ	antu-ka-humɨ-nĩ
3	anta-ku-ĩ	antu-ina-ku-ĩ	antu-ka-mataĩ	antu-ka-aha-mataĩ

There is reason to believe that Wambisa, Achuar, and Shuar continue the original system and that Aguaruna is the innovative one. 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 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 syncretism is that there is a formative *-taĩ* which appears in Aguaruna (Overall 2007: 397–8) but also, crucially, in Wambisa (Peña 2016: 812) in the first-person (both SG and PL) and in the third. This formative could thus have provided the analogical model in earlier Aguaruna to extend the suffix *-mataĩ* to the 1PL. In addition, the absence of 1SG marking (*-ha*

[34] 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.

elsewhere) in the DS verbal inflection makes the 1sg form look like the (unsuffixed) 3sg. That syncretism (1sg<3sg) could have been simply extended to the plural in Aguaruna, which would be the reason why today we find *antu-ina-ku-ĩ* in the 1pl instead of the expected **antu-ku-hi-nĩ*. By expanding its former domain (1sg/3sg/3pl) to the 1pl, this morph has transitioned from a more unnatural to a more natural distribution in Aguaruna.

Wambisa1: 3sg/pl
Wambisa2: 1sg/3

4.2.5.16 Zapotec, Yatzachi, and Texmelucan (Butler 1980; Speck 1984)

In some varieties of Zapotec (Otomanguean), 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 4.136) that is absent from the rest of the paradigm.

Table 4.136 Partial paradigm of 'study', progressive (Butler 1980: 147–8)

	sg	pl
1excl	ch-sed-a'[a]	ch-sed-to'
1incl	–	ch-sed-cho
2	ch-sed-o'	ch-sed-le
3	ch-sed-bo'	ch-**əsə'ə**-sed-bo'

[a] The progressive is marked with *ch-* and *-sed-* is the stem.

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 in other aspects. In Table 4.137, the 3sg.Completive *gw-**lez**-bo'* is opposed to 3pl.Completive *go-sə'ə-**bez**-bo'*.

Table 4.137 Stem of 'wait' (Butler 1980: 86)

	Progressive		Stative		Completive		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

This pattern must also have emerged as a result of sound changes operating in different environments (notice that the 3pl prefix occurs between the aspect prefix and the stem); however, these alternations are phonologically unmotivated synchronically (compare *bez/lez/cuez* in 'wait' to *bež/chež/cuež* in 'cry', or to *yis/dis/chis* in 'distribute'). Vocalic alternations are also found in vowel-initial stems, for example *ene'e/one'e/ene'e* in 'want', *on/en/on* (notice the reversed vowels) in 'do', or *ol/il/ol* in 'sing'. Some vowel-initial verbs also add a consonant to the stem in these same cells (e.g. *ao/dao/ao* 'come').

In other Zapotec varieties (e.g. Zaniza and Texmelucan, see Table 4.138), rather than being 'missing' from some cell where they might have been expected, these completive roots have spread in the paradigm to the first-person forms of all other TAMs (see Operstein 2002).

Table 4.138 Stem of 'distribute' in Texmelucan Zapotec (Speck 1984: 156)

	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

Different forms are involved in other verbs (e.g. *loo* vs *boo* 'remove', *dub* vs *ub* 'catch', *ruz* vs *az* 'be beaten'). The morphology involved is very similar to the one presented for Yatzachi Zapotec, which confirms that they are cognate alternations. The extension of the completive stem in Texmelucan is taken to have started in the 1PL. According to Operstein (2002), hortative/imperative forms (which have a close morphological affinity to the completive in Zapotec) would have begun to be used in the 1PL of other TAMs.[35]

4.3 Measuring cross-linguistic variation in morphomes

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 and 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 and Klamer 2018: 2–3)

[35] The state of affairs where the completive root appears in the 1PL but not in the 1SG seems to be documented in a Zapotec variety from the 16th century. A similar development took place in standard Italian, where the former 1PL subjunctive spread to the 1PL indicative.

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 differences between (paradigmatically) related words. The second refers to the morphosyntactic or semantic distribution[36] of these differences. These two dimensions of morphological signs are, however, complex, in that they subsume different and independent axes of variation.

In order to systematically analyse variation, 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 and Nichols 2002; Brown and Chumakina 2013) basically consist of taking a broad but relatively well-defined phenomenon (e.g. clause linkage, agreement, negation, gender) 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 checking whether all logically possible combinations are attested or whether naturally occurring examples actually cluster around a restricted set of frequent values or value combinations.

The challenges of applying this methodology to the study of the morphome are, obviously, considerable. First, whereas terms like 'agreement', 'negation', or 'gender' belong to the terminological toolkit of most theorists and field linguists, the term 'morphome' does not. Consequently, finding morphomes 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 analysed with some discrepancies). By contrast, the term 'morphome' has been applied to many different phenomena and objects in ways which are not always entirely consistent, and some linguists even reject the notion altogether. This makes it, therefore, a more difficult object of study than the average linguistic phenomenon, and may explain why there have not been any typological approaches to the morphome so far.

Taking as the starting point the operationalization of the morphome that was advanced in Section 4.1, this section will present the various ways in which morphomes may differ from one another. Following the spirit of the AUTOTYP[37]

[36] 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'.

[37] 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.

research programme, the various dimensions/categories/variables that are presented throughout this section have emerged inductively from the individual examples of morphomes that were presented in Section 4.2.

In this process, it was found that the overall **distribution** of a form can be decomposed in different, finer-grained 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, in how many lexemes, etc. Different aspects of a morphome's **form**, in turn, can also be identified: 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, therefore, have to be captured and operationalized separately. In the rest of this section I will present the underlying variables, and I will propose ways to measure this variation objectively. After a theoretical exposition of each variable I will present an overview of the empirical data in the morphome database of Section 4.2. The values of all morphomes for all variables can be consulted in the Appendix.

4.3.1 External morphosyntactic constraints

Not only morphemes, but also morphomes, can be circumscribed to particular inflectional subdomains. Even some of the most famous morphomes in the literature are somewhat unmorphomelike, as it were, in that they, like 'meaningful' formatives, are limited in their distribution to particular morphosyntactic or semantic contexts/values.

Consider, for instance, the paradigmatic distribution of the Spanish L-morphome, which occurs in the 1SG of the present indicative and through the present subjunctive. All of its cells share a tense value 'present'. This will be referred to as a 'strong' morphosyntactic constraint: all the cells within a morphome have a certain value in common. 'Weak' constraints, on the other hand, are those by which a morphome'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 morphome, thus, would be subject to a weak morphosyntactic constraint. The overall morphosyntactic constrainedness of a morphome, therefore, has been measured here by two different variables which register the number of distributional constraints of each kind that a morphome's exponents are subject to.

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.'

In the present database, morphomes have been found to range between complete morphosyntactic unconstrainedness (i.e. no restrictions of either type) and being subject to two strong and two weak constraints simultaneously. Consider, first, one of the most restricted morphomes in Table 4.139. The diphthongizations that often constitute the exponents of the N-morphome are highly paradigmatically restricted in this variety of Asturian. Its three constitutive cells are all 'present-tense' (strong constraint 1), 'indicative mood' (strong constraint 2), and 'non-1' (weak constraint). Despite all these morphosyntactic restrictions, the forms continue to be morphomic according to the criteria used in this book.

Table 4.139 Partial paradigm of *mur'der* 'bite' in western Asturias (Bybee 1985: 73)

	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

At the opposite pole of this variable, many morphomes have been found to be completely unrestricted in their paradigmatic distribution. Consider the one in Table 4.140. In Skolt Saami *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 first, second, and third-person. Its distribution is, thus, morphosyntactically unconstrained.

Table 4.140 Skolt Saami *kuullâd* 'hear' (Feist 2011: 115)

	PRS		**PAST**	
	SG	**PL**	**SG**	**PL**
1	kuul-am	kuull-âp	ku'll-em	kuul-im
2	kuul-ak	kuull-ve'ted	ku'll-ik̆	kuul-id
3	kooll	ko'll-e	kuul-i	ku'll-e

As for the overall numbers[38] in the morphome database, Figure 4.3 gives an overview of how the morphomes tend to fare according to their morphosyntactic restrictedness.

[38] In this section, only averages and other basic descriptive statistics will be presented. The analysis of correlations between variables, and statistical significance matters, are dealt with in Section 4.4.

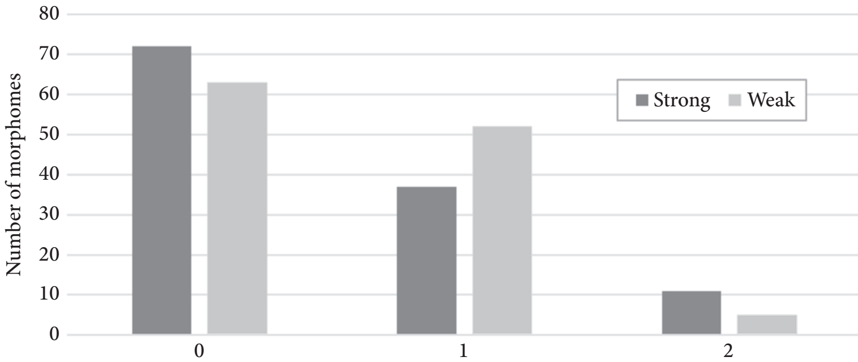

Figure 4.3 Morphomes and their morphosyntactic constraints

Most of the morphomes in this database (87, 65.2%) are characterized either by no constraint whatsoever or by just a single one. This is probably unsurprising, as with an increased number of morphosyntactic constraints, it becomes more and more difficult, logically, to stay morphomic. Notice, therefore, that any additional weak or strong constraint upon the distribution of diphthongization in the Asturian variety described in Table 4.139, would have resulted in a morphosyntactically impeccable (i.e. a morphemic) distribution.

4.3.2 Word-form recurrence

Another dimension along which morphomes may differ is the number of distinct word forms where they appear. A morphome, as defined here, is characterized by shared form. However, despite the sharing of segments or formatives, the cells constitutive of a morphome can also display differences at the whole-word level. A morphome can thus span both syncretic (see Table 4.141) and non-syncretic (Table 4.142) word forms.

Table 4.141 Possessive inflection of two Aguaruna nouns (Overall 2007: 200–202)

	yatsu 'brother (of a female)'		*yawaã* 'dog'	
	SG	PL	SG	PL
1	yatsu-hu	yatʃĩ	yawaã-hu	yawayĩ
2	yatsu-mɨ	yatsu-mɨ	yawai-mɨ	yawai-mɨ
3	yatʃĩ	yatʃĩ	yawayĩ	yawayĩ

Table 4.142 Person–number inflection in two Koasati verbs (Kimball 1985: 76, 80–81)

	míkkon 'be a chief'		*cákkin* '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

The morphome in Aguaruna constitutes a whole-word syncretism of 3 and 1PL. There is only one word form in all contexts and, for the purposes of this variable's measurement, the word-form recurrence of the morphome is 1. By contrast, the 2+1PL morphome in Koasati involves different word forms in each cell, which means its word-form recurrence is 3.

A clarification is in order concerning how the number of different word forms has been counted here in concrete cases. The total number 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 analysed here. 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 metric's calculation.

Consider the Basque paradigm in Table 4.143. This morphome (marked with the suffix *-tza* in this verb) appears in person–number values 2SG, 1PL, 2PL, and 3PL. This suffix and these values are orthogonal to other morphological distinctions in the language, like tense, a fact which would multiply (from four to eight) 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 four different word forms.

Table 4.143 Paradigm of Basque *ibili* 'walk'

	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

Figure 4.4 presents an overview of how the morphomes in the database classify according to this variable. Whole-word syncretism was the most common value, found in 24 (20%) of the morphomes in the database. From there, there is a downward trend according to which morphomes that span over a greater number of

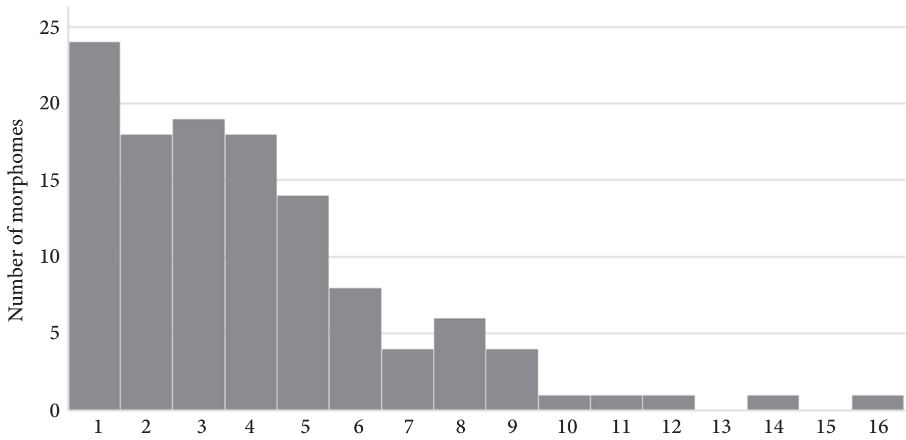

Figure 4.4 Morphomes and number of word forms

word forms are progressively less frequent. As many as 19 morphomes (15.8%), for example, spread over/contain three different word forms, 14 (11.7%) span over five different word forms, four (3.4%) extend over nine word forms, and only one has been found to span 16 distinct word forms, the maximum in the database.

4.3.3 Paradigmatic recurrence

Morphomes, as defined here, must be instantiated by more than one allomorph. This allomorphy, however, and the recurrence of a particular morphomic pattern, can take place at different levels. In the case of the morphomes that have been most frequently discussed in the literature, the different allomorphs occur in different lexemes. Thus, the L-morphome stem of Spanish *caber* 'fit' is *que**p***-, while that of *tener* 'have' is *ten**g***-. The different formal instantiations of the morphome are thus found by looking at the forms in different lexemes.

In other less frequently discussed cases, the different exponents of a morphome are found within a single lexeme's paradigm. We will say in these cases that the morphome recurs (i.e. occurs more than once) within the paradigm, with recurrence taking place in different subparadigms, i.e. under a different cross-cutting orthogonal feature value.

Consider the morphomes in Table 4.144. The morphomic affinity of 1PL/2SG/2PL occurs in Darma, with its own exponents, in both non-past (*-[h]e*) and past (*-s[u]*). Similarly, in the English verb 'be', the 2SG/1PL/2PL/3PL affinity is repeated in the paradigm in both present (*are*) and past (*were*). These morphomes will be thus classified as recurrent within the paradigm.

Table 4.144 Two single-paradigm-recurrent morphomes

	Darma *ra* 'come' (Willis 2007: 350–56)				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-ɟu	ra-ɟu	is	are	was	were

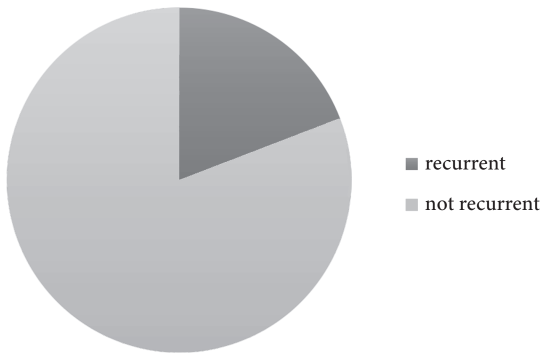

Figure 4.5 Morphomes recurrence within the paradigm

Notice how this differs from the Basque example in Table 4.143, where the 2SG/1PL/2PL/3PL morphome spans both present and past, with the same exponence, rather than being repeated with different forms in different tenses. It will thus be classified as a morphomic pattern that does not recur within the paradigm.

Figure 4.5 shows the frequency of each of these types in the database. Overall, 19.2% (N=23) of the morphomes recur within the lexeme, whereas the remaining 80.8% recur only across lexemes. It should be pointed out that there is a dependency relation of this variable on 'strong constraints' as defined in Section 4.3.1. Only those morphomes with one or more strong constraints can logically recur in the paradigm. Looking at these exclusively (which will be necessary in statistical analysis), there are only 48 morphomes which could potentially recur within the paradigm, of which almost half (48%) do.

4.3.4 Cross-lexemic recurrence

Morphomes can also differ in their 'grip' on the lexicon. Morphomes thus vary with respect to the number of lexical items they appear in, which can be easily measured as a percentage of the items in the relevant class of words. The most robust morphomes according to this variable (a) will be overtly present in every single lexical item and (b) will not have any exceptions. Note that these are slightly different things.

(a) Overt presence refers to those cases where the formal difference presupposed by the morphome is present (i.e. form A appears in the cells of the morphome and form A does not appear elsewhere). In the Spanish verb *calentar* 'heat up' (see Table 4.145), for example, the N-morphome is overt. In *orientar* 'orient', however, it is covert, since the stems within the N-morphome are indeed identical, but so are the stems in other cells of the paradigm.

Table 4.145 Present indicative of two Spanish verbs (I)

	***calentar* 'heat up'**		***orientar* 'orient'**	
	SG	PL	SG	PL
1	caliento	calentamos	oriento	orientamos
2	calientas	calentáis	orientas	orientáis
3	calienta	calientan	orienta	orientan

(b) The presence of exceptions refers to those cases where the affinity presupposed by the morphome is not observed i.e. the morphological identity that is supposed to hold within the cells of the morphome conflicts with what is actually found in the paradigm. Consider, for example, the Spanish verb *venir* 'come' (Table 4.146). Stem vowel identity within the N-morphome (SG+3PL) is broken in *venir* and 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*).

Table 4.146 Present indicative of two Spanish verbs (II)

	***calentar* 'heat up'**		***venir* 'come'**	
	SG	PL	SG	PL
1	caliento	calentamos	vengo	venimos
2	calientas	calentáis	vienes	venís
3	calienta	calientan	viene	vienen

These two variables (i.e. 'overt presence' and 'exceptions' in the lexicon) are obviously not independent, because if a lexeme, like *venir*, constitutes an exception (i.e. (b)), this entails that the morphome is not overtly present in that lexeme (i.e. (a)). Every lexeme is classifiable, thus, as either (1) showing a 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 probably preferred to type (2) lexemes, which are in turn preferred to type (3) ones. When operationalizing this variable of cross-lexemic recurrence, the ideal option might have been to measure the percentage/number of lexemes in each of the classes (1), (2), and (3). At the same time, however, a single metric seems desirable; in addition, data on exceptions has been found during the present research to be very seldom reported in descriptive grammars. Thus, 'overt presence' has been the only factor measured in this variable, also because it must the most important of these types in the 'discovery' of morphomes by either linguists or language users. As for the N-morphome in Spanish, for example, 426 verbs (see Herce Calleja 2016), or around 4% of the lexicon,[39] 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. At the opposite pole, the least recurrent possible morphome, limited to the paradigm of a single lexeme, has also been found, in English *be* (see Table 4.150).

Figure 4.6 presents an overview of the lexical recurrence of the morphomes in this database. It shows the recurrence in the lexicon of the morphomes in the present database, ordered from least to most recurrent. Of the 120 morphomes, 44

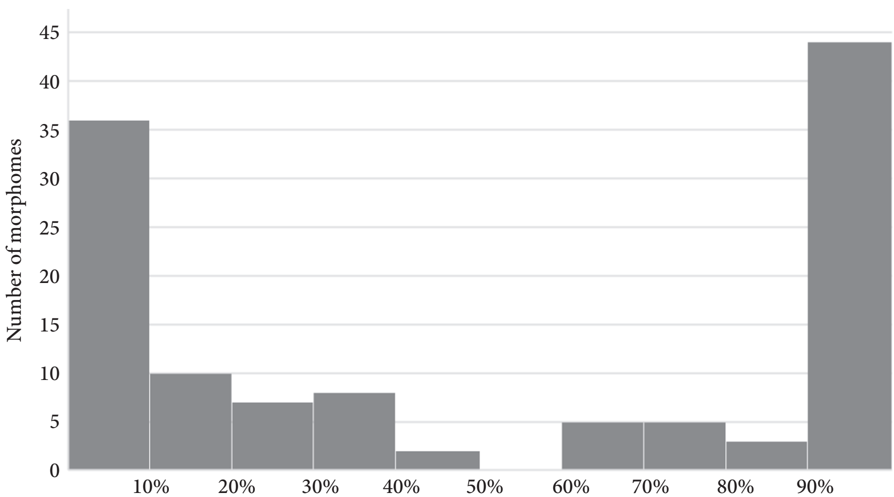

Figure 4.6 Lexical recurrence of the morphomes

[39] 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 morphome will be a finer or rougher estimation depending on the evidence (i.e. source or description) available.

appear in over 90% of the lexemes, with 38 (31.7%) of these appearing in every single lexical item in the relevant word class. At the opposite end of the scale, another 36 morphomes (30%) occur in 10% of the lexical items or fewer. There appears to be a tendency, thus, for morphomes (maybe valid for formatives in general)[40] to either occur everywhere where they possibly could, or else be limited to a comparatively small number of (irregular) lexemes.

4.3.5 Number of exponents

Patterns of morphomic exponence may vary in their formal diversity. Some unnatural patterns are instantiated by several allomorphs and others by just one form. Those morphomes that have several different forms in different lexemes or domains, are considered more systematic and robust.

Consider person–number agreement in Ayoreo in Table 4.147. The singular and third plural share form in many verbs in the language. The form of this suffix (or stem extension) differs from verb to verb. Besides the ones illustrated in Table 4.147, we also find *-gu*, *-si*, *-ru*, *-di*, *-ra*, *-ro*, *-su* ... 28 different form(ative)s are associated in Ayoreo with the context SG+3PL. An example of a morphome with somewhat less allomorphic diversity could be provided by the L-morphome of Spanish, which is found with the forms /g/ (e.g. in *tener* 'have'), /k/ (e.g. in *parecer* 'seem'), /ig/ (e.g. in *caer* 'fall'), and /ep/ (in *caber* 'fit'). In the lowest ranges of allomorphic diversity (see Table 4.148), a morphome may be instantiated by only two allomorphs.[41] The syncretisms of 2/3PL and 3SG.M in Ngkolmpu, for example, is only revealed/instantiated by two forms (*s-* and *y-*).

Table 4.147 Person–number agreement of some verbs in Ayoreo (Bertinetto 2009)

	'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

Figure 4.7 shows the properties of the morphomes in Section 4.2 by their allomorphic diversity (from fewer to more exponents). The most common value, occurring 34 times (28%), is to have two different 'allomorphs' only, which means satisfying the requirements for allomorphic diversity that were set up here minimally. Morphomes that are instantiated with more exponents are progressively

[40] Consider the possible relationship of this finding with proposed (cognitive) principles of morphological architecture like Carstairs-McCarthy's (1994) No-Blur principle.
[41] A minimum of two allomorphs was set (somewhat arbitrarily, although see Section 2.11) as the threshold to classify a pattern as systematic for the purposes of the morphome database in this book.

Table 4.148 Some durative forms of the copula in Ngkolmpu (Carroll 2016: 245)

	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

less frequent: 21 (17.5%) have three, 14 (11.7%) have four, 11 (9%) have five different allomorphs, etc. The highest value corresponds to a morpheme in Nimboran (Section 4.2.4.15) that boasts up to 30 different realizations.

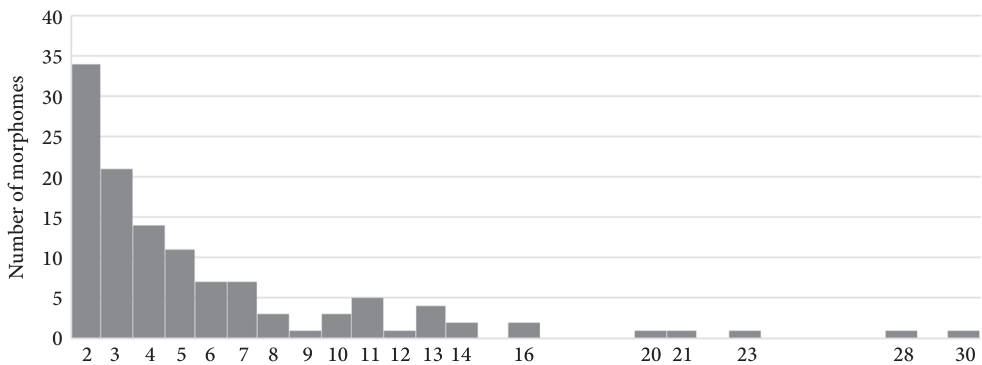

Figure 4.7 Allomorphic diversity of the morphemes

4.3.6 Shared form

Another variable that might be relevant to register in relation to morphomic variation is the 'amount' of morphological substance shared between a morphome's cells. A considerable phonological/segmental length in concrete cases increases our confidence that a pattern is something significant (rather than e.g. a segmentation glitch, see Table 2.44). It must also make a pattern more 'salient' for language users' perception and acquisition of morphological generalizations.

Table 4.149 presents four different Italian verbs, three of which contain a stem alternation pattern between 1SG/3SG/3PL and the rest of the cells. In the first verb, three segments /ɔls/ are shared by the three cells in PYTA to the exclusion of other cells. In the second verb, the number of shared segments is two /ɔl/, and in the third this is just one /e/. The last verb does not show any segments shared by these cells and would thus not count as showing the morphome. A morpheme-level measure of the amount of shared morphology can be then obtained by averaging across all its possible allomorphs, with the average for the Italian morphome in Table 4.149 calculated, in this case, at 1.5 segments on average, across its 11 distinct attested forms.

Table 4.149 PYTA stem alternations in the past-tense of three Italian verbs

	Cogliere 'pick'	*Volere* 'want'	*Fare* 'do'	*Cantare* 'sing'
1SG	ˈkɔlsi	ˈvɔlli	ˈfetʃi	kanˈtaj
2SG	koʎˈʎesti	voˈlesti	faˈtʃesti	kanˈtasti
3SG	ˈkɔlse	ˈvɔlle	ˈfetʃe	kanˈtɔ
1PL	koʎˈʎemmo	voˈlemmo	faˈtʃemmo	kanˈtammo
2PL	koʎˈʎeste	voˈleste	faˈtʃeste	kanˈtaste
3PL	ˈkɔlsero	ˈvɔllero	ˈfetʃero	kanˈtarono

The range of variation found in the present study is quite large, as morphomes have been found to range between an average of 3.7 shared segments for the one in Páez (that morphome has the allomorphs *-iʔkwe*, *-kwe*, and *-we*, see Section 4.2.5.12) and 1 (e.g. the morphome in Sobei, see Section 4.2.4.16, which has the allomorphs /o/ and /i/). The numbers for this variable are summarized in Figure 4.8.

In a way similar to several previous variables, morphomes in the database cluster towards the lower morphological robustness end of the distribution. Many morphomes in this database (46, 38.3%) are evidenced always (i.e. under all of

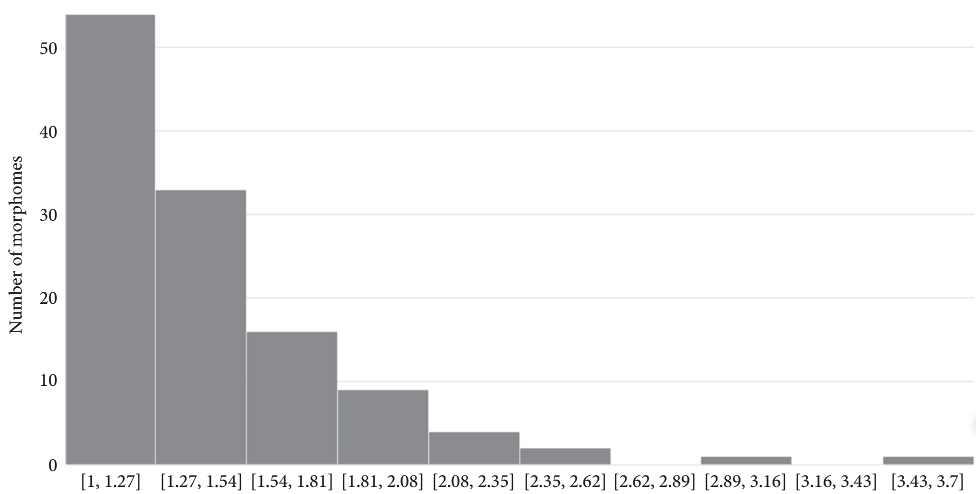

Figure 4.8 Average number of segments instantiating the morphome

its allomorphs) by just a single segment. Since only segmental morphological correlates were considered, an average of one segment is the logical minimum. Longer exponents are progressively less frequent, following a Zipfian distribution.

4.3.7 Informativity

The diversity of the patterns in Section 4.2 has also revealed that morphomes often differ in their informativity, i.e. in the extent to which they participate in the overall system of morphological contrasts in a language.

Table 4.150 presents examples of an informative (Yagaria), an uninformative/redundant morphome (Jabuti), and an intermediate one (Jerung). The 2SG+1PL morphome in Yagaria (i.e. the alternation between *-ve* and *-pe* in this particular paradigm) may be morphosyntactically unnatural but is as 'functional' as it can possibly get. Because of its perfect orthogonality to the other morphological distinction in the paradigm (*-u* vs *-a*), the morphome in Yagaria plays a fundamental role in the expression of person–number categories in the language because its presence is the only thing that distinguishes 1PL from 1SG, and 2SG from 2/3PL. It is, thus, as 'useful' as it gets (like canonical morphemes) because, like them, it is completely orthogonal to other formatives in the paradigm.

Table 4.150 Three morphomes with a different degree of informativity

	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

Contrast this to the morphomic alternation in Jabuti. The formal contrast between the stems *habä* and *rabä* is completely redundant in the language in that it does not increase the number of morphological distinctions in the paradigm. More restricted affixes (*hi-* and *a-*) occur in subsets of the 2+1PL morphome and, because they make finer-grained distinctions, they render the stem alternation functionally superfluous.

Other morphomes are intermediate between these two extreme types in that they are informative in some of its cells but redundant in others. The morphomic stem alternation in Jerung, 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-* (rather than *go-*) is the feature that distinguishes 2DU and 3DU. Figure 4.9 shows how the morphomes in the present database classify according to their informativity.

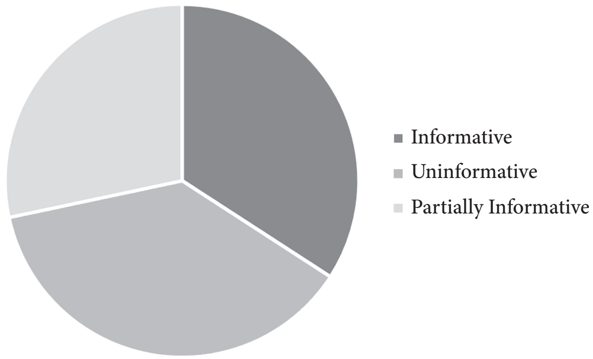

Figure 4.9 Informativity of the morphome

All three types defined here are similarly common, with 45 (37.5%) uninformative morphomes, 41 (34.2%) informative morphomes, and 34 (28.3%) partially informative ones. It may be surprising to find that, although the most frequently discussed morphomes (stem alternations in Romance) are usually redundant within the broader system of morphological contrasts, the majority of morphomes (62.5% N=75) in the present database are at least partially informative.

With regard to this variable, it may be useful to explicitly reflect on the case of morphomes where whole-word syncretism holds between their different paradigm cells (consider the verb 'sleep' in Alpago in Table 4.151). When this happens, the morphome must always be understood as completely informative since it is 'all there is', i.e. it constitutes an atomic whole as far as the morphological contrasts in the language are concerned. The morphological correlates of the N-morphome in Alpago (i.e. rhizotony and stem vowel /ɔ/) are precisely what distinguish this word form ˈ*dɔrme* and the morphome cells from others like e.g. the 2PL indicative *dorˈme*.

Table 4.151 Partial vs whole-word syncretism within the N-morphome

	Verb 'sleep' in Alpago (Zörner 1997)		**Verb 'sleep' in Spanish**	
	IND	**SBJV**	**IND**	**SBJV**
1SG	ˈdɔrme	ˈdɔrme	ˈdwermo	ˈdwerma
2SG	ˈdɔrme	ˈdɔrme	ˈdwermes	ˈdwermas
3SG	ˈdɔrme	ˈdɔrme	ˈdwerme	ˈdwerma
1PL	dorˈmɔŋ	dorˈmone	dorˈmimos	durˈmamos
2PL	dorˈme	dorˈmede	dorˈmis	durˈmajs
3PL	ˈdɔrme	ˈdɔrme	ˈdwermen	ˈdwerman

Single-word morphomes are, therefore, fully informative at the level of the morphological contrasts of the language. The fact that the value of 1 word form, in the variable defined in Section 4.3.2, logically entails a value 'informative' here will need to be addressed in the statistical analysis in Section 4.5. Excluding single-word morphomes, the number of (fully) informative morphomes is reduced to 17 (17.7% of those that can logically adopt any informativity value).

4.3.8 Morphosyntactic coherence

Another dimension along which the morphosyntactic distribution of morph(ome)s may vary concerns the internal morphosyntactic feature value relations within the morphome. 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 (see Table 4.152), we would not refer to those entities as morphomic in this book.

Table 4.152 Two Hinuq noun paradigms (Forker 2013: 55)

	'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-ɬ	mal-i-ɬ	xalq'i-la-ɬ	xalq'i-mo-ɬ
ESS2	malu-qo	mal-i-qo	xalq'i-la-qo	xalq'i-mo-qo

All non-absolutive forms of the noun in Hinuq are formed on the basis of the same stem. The so-called oblique stem may differ from the absolutive form in many different ways: by the addition of various suffixes (*-i*, *-la*, *-mo* above), ablaut, shift of stress, deletion of the final consonant, etc. 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 this, their distribution need not be described disjunctively. It displays a rectangular shape when represented in a paradigm and it does not count as a morphome for this database.

However, within the distributions that cannot be described as the realization of a value or two or more values conjunctively (i.e. within morphomic/unnatural distributions) there is still a great amount of variation. Some forms' distribution (e.g. GEN.SG and NOM.PL in Irish in Table 4.153) is such that the associated morphosyntactic contexts do not share any value whatsoever. These forms are the least natural because they effectively 'mean' one thing and the opposite. Other unnatural forms' distribution (e.g. the perfective positive suffixes from Northern

Akhvakh, where the same form is used to agree with singular and with neuter plurals) is comparatively more ‘natural’ (see Table 4.153).

Table 4.153 Morphomes of Irish and Northern Akhvakh

	Irish ‘woman’		N. Akhvakh ‘PFV.POS’		
	SG	PL		SG	PL
NOM	bean	mná	M/F	-ari	-iri
GEN	mná	ban	N	-ari	-ari

More complex two-dimensional patterns can also be found (see Burmeso in Table 4.154), and three-dimensional patterns are also attested in the database (see Menggwa Dla).

The variable that this section is presenting, i.e. the paradigmatic ‘shape’ or relative (un)naturalness of a morphome, will be operationalized here as the average percentage of feature-values shared between its cells. In the case of Irish in Table 4.153, there is only one pair of morphome cells, with 0 values shared (i.e. 0/2=0). In the case of the Northern Akhvakh pattern, there are three pairs of cells within the morphome: (M/F.SG, N.SG), (M/F.SG, N.PL), and (N.SG, N.PL), whose shared values are 1, 0, and 1 respectively out of the logical maximum of six in total. The average proportion of shared values is thus (1+0+1)/6=33.33%.

In the case of the four-cell morphome of Burmeso in Table 4.154, there are six pairs of cells within it: (II.SG, VI.SG), (II.SG, VI.PL), (II.SG, V.PL), (VI.SG, VI.PL), (VI.SG, V.PL), and (VI.PL, V.PL), whose shared values are 1, 0, 0, 1, 0, and 1 respectively, three in total out of the logical maximum 12, the average thus being 3/12=25%. In the case of the three-feature morphome of Menggwa Dla, describing its paradigmatic distribution requires reference also to four cells, with six possible pairings between them: (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 number of shared values of each of these pairs is 1, 2, 0, 2, 2, and 1 respectively from a total of 18, the average percentage of shared values thus being (1+2+0+2+2+1)/18=44.44%.

Table 4.154 Burmeso and Menggwa Dla morphomes

Gender	Burmeso Conjugation 1			Stem of ‘sleep’ in Menggwa Dla					
				Masculine			Feminine		
	SG	PL		SG	DU	PL	SG	DU	PL
II	g-	s-	1	e-	e-	e-	e-	e-	e-
VI	g-	g-	2	e-	e-	ap-	e-	e-	e-
V	j-	g-	3	ap-	e-	ap-	ap-	e-	e-

The patterns presented throughout this Section 4.3.8 vary between 0% and 50% naturalness: 0% Irish, 25% Burmeso, 33.33% Northern Akhvakh, 44.44% Menggwa Dla, 50% Hinuq. As defined here, this variable can vary between 0% and 50%. However, because of the present morphomehood requirements (see Section 4.1.1), structures of the Hinuq type have been excluded, so no morpheme here reaches the maximum of 50%.

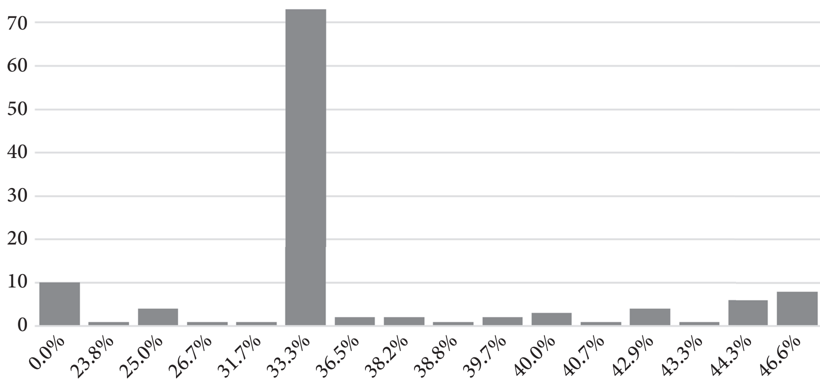

Figure 4.10 Morphosyntactic coherence (MC) of the morphomes

Figure 4.10 displays the morphosyntactic coherence of the morphomes in the present database. Although this is a continuous variable in principle, only 16 different values of MC have been found in the present morphome database, ranging from 0% to 46.6%. This is so because some cell-geometrical patterns (the simplest ones) are very common. The one illustrated by Northern Akhvakh in Table 4.153 (i.e. MC=33.3%) is the most common one in the database by a large margin (73 morphomes, 60.8%). This is followed by the patterns with 0% MC (10, 8.3%), with 44.6% MC (8, 6.7%), and with 44.3% MC (6, 5%). The most relevant findings are thus the prevalence of the value 33.3%, and the clustering of morphomes towards the higher naturalness end of the spectrum (i.e. they appear closer to the logical maximum of MC=50% than to MC=0%).

4.3.9 Morphome paradigm size and others

Although the previous metric, morphosyntactic coherence, captures most information about a morphome's distribution in the paradigm, there are some aspects that it does not. It does not address, for example, the number of cells in a morphomic paradigm.

Based on the presence of the morphomic exponents (e.g. *kɔɔn* in Table 4.155), one would need a content paradigm with just four cells (see Table 4.156) to capture these morphomes' distribution.

Table 4.155 Partial paradigms of three Fur verbs (Waag 2010)

	'tie' imperfective		'hang' imperfective		'grind' imperfective	
	SG	PL	SG	PL	SG	PL
1	ʔirg-ɛl	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	lɪg-ɛl	kalg-ɛl-ɪ	kɔɔn	kawn-ɛ
3.NHUM	rig-ɛl	rig-ɛl-ɪ	lɪg-ɛl	lɪg-ɛl-ɪ	kɔɔn	kɔɔn-ɛ

Table 4.156 Fur morphome content paradigm

	SG	PL
HUM		
NHUM		

For the purposes of cell-counting in the morphome content paradigm, values that behave identically concerning the presence or absence of the morphome will be combined into a single one, independently of these values' semantic content.

Because, with respect to the presence or absence of the morphomic exponence, the first-person in Table 4.157 behaves like the third, the morphomic affinity in Me'phaa is therefore also reducible to a four-cell content paradigm (see Table 4.158) identical to the one in Table 4.156 when one abstracts away from concrete values and row and column order.

Table 4.157 Some inflectional forms in Me'phaa (Suárez 1983: 155, 158, 160)

	'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 4.158 Me'phaa morphome content paradigm

	SG	PL
1/3		
2		

More complex morphomes will require reference to a greater number of features and values, and will contain more cells in their content paradigm. Consider the Italian L-morphome. Based on the presence or absence of the shaded stems, this

Table 4.159 Present-tense of two Italian verbs (Maiden and Robustelli 2014)

	cogliere 'pick'		*dire* 'say'	
	IND	SBJV	IND	SBJV
1SG	colgo	colga	di[k]o	di[k]a
2SG	cogli	colga	di[t͡ʃ]i	di[k]a
3SG	coglie	colga	di[t͡ʃ]e	di[k]a
1PL	cogliamo	cogliamo	di[t͡ʃ]iamo	di[t͡ʃ]iamo
2PL	cogliete	cogliate	dite	di[t͡ʃ]iate
3PL	colgono	colgano	di[k]ono	di[k]ano

Table 4.160 Italian L-morphome content paradigm

	IND		SBJV	
	SG	PL	SG	PL
1				
2				
3				

morphome (see Table 4.159) is irreducible and requires a content paradigm with 12 different cells (see Table 4.160) because all values of person (1, 2, 3), number (SG, PL), and mood (IND, SBJV) behave differently with respect to the analysed morphomic structure.

Because of the way morphomehood has been defined here (allowing only morphomes in tabular paradigms), this variable of morphome paradigm size can only take a discrete number of values: 4 (2×2), 6 (2×3), 8 (2×2×2), 9 (3×3), 12 (2×3×2, 4×3), 16 (4×4), 18 (3×3×2), etc.

Figure 4.11 shows how the morphomes in the database pattern according to this variable. The majority of the morphomes in the dataset (65.8%, N=79) can

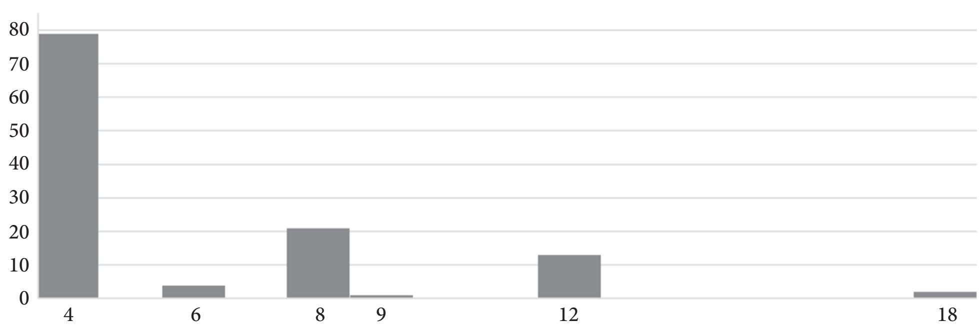

Figure 4.11 Size of morphomic paradigms

be captured in a 2×2 paradigm and have therefore the lowest possible complexity for a morphome according to this metric. Higher-complexity distributions are generally less frequent.

In Sections 4.3.8 and 4.3.9 I have tried to operationalize paradigmatic distributions into numeric variables. Although the proposed metrics are not independent[42] (e.g. 25% MC implies a six-cell morphome paradigm, a four-cell morphome paradigm implies either 0% or 33.3% MC), they do capture different facts about a morphome's paradigmatic distribution. The morphomes in Table 4.161, for example, have an identical morphosyntactic coherence of 33.3% but can be distinguished by their different morphome paradigm size, the former being 2×2×2=8, while the latter is 2×2=4.

Table 4.161 Content paradigms of two morphomes in Skolt Saami (left) and Spanish (right)

	njorggad 'whistle' (Feist 2015: 204, 210)				*perder* '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

Although these measures are useful and provide complementary information, they do not exhaust the variation found in the broader domain of paradigm distributions. Thus, knowing a morphome's morphosyntactic coherence and its paradigm size does not suffice, by itself, to capture its full distribution and geometrical shape in the abstract paradigm. Morphome distributions also differ in the number of features required in their description (e.g. 2×3×2 and 4×3 paradigms both have 12 cells but differ in this respect), and in the number of paradigm cells that the morphome spans (e.g. within a 2×2×2 paradigm, a morphome could span anywhere from two to seven cells).

Even all these four variables, however, do not suffice to capture a distribution unmistakably. The morphomes in Table 4.162,[43] for example, take identical values as per morphosyntactic coherence (46.6%), morphome paradigm size (8), features involved (3), and paradigm cells spanned by the morphome (5). However, they still constitute different configurations, by which I mean that they are not rotational or row/column-order variants of each other.

[42] For this reason, only MC will be considered in the statistical analysis in Section 4.4.

[43] The same applies to the following morphomes: Daasanach1 vs SkoltSaami2; Tol6 vs Iraqw1, Sobei2, Spanish2, Toposa, Nimboran2, Thulung2, and Udmurt; and SkoltSaami1 vs Maranunggu1, and Wubuy2.

Table 4.162 Streamlined paradigm of two morphomes

	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

This suggests, merely, that there is still work to be done with regard to the measurement and typologization of paradigmatic distributions, which I shall leave for future research.

4.3.10 Locus of marking

A salient aspect on which different morphomes may vary is the syntagmatic locus/status of the exponence that reveals them. The most well-known Romance morphomes (i.e. N-morphome, L-morphome, and PYTA) all involve changes in the stem. Some other morphomes that have been presented in this book are revealed by affixes instead (see e.g. Ngkolmpu in Table 4.148, or Darma in Table 4.144). Some morphomes involve both affixal and stem-alternation exponents, or are of uncertain classification regarding their status (see e.g. the morphomes of Basque (4.2.3.2), and Svan (4.2.2.13)).

The morphomes in the current dataset classify as in Figure 4.12 according to this. A majority of the them (52.5%, N=63) are morphological alternations in the stem, while a third (34.2% N=41) are affix-based. The remaining 16 cases (13.3%) are either mixed or borderline cases where it is difficult to decide on the 'right' status of exponents.

The theoretical status of a form as stem or affix has been taken over from the analysis in the respective language sources. From an empirical perspective, the nature of the stem–affix distinction is controversial and certainly not independent from some of the variables that have been presented in this Section 4.3 (e.g. lexical recurrence, strong morphosyntactic constraints, number of allomorphs). My understanding is, in fact, that a combination of these factors is precisely what motivates the analysis of a form as either stem or affix. Due to its theoretical nature and its lack if logical independence from the rest of the variables, the stem–affix dimension will not be included in the statistical analysis of Section 4.4, although it will still be informative to see what variables underlie linguistic analyses.

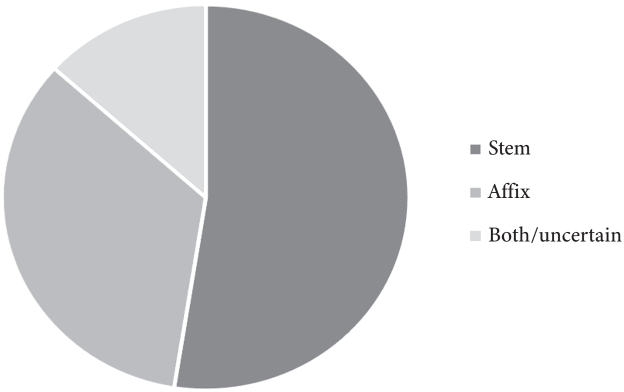

Figure 4.12 Proportion of affix and stem-based morphomes

4.3.11 Cross-linguistic morphome recurrence

4.3.11.1 Recurrent values

Although typological uniqueness has sometimes been thought of as a diagnostic of morphomicity (Maiden 2018b: 22), there is no reason whatsoever to believe that being typologically unique should be a requirement involved in the identification of any linguistic phenomenon (see Section 2.6). Under the approach espoused here, therefore, morphomic structures can 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 paradigmatic-distributionally identical morphomes in the database, all of which will be noncognate.

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 features characterized by a certain degree of universality. Grammatical cases and tenses vary quite drastically across languages in their number and the way they divide the functional–syntactic space. Grammatical genders can also split 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, dividing the domain into one (SG) vs more than one (PL) individuals seems to be the basic distinction upon which languages may occasionally add others. The following person–number morphomic patterns have been found to be recurrent:

SG+3PL

Morphomes involving SG+3PL have been found in the present sample in a total of eight languages from six different stocks: Ayoreo (Zamucoan), Bororo (Bororoan), Daju (Dajuic), Greek (IE), Jerung (Sino-Tibetan), Spanish (IE), Sunwar (Sino-Tibetan), and Tol (Jicaquean). A streamlined presentation of two of these structures is found in Table 4.163.

Table 4.163 Two languages showing SG+3PL morphomes

	Daju 'drink'		Ayoreo 'fill up'	
	SG	PL	SG	PL
1	uɾ-o	uɾ-ciga	ɲĩ-rate	ɲĩ-ra-ko
2	uɾ-o	uɾ-cini	mã-rate	wakã-ra-tɕo
3	uɾ-o	uɾ-o	tɕĩ-rate	tɕĩ-rate

Note: For the sources and for additional information on each of these morphomes, please consult the corresponding language's section in section 4.2.

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 more 'marked' values. 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 morphological 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 morphome 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. It has been found in four languages: Italian (IE), Chiquihuitlán Mazatec (Otomanguean), Tapieté (Tupi-Guaraní), and in Wambisa (Chicham).

PL+1SG/2SG/3SG

Morphological affinities between all plural persons and one of the singular ones have also been found here to be common. PL+1SG occurs in six languages: Barai (Koiarian), Luxembourgish (IE), Nivkh (Isolate), Nuer (Nilotic), Vitu and Vurës (Austronesian). A morphome spanning the values PL+2SG has been found in four languages in the sample: Basque (Isolate), English (IE), Koiari (Koiarian), and Malinaltepec Me'phaa (Otomanguean). Finally, a morphological exponence whereby PL+3SG systematically share form is also recurrent in the present sample and has also been found in four languages: Kele (Austronesian), Nen (Yam), Svan (Kartvelian), and Wambisa (Chicham).

These three-person–number patterns, illustrated in Table 4.164, share some very obvious similarities, which is the reason why they have been presented together here. The first of these is, of course, that all of them involve the similarity/identity of all the plural cells with a single singular cell. Another one is that they 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.

Table 4.164 Three languages showing PL+SG-cell morphemes

	Luxembourgish 'be'		Me'phaa 'carry' past		Wambisa 'daughter'	
	SG	PL	SG	PL	SG	PL
1	sinn	sinn	ni-gongo:	ni-rango:	nauantu-ru	nauantu-rĩ
2	bass	sidd	ni-rango:	ni-rango:	nauantu-rumɨ	nauantu-rĩ
3	ass	sinn	ni-gongo:	ni-rango:	nauantu-rĩ	nauantu-rĩ

The explanation I would like to offer here concerning the recurrence of these patterns might be more tenuous than in the case of the other recurrent patterns, as it relies partially on chance. Whole natural classes like PL will frequently share forms/morphemes. Relatively 'marked', more infrequent classes like PL will tend to do so to a greater extent than more 'unmarked'/frequent natural classes like SG. Diachronic accidents[44] 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 ones. Such paradigmatic configurations, once in place, might also be somewhat more stable than patterns like 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 better learnability and greater resilience.

As further proof that this explanation might be on the right track, it might be useful to observe that other morphomic patterns that involve a relatively infrequent natural class falling together with a relatively frequent cell outside of it can be found quite often in the present database (consider e.g. the morphomes of Sobei [Irrealis+3SG.Realis], Udmurt [Future+3PL.Present], and the Spanish L-morphome [Subjunctive+1SG.Indicative]). The opposite patterns (e.g. Present+2PL.FUTURE, Realis+1PL.Irrealis) are less common. The recurrence of the PL+1SG, PL+2SG, and PL+3SG morphomes may therefore be the combined result of learnability pressures by which these patterns are more likely to arise and less likely to be lost (for example, by falling back analogically to the closest natural class).

[44] See the sections on Basque, and Me'phaa for two different kinds of 'accidents'. Nen and Wambisa (see also Tables 4.112–4.114) also show that morphological affinities between two plural cells seem to be readily extended analogically to the third.

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. It appears (see e.g. Table 4.165) in seven languages: Darma (Sino-Tibetan), Jabuti (Macro-Je), Karamojong (Nilotic), Koasati (Muskogean), Mazatec (Otomanguean), Nuer (Nilotic), and Tol (Jicaquean).

Table 4.165 Three languages showing 2+1PL morphomes

	Karamojong indicative		Koasati 'be a chief'		Jabuti 'fall'	
	SG	PL	SG	PL	SG	PL
1	aka-	iki-	mikko-lí	mikko-t-il-ká	hõkü	hi-nõkü
2	iki-	iki-	mikko-t-is-ká	mikko-t-as-ká	a-nõkü	a-nõkü
3	a-	a-	mikkó	mikkó	hõkü	hõkü

The explanation I would like to propose here for the cross-linguistic recurrence of this particular pattern has to do with its proximity to a natural distribution. As I have argued elsewhere (Section 2.12.1), 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 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 (Trans-New Guinea).

SG+1PL

Morphomes involving SG and 1PL have been found in three unrelated languages: Yele (Isolate), Tol (Jicaquean), and Benabena (Trans-New Guinea). The same 'problematic' nature (i.e. multiple affiliation) of the 1PL (see Section 2.12.1) may also be the reason behind the recurrence of this particular pattern. In many languages, 1PL.INCL forms behave morphologically like singulars. These systems have been called minimal-augmented, and, along with clusivity itself, are most common in Circum-Pacific languages (see e.g. Cysouw 2003; Bickel and Nichols 2005). In languages (particularly from this area) with no inclusive/exclusive distinction, the undifferentiated 1PL could thus be expected to share some of its morphological properties with the other 'minimal' (i.e. SG) forms.

4.3.11.2 Recurrent patterns

As shown throughout this section, person–number agreement inflection is a morphological domain particularly suitable to finding cross-linguistically recurrent morphomic structures. This is probably just the by-product of person and number representing the most cross-linguistically frequent orthogonal features. Recurrence in other domains can be found as well, however, provided that some degree of flexibility is allowed with respect to the actual values and categories involved.

The genders of Burmeso are obviously different from the ones in Khinalugh, 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 (see Table 4.166). Although the values and categories vary, then, the patterns are still 'the same' at an abstract level. This same pattern is also found in the gender–number inflectional system of Mian (see Section 4.2.4.12). The gender–number morphomes found in Burushaski (4.2.2.3), Ket (4.2.2.6), and Northern Akhvakh (4.2.2.11) are also the same in that all three merge the singular of some gender with the singular and plural of some other (inanimate) gender.

Table 4.166 Gender–number inflections in two languages

Burmeso Conjugation 1 gender affixes			**Khinalugh Set 2 gender affixes**		
Gender	SG	PL	**Gender**	SG	PL
II Female, animate	g	s	II Female	z	v
III Miscellaneous	g	j	III Animate	v	j
IV Mass nouns	j	j	IV Inanimate	j	j
I Male	j	s	I Male	j	v

A complete abstraction away from the concrete values and categories involved in a morphome will allow us to focus on the geometric patterns exclusively and observe another potential sort of (more abstract) cross-linguistic recurrence. The person–number morphomes presented in Section 4.3.11, for example, are all geometrically the same in that they instantiate 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 (see Figure 4.13).

Thus, if we abstract away from the concrete features and value-sets involved, the SG+3PL morphome, the PL+2SG one, and many others (e.g. Northern Akhvakh's M.SG+F.SG+N) will constitute instantiations of the pattern in Figure 4.13 because they are all merely rotational or row-order variants of each other (see Table 4.167).

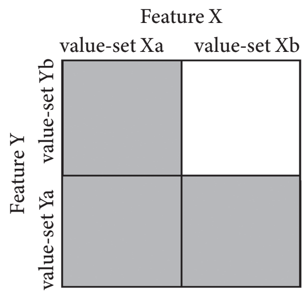

Figure 4.13 Schematic representation of Pattern A

Table 4.167 Some Pattern A type morphomes

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-	M	-ari	-ari

This, which I will call here Pattern A, is by far the most prevalent one in the sample. It is found in a total of 69 different morphomes (57.5%). 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.

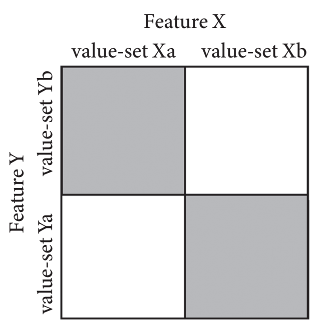

Figure 4.14 Schematic representation of Pattern B

The geometrical pattern in Figure 4.14 is found in a total of 10 morphomes (8.3%) in the present database. This is the case, for example, with the morphomes in Table 4.168.

Table 4.168 Some Pattern B morphomes

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

To present the third most recurrent morphomic pattern in this database, we need to enter the realm of the three-dimensional. In a total of seven morphomes (6.4%), morphological affinities in the paradigm follow the pattern in Figure 4.15. Because of the obvious limitations of a three-dimensional visualization, these patterns will be represented in 2D.

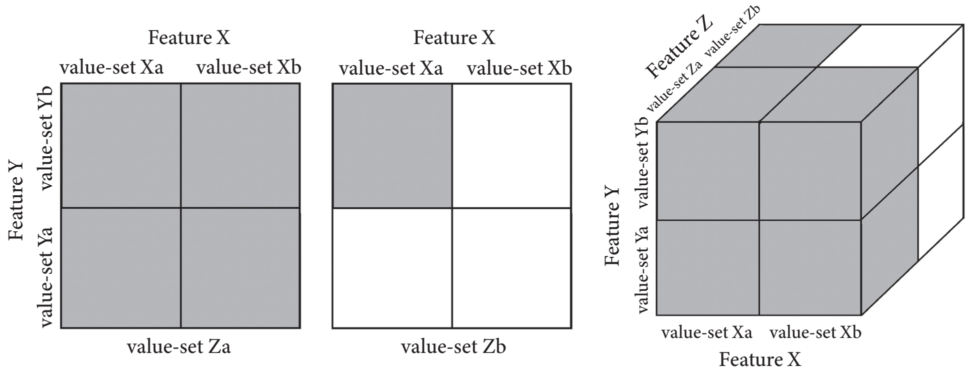

Figure 4.15 Schematic representation of Pattern C (3D left; 2D right)

This abstract morphomic pattern has been found among others in Spanish (IE) and in Udmurt (Uralic). These two morphomes are schematically represented in Table 4.169.

Table 4.169 Two pattern C morphomes

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	–	–

Beyond these most frequent ones, 7 other geometrical patterns (see Figure 4.16) have been found here to be represented by more than one morphome. In addition, another 14 patterns have been found instantiated by just a single morphome, for a total of 21 different geometrical paradigmatic patterns instantiated in this database.

It is very revealing that a single one of them, Pattern A, accounts for 56.7% (N=68) of the morphomes. Another circumstance that may be noted in Figure 4.16 above is that there appears to be a correlation between the **size** of a pattern (as defined by its morphome paradigm size) and its **naturalness** (as defined by morphosyntactic coherence), and a pattern's cross-linguistic recurrence.

The two smallest possible patterns (i.e. those that can be captured in a 2×2 table) are the two most frequent ones, and the next smallest one (2×3

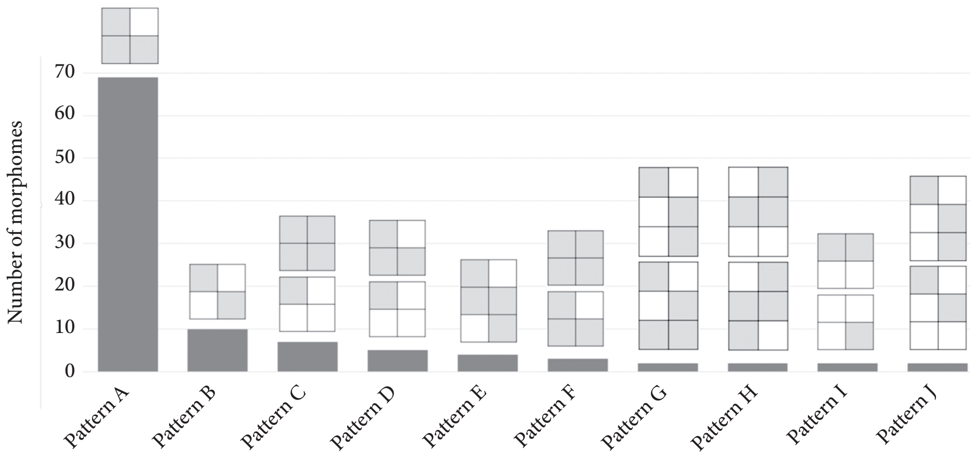

Figure 4.16 Geometrical morphomic patterns and their recurrence

pattern E) is also among the most frequent. The larger geometrical patterns (2×3×2 patterns like G, H, I, and 3×3×2 patterns), by contrast, seem to be predominantly unique.

Naturalness, in turn, can explain asymmetries between patterns of the same size. It would explain, for example, the substantially higher frequency of Pattern A relative to Pattern B. It also seems to be the decisive force driving the cross-linguistic frequency of different 2×2×2 patterns (C, D, F, and I in Figure 4.16), as this (7, 5, 3, 2) mirrors perfectly their relative degrees of morphosyntactic coherence, i.e. naturalness (46.6%, 44.3%, 42.9%, 33.3%).

4.4 Statistical analysis

Examining and discussing the data in each of the surveyed variables is interesting in itself, since it gives us information on the properties of morphomes cross-linguistically. The statistical analysis of their overall properties and correlations promises to be another avenue for empirical discoveries that may shed light on some aspects of morphological architecture and/or linguistic cognition.

A sensible first step to analyse the wealth of data in this morphome database is to assess how similar each of the variables presented in Section 4.3 is to the others with respect to how they classify or order the morphomes. Hierarchical clustering (function hclust in R (R Core Team 2021))[45] reveals the picture in Figure 4.17.

[45] On the basis of Kendall's τ correlation coefficients in absolute numbers between the different variables, a Euclidean distance matrix was obtained. The Ward's method for hierarchical clustering was used on this distance matrix, but the method used did not substantially impact the results.

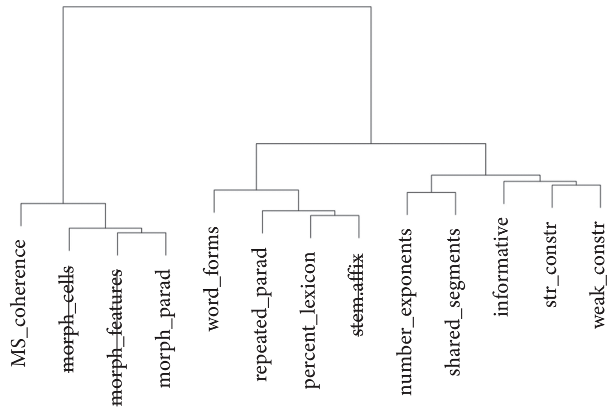

Figure 4.17 Hierarchical clustering of morphomic diversity variables

As already advanced in Section 4.3.9, morphome cells, morphome features, and morphome paradigm size are logically (and empirically) dependent variables. To avoid problems of multicollinearity in statistical analyses, only one of these three closely related variables (the richest one: morphome paradigm size) will be included in subsequent analysis. The stem–affix variable has also been excluded from later analysis due to its close association with others like the percentage of lexicon (Section 4.3.4), and the repetition of a morphome in a single paradigm (Section 4.3.3). I consider stem–affix status to be a theoretical construct based on these very variables, which means this is not something that should be explored on a par with them. Although they also promise to throw light onto the processes that shape morphomes, issues concerning cross-linguistic recurrence (Sections 4.3.11 and 4.3.12) or historical origin (Section 3.1) have ontologically no place either among the structural traits of morphomes that could possibly be relevant to morphological architecture.

On the basis of the values that the morphomes in this database take in the remaining variables (those described in Sections 4.3.1 to 4.3.9), a Principal Component Analysis (PCA) can help us observe whether morphomes tend to cluster in internally comparatively homogeneous groups, and also how representative the most commonly researched morphomes of Romance are of the phenomenon as a whole.

As Figure 4.18 shows, Romance morphomes cover only a reduced area within the overall design space of morphomes cross-linguistically (consider e.g. that they are all lexically restricted and not repeated within the paradigm). This strongly suggests that exploring Romance morphomes exclusively is not enough to understand the phenomenon in its entirety, which confirms that the database this book presents, and its overall line of research, were urgently needed.

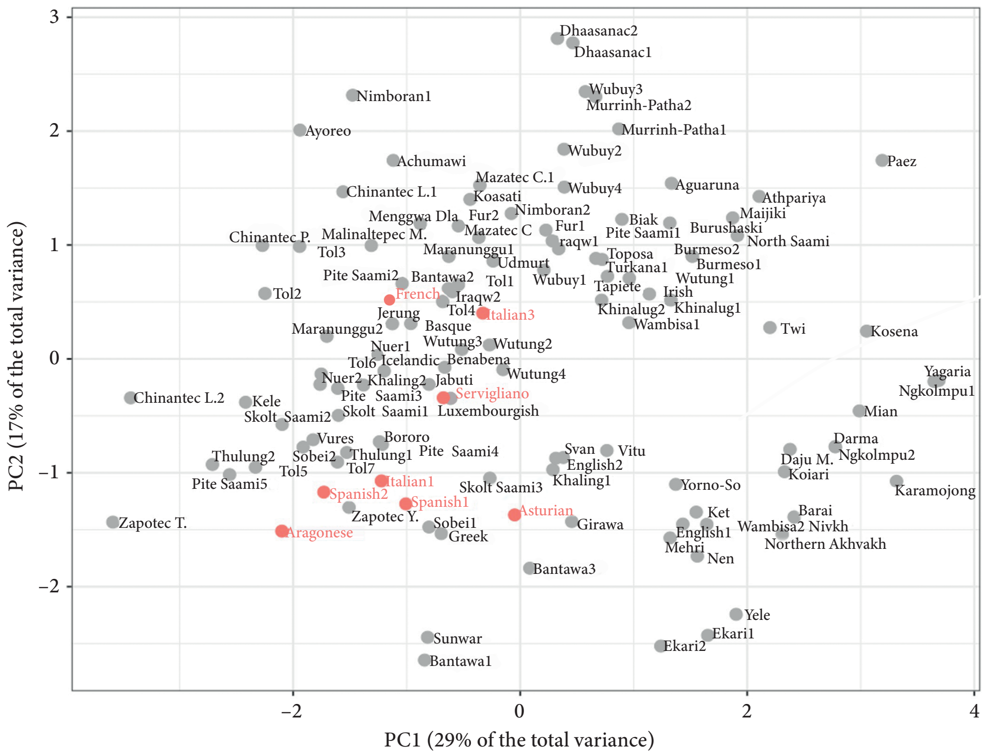

Figure 4.18 PCA of all morphomes, with those of Romance in red

Related to this, and maybe unsurprisingly, it has been found that morphomes from the same language, cognate morphomes (see Figure 4.19), and (to a lesser extent) morphomes from the same family and geographical macro-area, tend to be more similar to each other than morphomes in genetically and geographically unrelated languages. This probably holds for most linguistic phenomena and traits with some level or heritability and/or horizontal transmissibility. Morphomes, whose heritability has been well documented in Romance, are not an exception. The present set of data confirms this property is not limited to this language family. Even though cognate morphomes were only included in the database when they differed in their paradigmatic distribution, these (see e.g. Kele and Vures, Skolt Saami2 and Pite Saami3, Koiari and Barai, Kosena and Yagaria, Aguaruna and Wambisa1, Nen and Ngkolmpu2) still preserve a high degree of overall similarity across all variables.

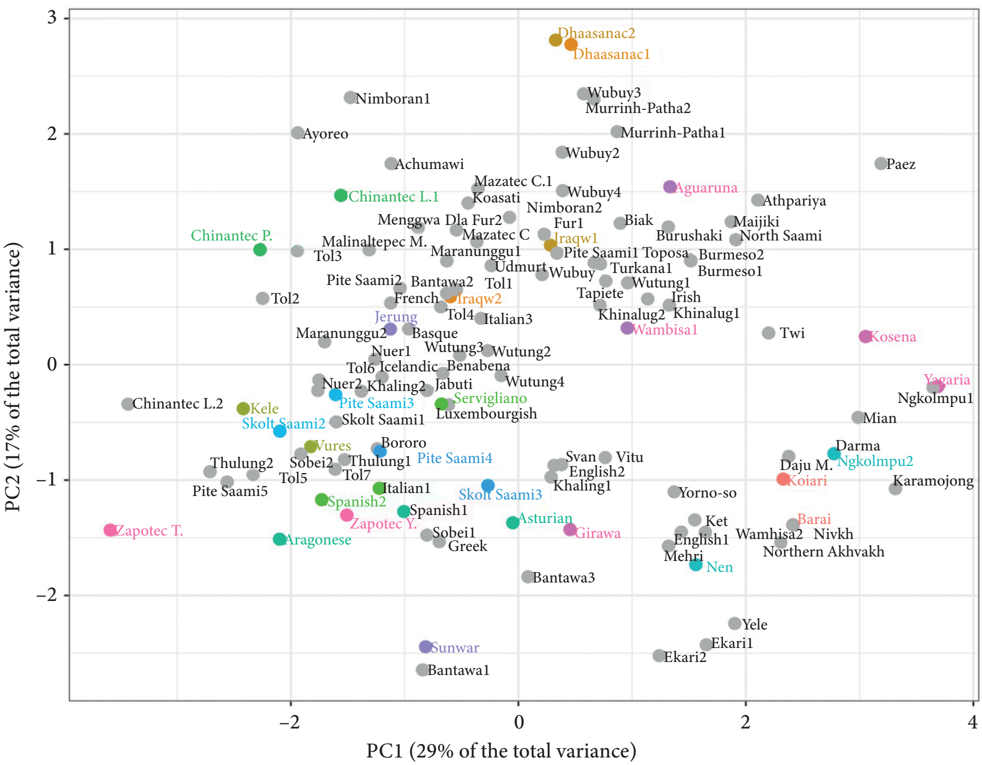

Figure 4.19 Overall similarity of cognate morphomes (same colour, grey = noncognate)

Figure 4.20 can help us understand the correlations between the different variables. It shows significant correlations above the diagonal (according to Kendall's Tau statistic)[46] and their P-values below the diagonal. These correlations will be discussed further in the remainder of this section.

−0.227 Number of different word forms and informativity. There is a highly significant (P<0.01) inverse correlation between these two variables by which a morphome that spans a greater number of different word forms tends to be associated with lower levels of informativity. This might be one of the main candidates for an empirical difference between morphemes and morphomes.

[46] This a non-parametric statistic that measures the similarity of the data ranking by different variables. It takes numbers between −1 and +1, with numbers close to zero suggesting no correlation, and numbers close to −1 or +1 a strong correlation.

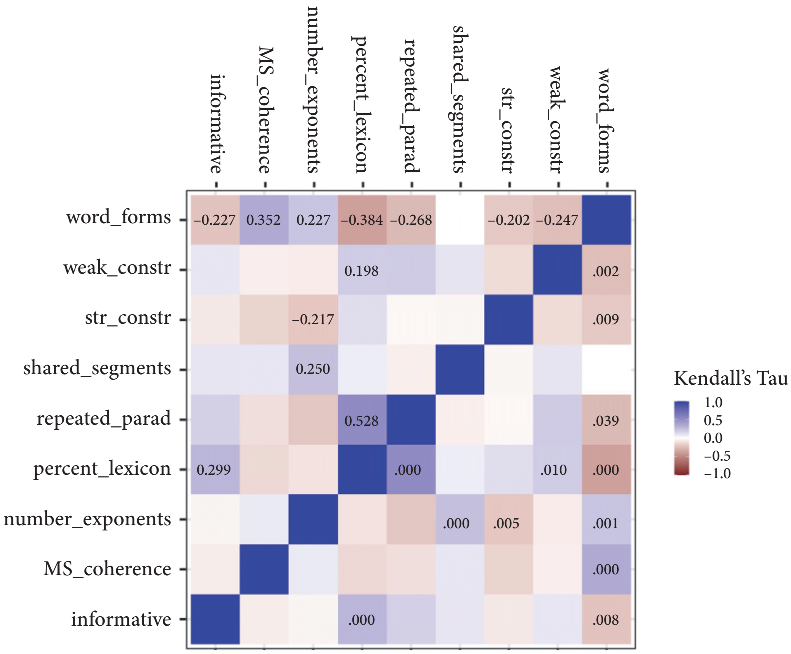

Figure 4.20 Kendall Tau's correlation coefficients between the variables

Consider the paradigms in Table 4.170. It is very common (one could even say that 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, in the simplest possible way, 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, five suffixes are deployed to produce eight different word forms.

Table 4.170 Partial paradigm of Georgian 'fly' (Aronson 1991: 228–32)

	Georgian morphemic system		Pseudo-Georgian morphomic 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

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 (see Pseudo-Georgian in Table 4.170) without giving up the economy that comes from orthogonality.

The data gathered here, however, suggest a tendency for morphomes not to be orthogonal to other morphological elements in the language. Although exceptions can be found (see Tol in Section 4.2.5.14, and Yagaria in Section 4.2.4.21), systems like the Pseudo-Georgian one above are extremely 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 therefore 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, morphological 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 rendered the suffix uninformative (i.e. redundant) as far as the morphological contrasts in the language are concerned. Because morphomes, by definition, cannot rely on semantic/syntactic values for their distribution, they can only reduce or keep in check their distributional complexity by recourse to other morphological cues. Orthogonality to other form(ative)s prevents this, and may thus be dispreferred in morphomes, but not in morphemes, which make use instead of the niches provided by feature values.

0.352 Number of distinct word forms and MS coherence. There is a highly statistically significant ($P<0.001$) positive correlation between the number of different word forms that a morphome spans and that morphome's morphosyntactic coherence. The greater the number of word forms, thus, the more the morphome's morphosyntactic distribution tends towards naturalness. This seems an understandable correlation, since, if a morphome spans many different word forms, a rationale of some sort seems to be most useful to keep the overall distributional complexity in check.

Table 4.171 Distribution of diphthongization in Spanish N-morphome verbs

	PRS.IND	PRS.SBJV	IPF	PAST	IPF.SBJV I	IPF.SBJV 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 4.172 Distribution of diphthongization in randomized Pseudo-Spanish

	PRS.IND	PRS.SBJV	IPF	PAST	IPF.SBJV I	IPF.SBJV 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

Even in morphomes, this rationale can be partially semantic. Consider the paradigms in Tables 4.171 and 4.172. As evidenced in the former, the Romance N-morphome is far from being semantically random. The word forms it spans all share a value PRS, and in addition are expressible quite succinctly as the SG and/or 3 cells within that domain. This must surely aid the functionality and learnability of a pattern. A comparison to the more unstructured hypothetical morphome in Table 4.172 shows just how rare true morphosyntactic incoherency really is, even among morphomes. The reason for this must be related to both diachronic and learnability constraints (see the discussion in Section 4.5.1). The need for structure becomes greater, of course, the larger the number of word forms or contexts a morphome spans.

0.227 Number of exponents and number of word forms. A significant positive correlation has been found between the number of different word forms a morphome spans and that morphome's allomorphic diversity. This correlation will be discussed along with the next one.

–0.384 Number of word forms and percentage of the lexicon. A significant negative correlation exists between the number of word forms a morphome spans and its recurrence in the lexicon. This correlation and the previous one must be related to a tradeoff between lexical and grammatical informativity (see Section 4.5.2), and could also be explained diachronically by looking at the properties of morphomes derived from sound changes. These tend to originate in lexical rather than inflectional material. Stems must evidently be highly morphologically diverse across lexemes for them to perform their communicative roles (i.e. transferring lexical meaning). When sound changes generate stem alternations, these alternants will (i) tend to be morphologically highly diverse, (ii) tend to span over many distinct word forms (since lexical stems tend to be identical across the paradigm), and (iii) may be restricted in their application across the lexicon. In fact we find important differences in this respect between the properties of morphomes generated by sound changes and others (see Figure 4.21).

–0.268 Number of word forms and repetition within a single paradigm. A correlation has been found whereby morphomes which span more different word

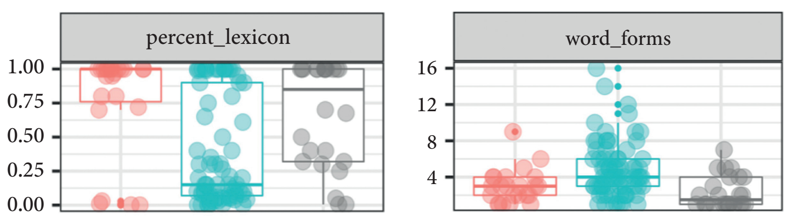

Figure 4.21 Lexical (left) and word-form (right) recurrence of sound-change-generated morphomes (blue) relative to those where sound change was not involved (red) and those whose origin is unknown (grey)

forms tend not to be repeated within a single paradigm. This points in the same direction as other correlations that suggest morphs vary in their informativity only within a narrow range (see also the notion of 'Morphological Equilibrium' e.g. in Milizia 2015), so that a tradeoff exists between providing grammatical or lexical information (see the discussion in Section 4.5.1).

–0.202 and –0.247 MS constraints and number of word forms. There is a significant inverse correlation between the number of constraints that a morphome is subject to (of both strong and weak type) and the number of different word forms that the morphome spans. This makes sense since the less constrained a formative is within the paradigm, the greater the chances are that it cross-cuts or constitutes a superset of the distribution of another (compare the N-morphome in Asturian and Spanish in Section 4.2.3.13).

0.198 Weak MS constraints and recurrence in the lexicon. The correlation between more MS constraint 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, again, due to the tradeoff between a formative's spread in the paradigm (what is measured, really, by MS constraints) and its spread in the lexicon.

–0.271 Strong constraints and number of exponents. Like the correlation above, negative correlation holds for both strong and weak constraints, and for number of exponents, although this is only significant for strong constraints. The reason why more constrained morphomes appear to have less morphological diversity must be related to the grammatical–lexical tradeoff mentioned previously and discussed in Section 4.5.2 (see also Figure 4.23).

0.250 Number of different exponents and average number of segments. A significant positive correlation has been found between the number of different allomorphs of a morphome and their average length in segments. This finding might not have been necessarily anticipated, and it is not clear why this correlation should exist. I will advance here one hypothesis: that some of the morphomes that are less morphologically diverse (i.e. repeated with only two or three different allomorphs) might be 'spurious' in that some might constitute accidental homophonies rather than synchronically relevant grammatical categories. Shorter exponents (e.g. a single vowel) are more likely to be formally identical by accident. The requirement for an identity to be repeated with exactly the same paradigmatic extent with two different formatives (see Section 4.1.2) is intended to make it much less likely for spurious morphological identities to make it into the database. However, this risk is probably still not zero, and may only become progressively reduced as the number of allomorphs increases.

A closely related way of explaining this correlation might be to say that those morphomes that have a higher number of morphological realizations are more likely to be learned as grammatical abstractions truly independently of their concrete formal exponents (much as morphomes are usually formalized, with phonologically 'blind' syncretic indexes like λ in Section 2.11). It may be, therefore, that only these more robust and morphologically diverse morphomes are independent of their phonological instantiations and thus productive in the strictest sense of the word, i.e. in a way allowing them to motivate new (e.g. longer and suppletive) alternations.

0.528 Recurrence in lexicon and in paradigm. There is a highly significant correlation between a morphome's ability to recur within the paradigm of a single lexical item and its generality across the lexicon.

0.299 Recurrence in the lexicon and informativity. At the same time, there is a significant correlation between the lexical recurrence of a morphome and its informativity (as defined in Section 4.3.7). Informative morphomes (i.e. those which discriminate between different word forms) might be preferred because of their greater usefulness/functionality. Morphomes that recur in a paradigm might also contribute more (grammatical) meaning, while at the same time being more salient due to the greater combined token frequency of the different instantiations of the morphome within a single paradigm.

4.5 Discussion

4.5.1 A naturalness bias in morphomes?

Naturalness, as shown in Section 2.2, and as measured by morphosyntactic coherence (see section 4.3.8), is a gradient property. As shown in Figures 4.10 and 4.16, morphomes that adopt comparatively less unnatural distributions seem to be more frequent. These (e.g. geometrically contiguous) patterns must either emerge more frequently during language change or enjoy a greater learnability and diachronic stability once they arise (see Pertsova 2011 and Saldana et al. 2022).

That very unnatural paradigmatic distributions must arise less frequently than more natural ones is evidently the case when a pattern arises through analogical processes. As will be shown in Section 5.1, features and values are important structuring forces in grammar and in the paradigm. Thus, when forms are extended analogically, or by way of secondary grammaticalization processes, to other values/paradigm cells (see Table 4.173), the source and the target are usually adjacent in terms of feature values.

In Biak, as explained in Section 4.2.4.3, what must have been initially trial forms have been extended to cover larger numbers. A run-of-the-mill morphosyntactically driven analogical process (i.e. an incipient consolidation of the trial and plural numbers) has resulted in Biak in a morphomic pattern. Similarly, in Basque, through a politeness-driven process like the one that caused the loss of *thou* in English, an originally 2PL form was extended to the 2SG, which resulted in an unmotivated affinity between the 2SG and PL.

Table 4.173 Two geometrically contiguous morphomes

	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

In these processes, language users deployed forms in contexts where they could not be used before. However, speakers are obviously sensitive to the former 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 with respect to their value(s) and paradigmatically 'contiguous'. Non-contiguous, very unnatural morphomes can thus usually only emerge analogically by way of an intermediate contiguous-morphome stage (see e.g. the diachrony of the morphome in Twi proposed in Section 4.2.1.8).

It is my contention, however, that, even in the case of morphomes emerging in a seemingly more accidental manner (e.g. from the morphologization of sound changes), contiguity may be more frequent an outcome than would be expected from chance in a naïve way. After all, the forms in the paradigms where the sound changes apply are far from being completely random. They are riddled 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 morphomes/alternations derived from sound changes cannot be expected to be completely random regarding their paradigmatic distribution, as cells with similar content or frequency will have a higher chance of sharing forms, and hence sound changes as well.

Source constraints are of course likely to represent, ultimately, functional/cognitive preferences towards more natural morphomes. Natural classes (e.g. PL) enjoy a learnability advantage over morphomic classes, and in turn, morphomes with a higher morphosyntactic coherence (e.g. 2PL+3PL+3SG) may be preferred (i.e. might be more learnable and diachronically resilient) compared to less coherent ones (e.g. 2PL+3SG) (see Saldana et al. 2022;). 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 morphological 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 than doing so if semantic adjacency does not hold (e.g. DU=PL≠PC or 2PL=3SG≠3PL≠2SG). This could reasonably make more unnatural morphomes comparatively more difficult to learn, and more vulnerable to change (into a contiguous pattern), to disintegration, or to levelling when subsequent sound changes or analogical changes interfere with the paradigmatic distribution of formatives. Some diachronic developments (see Tables 4.174 and 4.175) have been found to support this view.

Table 4.174 Getting rid of a non-contiguous morphome in Nen

	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 4.175 Getting rid of a non-contiguous morphome in Lak

	Proto-Lezgian		Lak (Žirkov 1955)	
	SG	PL	SG	PL
I Male	*w	*b	Ø	b
II Female	*r/j	*b	d	b
III Animate	*b	*d	b	b
IV Inanimate	*d	*d	d	d

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 morphome, however, has been disrupted in Nen. Here, the more natural morphome that extended through 3+2PL has been extended to the 1PL as well, thus breaking the morphomic syncretism of 2SG and 1PL.

Something similar in its result (but different in its implementation) can be found in Lak. In the ancestral language (represented in Table 4.175 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 completely unnatural. 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 a greater degree of naturalness/geometrical contiguity.[47] The incorporation of the animate plural cell

[47] Alternative analyses of this change are also possible, of course (e.g. a more semantically oriented extension of a human-denoting exponent to other animates). See, however, the change from Amele to Girawa (Tables 4.81 and 4.82) for a similar development that is less easily accounted for in the same way.

into the original pattern, therefore, increases the coherence of the forms' morphosyntactic distribution, which may make it a more 'viable' meaning for a lexical entry. This may therefore increase the chances of *b-* being conceived by language users as a unit, i.e. a single prefix, as opposed to two homophonous prefixes.

4.5.2 The nature of the stem–affix distinction

Although the stem or affix status of a morphome have been reported in the database as in the source descriptions, this variable was excluded from statistical analysis because it was deemed a theoretical notion derived, precisely, from some of the descriptive variables analysed in the database. Consider the differences displayed in Figure 4.22.

As some of the plots in Figure 4.22 suggest, several independent variables (e.g. lexical recurrence, informativeness, number of word forms, repetition within the paradigm) seem associated with the analysis of a morphological entity as a stem or an affix. Consider, for example, the status of the segment /g/ as an exponent of the Romance L-morphome.

The formative *-g-*, in alternations like those in Table 4.176, has traditionally been regarded as part of the stem. Thus, morphologists usually say that the Spanish verb 'put' has two different stems: *pon-* and *pong-*. The assignment of this segment to the stem is due not to one but to multiple factors. Very important among these, I believe, is the fact that the segment occurs in a superset of the contexts of

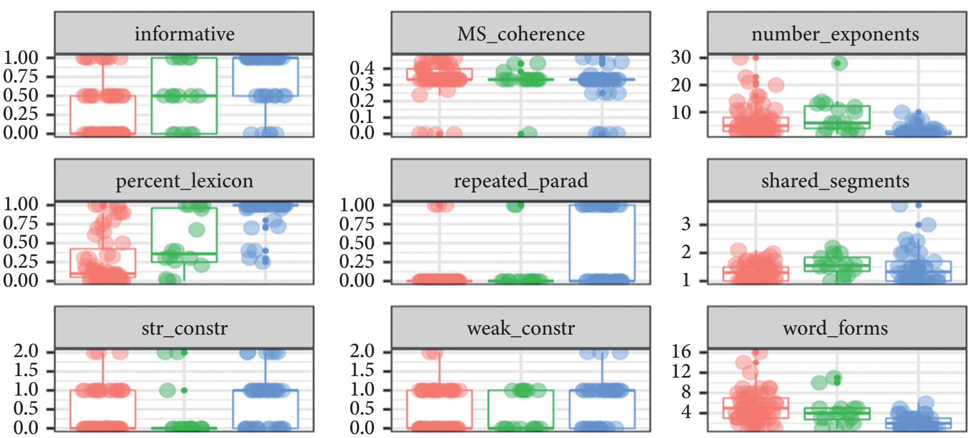

Figure 4.22 Properties of stem-based (red), affix-based (blue), and mixed (green) morphomes

Table 4.176 Present-tense of some Spanish verbs with L-morpheme /g/

	poner 'put'		*caer* 'fall'		*oír* 'hear'	
	IND	SBJV	IND	SBJV	IND	SBJV
1SG	pon-g-o	pon-g-a	cai-g-o	cai-g-a	o[j]-g-o	o[j]-g-a
2SG	pon-es	pon-g-as	ca-es	cai-g-as	o[j]-es	o[j]-g-as
3SG	pon-e	pon-g-a	ca-e	cai-g-a	o[j]-e	o[j]-g-a
1PL	pon-emos	pon-g-amos	ca-emos	cai-g-amos	o-ímos	o[j]-g-amos
2PL	pon-éis	pon-g-áis	ca-éis	cai-g-áis	o-ís	o[j]-g-áis
3PL	pon-en	pon-g-an	ca-en	cai-g-an	o[j]-en	o[j]-g-an

various other finer-grained suffixes (*-o* and *-a*)[48] so that *-g-* occurs redundantly, rather than informatively, and recurs across multiple word forms. The fact that it is found in a small subset of the lexicon, and that it has an unnatural distribution, must also contribute to classify *-g-* as part of the stem, i.e. as lexical in nature rather than inflectional (compare with PRS.SBJV *-a-*, which is usually considered inflectional and/because it occurs in a slightly larger subset of the lexicon with a natural paradigmatic distribution).

This quagmire reminds us of the need to reach unified and well-grounded definitions of the primitive notions in our discipline. An optimal definition for a linguistic concept should, in my opinion, be concise and make reference to as few distinct variables as possible—ideally to just one. The delimitation/definition of stem and affix is particularly troublesome in this respect because it usually (and subjectively, depending to a large extent on the individual linguist) relies on (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 between formatives), and even on (iv) lexical generality (i.e. the number or proportion of lexical items that show the form). This is, obviously, very unfortunate because this definitional interwovenness of logically different variables prevents us from analysing correlations in a meaningful way (see Herce 2019b for a similar point regarding the notion of (ir)regularity in language). This is why stem–affix status was not considered a cross-linguistically valid empirical property of morphomes in this book (see Haspelmath 2010).

Despite this initial precautionary scepticism, it should be stressed that if the correlations between some of these component variables were shown to be stable across languages and across time, these might be good enough reasons to keep notions like 'stem' and 'affix' into our descriptive and analytical toolkits as linguists

[48] Note that, if it were analysed as a suffix, -g- would not abide by the Superset Principle (see Section 2.7)

(see Spike 2020). There are strong pressures that shape the recurrence and distributions of morphs in a language. A tradeoff should necessarily exist, for example, between communicative efficiency (a language has to remain an adequate vehicle for the transmission of information) and learnability (a form in a language cannot be learned if it is hardly present in natural input). From the perspective of information transfer, one might have wished, for example, for a system where every single possible combination of lexeme+morphosyntactic content would be expressed by a different morph. However, even if very precise, a cumulative word meaning 'hesitate.1PL.PST.SBJV' would almost certainly be absent from even very large samples of natural language, and hence would be unlearnable. From the perspective of learnability, the ideal morph is one that is repeated ubiquitously, but a morph which occurs in every single lexeme under every single inflectional value would be completely uninformative and just a burden to the efficient transfer of information.

We could therefore entertain the hypothesis that the (morphological) units in a language will tend to provide a roughly similar amount of information (see the Uniform Information Density principle in Jaeger (2010) and Coupé et al. 2019), which we could arbitrarily set, to facilitate the present discussion, at 10 bits (this equals the information gain in knowing 10 choices between two equiprobable values, or equivalently, one choice among 2^10 = 1,024 different values). Imagining an inflectional system with 1,024 lexemes (e.g. 'work', 'guess') and 1,024 morphosyntactic value specifications (e.g. 1.DU.PST.SBJV.ACT, 3.SG.PRS.IND.PASS), an inflectional system would stick to this strict 10-bit-per-item constraint by having morphs located along the line in Figure 4.23.

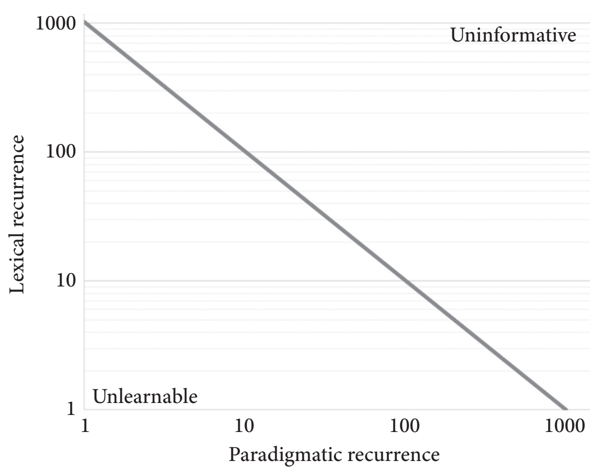

Figure 4.23 Number of cells and lexemes for morphs with information content of 10 bits

Some of the correlations found in Section 4.4. (e.g. the negative correlations between lexical recurrence and number of word forms, and the positive correlation between lexical recurrence and strong and weak morphosyntactic constraints) must derive from this tradeoff between lexical and grammatical informativity. A canonically lexical morph (which we would call a stem) would provide complete information about lexical semantics and identify the lexical value unmistakably (e.g. Spanish *trabaj-* 'work') while failing to provide any semantic or syntactic information (i.e. as a canonical stem, it will appear everywhere in the paradigm, regardless of the associated grammatical value). A canonically grammatical morph (which we would probably call an affix) would provide fine-grained grammatical information (e.g. Spanish *-o* '1SG.PRS.IND'), while failing to provide lexical information.[49]

Many morphs don't fall into either extreme and will provide some amount of both lexical and morphosyntactic information. Towards the latter extreme, a primarily morphosyntactic marker may fail to apply to all lexemes across the board (e.g. Spanish *-é* '1SG.PST.IND' only occurs in first-conjugation verbs), thus providing a bit (exactly one bit if it occurs in 50% of the lexemes) of lexical information too. Closer to the lexical extreme, a polysemous root (or one that is used in multiple lexemes (e.g. Spanish *fui-* 'was'/'went') can restrict the lexical value dramatically without fully specifying a lexical value. Other morphological elements are found between the lexical and the grammatical world. In our tool system of Figure 4.23, this would be a morph that occurs in 32 paradigm cells across 32 lexemes (notice the similarity with some classically morphomic exponents like *-g-* in Table 4.82).

This understandable tradeoff between the lexical and grammatical informativity of morphs might therefore be what makes the descriptive labels 'stem' and 'affix' useful. From the empirical cross-linguistic perspective, it would also be interesting to assess whether the notions of the (canonical) stem and affix are empirically supported abstract categories.[50] Some of the results (Figure 4.6), and correlations (such as between lexical recurrence and the presence of morphosyntactic constraints in Figure 4.20) described here suggest there is a tendency for morph(ome)s to be either eminently grammatical or eminently lexical.

[49] Token frequency can be added as a third crucial dimension: a morph that is both lexically and grammatically highly informative (e.g. English *am*, *is*, *be*) can exist if it expresses frequent lexical and grammatical values. Low token frequency values, by contrast, are associated with more syncretism and separative morphology. The token frequency of different cells and lexical items would have been a highly relevant variable to explore in relation to morphomes in this book. The reason why it has not been included in Section 4.4 is the impossibility of finding representative corpora of most of the languages in the database. Based on our experience with other languages, educated guesses and approximations about relative 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') and have been made (e.g. in Section 4.3.11) because frequency is one of the most powerful sources of explanation in language.

[50] For a similar enquiry about a different morphological hot topic, namely the notion of wordhood within and across languages, see Tallman and Auderset (2022).

Much more work would be needed to replicate this finding and to ascertain the origin of this apparent polarization into lexical vs grammatical informativity (e.g. by looking at the ease of processing and learning of morphs that pack different types of information). Of course, the cognitive biases towards learnability and informativity discussed around Figure 4.23 would manifest (i.e. affect the morphological architecture of natural languages) through language change. Because the morphosyntactic restrictions on a morph, its occurrence in a smaller number of word forms, and its participation in a language's overall system of morphological contrasts all make a morph(ome) more valuable in the transfer of grammatical information, these properties may favour the greater success (e.g. via generalization across inflection classes, or greater resilience) of morp(ome)s with these traits (and even of those parts of morphomes with these traits, see Herce 2021b).

4.6 Conclusion

Chapter 4 has presented a cross-linguistic database, the first of its kind, containing 120 morphomes from 79 genetically and geographically diverse languages. Each of these structures has been described for a dozen different quantitative variables which capture different aspects of each morphome's form and distribution. Information about the morphomes' diachronic origin (sound change, analogical change, etc.) and cross-linguistic recurrence (of various kinds) has also been presented. All this information is freely accessible through the supplementary materials that accompany this book. This database contributes thus to the fields of morphology and typology. Regarding the latter, typological approaches to morphomicity were not only lacking, but were sometimes not even considered possible because of the idiosyncratic nature of the phenomenon. Regarding the former, this study constitutes the first and only lengthy piece of research that deals with morphomic structures beyond Romance, which has nearly monopolized the literature on morphomes to date.

The findings that can be extracted from this database are many and varied, and are by no means limited to those that have been specifically flagged and discussed in this book. Among these, however, it has been found that morphomes as defined for this synchronic study are present in around 15% of grammatical descriptions.[51] This makes them a relatively infrequent morphological phenomenon, but not so infrequent that they can safely be ignored. The present study has revealed the need to analyse morphomic structures beyond Romance in a similarly detailed way, as these (i.e. N, L/U, PYTA) span only a comparatively small range of

[51] 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 in this assessment, the proportion of languages that display these structures is likely to be somewhat higher than this.

possibilities within the overall variability of morphomes cross-linguistically (see Figure 4.18). This design space is, essentially, unlimited: morphomes can be completely unconstrained in their morphosyntactic distribution, or can be subject instead to multiple 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 analysed (concretely: number of word forms, number of exponents, shared form, and paradigm size) follow Zipfian distributions.

Statistical analysis shows that, although they are logically independent, some of the analysed variables appear to be significantly correlated. The causes of most of these correlations must be many and complex, and each of them could well be the topic of a whole separate book. Some of the most interesting correlations found here involve the greater lexical extension of more morphosyntactically restricted morphomes (which hints at a tradeoff between lexical and grammatical informativity, see Figure 4.23), the tendency of morphomes to not be orthogonal to other formatives (see Table 4.170), and the increased naturalness of morphomes that spread across more cells or word forms (which seems to reflect complexity limitations by which morphomes cannot easily be both 'big' and 'messy', see Tables 4.171 and 4.172).

Although their variety, in terms of features and values involved, is outstanding, this study has also found that some morphomic structures are far from being typologically unique, but are instead found across 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, and SG+1PL. These groupings, and their relative frequency across languages, can contribute to current theoretical discussions, for example on the architecture of person and number (see e.g. Harbour 2019). Surveying morphomic patterns at a more abstract level, it has been found that simpler and more natural patterns (e.g. all of the above) are more frequent. Possible explanations for this bias towards more natural syntactic–semantic structures have been presented throughout this section and throughout the diachronically oriented Chapter 3. In general, the finding reminds us of the power of syntactic and semantic values to influence morphological structure in paradigmatic systems—a fact which has sometimes been downplayed recently (more on this in Section 5.1).

The present findings notwithstanding, there remains much work to be done concerning morphomes and the variables and correlations analysed 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 would help to put the present findings in a broader perspective and to find an answer as to whether or to what extent morphomes 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 2020a). Although it is likely that morphomes will tend to be less informative than morphemes (see Figure 4.9) and will tend to have different diachronic origins (e.g. sound change, see Figure 3.1), my contention is that other empirical differences between morphomes and morphemes that go beyond the definitional ones might be few overall.

5
Implications
Features and forms

This chaper reflects on the importance of morphosyntactic features (Section 5.1) and form-to-form predictive relations (Section 5.2) for the evolution of morphology in the paradigm. Even if discussion has been understandably focused here on unnatural patterns and thus on other sources of external motivation, values and meaning deserve to be pondered against the autonomously morphological templates that constitute the topic of this book.

5.1 The importance of features

Morphological elements, whether stem alternants or affixes, whether morphemes or morphomes, 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/meanings, and often pattern into natural classes. The fact, for example, that the dental suffix *-te* in German conjugation appears in every paradigm cell of the past-tense and nowhere outside the past is probably a 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-ed*) it left the realm of syntax to enter that of morphology but preserved its earlier distribution. Thus, even if the organizing principles of morphology and syntax differed substantially (and if, for example, morphology 'didn't care' about features or values), a great deal of form–meaning correlation might be expected nonetheless in synchrony. If we believe morphology might 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 in accounting for the distributions of morphological elements, not only because of their significant synchronic correlation but also because they play a big role in analogical change. The fact that this book has been devoted to patterns at odds with morphosyntactic values cannot lead us to think that these are irrelevant in morphological architecture.

The Typological Diversity of Morphomes. Borja Herce, Oxford University Press. © Borja Herce (2023).
DOI: 10.1093/oso/9780192864598.003.0005

Table 5.1 Diachronic spread of the formative *-r* in Scandinavian, present-tense

	Old Norse (Rask 1976: 121)		Old Swedish (Noreen 1904: 471–3)		Modern Swedish (Holmes and 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

The present-tense paradigms of 'burn' in several stages of Scandinavian (see Table 5.1) show that, in analogical extension, morphosyntactic feature values (e.g. SG.PRS in Old Swedish or PRS in Modern Swedish) often act as niches (Gause 1934; Aronoff 2016) where a single form may come to predominate. Morphosyntactic and semantic values, therefore, often constrain/drive the expansion of formatives to new environments.

Feature values are therefore assumed to be important in morphology 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).

The shaded change in *blōta* (see Table 5.2) 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 that a syncretism of 2PL and 3PL that had resulted from regular sound change (3PL **-anþ>-aþ*) was extended in English to the remaining cell of the plural. Feature values like 'singular' or 'plural', therefore, constitute grammatical templates of the utmost importance. This means that they should be allowed to feature prominently in morphological description, theory, and formalization. A particularly striking example of how feature values can act as niches in morphological change can be found in Yakkha (Kiranti).

Table 5.2 Two similar morphological changes in two Germanic verbs

	blōta 'sacrifice' (Wurzel 1980: 451–2)				*beran* '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	bløt	blótum	bere	*berams	bere	beraþ
2	bløtiR	blōteð	bløtr	blóteþ	birest	beraþ	birest	beraþ
3	bløtiR	blōta(n)	bløtr	blóta	bireþ	beraþ	bireþ	beraþ

Table 5.3 Yakkha agreement suffixes, partial paradigm (Schackow 2016: 223)

	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

In a relatively brief period of time, quite dramatic changes seem to have taken place in the agreement patterns of the language (see Table 5.3) whereby many morphological distinctions and forms have disappeared (see also Lynch 2000: 91–5 for a similar example in Anejom, Oceanic). Syncretization has not been random. The resulting paradigm is one where, unlike in the earlier system, there are robust one-to-one form–meaning relations.

Diachronic changes like these suggest, thus, that morphosyntactic features and values are paramount in morphology. I cannot therefore 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, on the basis of the behaviour of stem alternation patterns in Romance, he argues that morphomic patterns are not dispreferred. This claim may arguably fit the evidence from Romance morphomes, but should be considered incompatible with the paradigmatic changes presented in 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 those in Table 5.4.

Table 5.4 Hypothetical alternative morphological changes

	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	bere	berams	**beraþ	berams
2	bløtiR	blōteð	bløtr	**bløteþ	birest	beraþ	birest	beraþ
3	bløtiR	blōta	bløtr	blóta	bireþ	beraþ	bireþ	beraþ

Probably all linguists would agree that changes like the hypothetical changes in Table 5.4 are less likely than syntactically/semantically motivated ones. Because they play, by definition, 'on the same team' as feature values, natural-class patterns always have an advantage over morphomic patterns in possessing this source of external motivation that morphomes lack by definition. There is abundant experimental evidence (e.g. Kirby et al. 2008; Silvey et al. 2015) that, when morphological distinctions are lost (e.g. during the iterated learning of an artificial language), the conflation of values is highly structured, and largely follows the tendency documented in natural languages like Yakkha in Table 5.3.

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 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: 10, 685). Although one might have wished to evaluate preferredness on two patterns in the same language, one natural and one morphomic, matched for every single other property, this is not possible. The experimental findings reported above, as well as the diachronic analogical changes discussed in this section, are difficult to reconcile with a theory of grammar where morphomic patterns are not dispreferred to some extent.[1]

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 3.2.4.1). It is interesting, for example, that palatalization before front vowels produced stem alternations only in those conjugations where the resulting pattern was morphomic (e.g. Spanish *hacer*, *decir*). By contrast, the alternations are not found in the productive conjugation (e.g. *pagar*, *colgar*), precisely where they would have resulted in a stem alternant isomorphic with a natural class.

Forms in Table 5.5 preceded by the asterisk contain the velar in the modern languages (e.g. Spanish *pa[g]e*). The sound change was thus either turned back

Table 5.5 Expected paradigmatic results of velar palatalization in Romance

	hacer 'do'		*pagar* '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

[1] A few cases have been presented throughout this book (Biak, Basque, Occitan, Slovene) of morphosyntactically motivated changes that gave rise to morphomic patterns. The preference for morphosyntactically motivated morphological extensions could thus be argued to be a more localized constraint on **change** independent of the naturalness of the more general pattern to which the change gives rise as a result. A bias towards morphosyntactically motivated changes without a similar bias towards morphosyntactically motivated patterns seems to me, however, unlikely.

analogically or resisted ab initio in these first-conjugation verbs, precisely those, 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 maybe even for a bias against natural classes. However, there are many and important confounding factors. First, the stem alternant *hag-* represents a greater proportion of the total use tokens of the verb compared to hypothetical **pac-* (see Table 2.70). Other asymmetries could have also favoured the alternation in *hacer*. For example, in the conjugations where it happened, the sound change affected the majority of the forms in the paradigm, whereas in the first conjugation it would only have affected a minority. Other confounding factors could have been that maybe too few verbs ended in the 'right' consonants in the *-ar* conjugation, or maybe the high token frequency of a few /k/-final verbs like *decir* 'say' or *hacer* 'do' could have favoured stem alternation in verbs from their same conjugations ... All these heterogeneous factors might have plausibly favoured the stem alternation pattern that survived even 'compensating' for its unnaturalness. With a single example (or with a few related examples from a single family), there is simply no way to tell. This is the reason why a cross-linguistic approach to the morphome was urgently needed.

Features and their values, cross-linguistic evidence suggests, are paramount in morphological structure. This 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. Morphomes show clearly, indeed, that 'the impulse toward greater isomorphism is not an irresistible one' (Stump 2015: 268). It has been my goal in this book 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.

5.2 The importance of form

Morphology (i.e. the internal structure of words and paradigms) is, as I argued in Section 5.1, 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 the Turkish case–number inflection (Kornfilt 2013), 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 syntactic and semantic values (consider the discussion in Section 5.1 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 Old High German) that had no number-dedicated morphology whatsoever). That morphology is often about conveying meaning (i.e. values and categories) is also hardly new or surprising considering the communicative needs that language as a whole has to serve.

As this monograph and others have 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) can result in morphological affinities that do not match semantic 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, morphological predictabilities within and across words are registered and actively employed by speakers to cover the gaps that a Zipfian input does not fill. This leads to morphologically driven analogies that perpetuate or reinforce the paradigmatic results of former historical accidents, or even create new categories (see sections on formally motivated analogy and pattern interactions in Section 3.1) based on more or less accidental morphological affinities.

Similarity and covariation in morphological 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, morphological implicational patterns tend to be reinforced at both the paradigmatic and the syntagmatic levels.

The Old Norse verbal inflectional system, for example (see Table 5.6), led one to expect a morphological identity between the infinitive and the 3PL present indicative forms. For the vast majority of verbs, thus, one could correctly predict the infinitive (e.g. *fara*) from 3PL (also *fara*) and vice versa. This vast generalization was perceived by language users, who thus had the capacity to overgeneralize this

Table 5.6 Some predictability-driven morphological changes

Old Norse			Spanish			
Lexeme	INF	3PL.PRS.IND	**Cell**	**Stem**	**Suffix1**	**Suffix2**
'drive'	*fara*	*fara*	3PL.PRET	*pus-*	*-ie-*	*-ron*
'must'	*skulu*	*skulu*	1SG.IPF.SBJV	*pus-*	*-ie-*	*-ra*
'owe'1	*eiga*	*eigu*	GER1	*pon-*	*-ie-*	*-ndo*
'owe'2	*eiga*	*eiga*	GER2	*pus-*	*-ie-*	*-ndo*

rule (see the change from owe1 (Old Norse) to owe2 (Icelandic)) whenever an exception was not successfully acquired from the input.

Implicative relations can 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 (exactly in 13 out of 14 word forms). This implicative pattern is perceived by language users, who may then strengthen it further when they occasionally extend it to the one context where the rule did not apply originally (consider the change from Gerund1 to Gerund2 in some Spanish varieties, see Pato and O'Neill 2013).

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 time interpreted and explained exclusively with reference to morphosyntactic features and their interaction. Morphological identities that correlate well with morphosyntactic values are deemed to be significant, while those which do not are either straitjacketed into better behaviour (e.g. by underspecification and blocking) or dismissed as 'accidental homophonies'. Yet there is abundant evidence that morphological differences do not always correspond to differences in semantic values (e.g. inflection class distinctions, overabundance) and conversely, that differences in semantic values do not always align with morphological differences (e.g. syncretism, deponency). These are examples of structures that exist at odds with meaning and values, which undermines 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 morphological similarity and knitting it into grammatical structure will surely be facilitated by the existence of some overarching meaning or morphosyntactic affinity, as this provides 'extra evidence' for the importance of the morphological pattern and for predicting its distribution. However, doing the same thing with semantically unrelated forms is likely to optimize cognitive resources too, and allow language users to solve the paradigm cell-filling problem (Ackerman et al. 2009).

This is shown quite nicely in the examples in Table 5.7. 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

Table 5.7 Partial paradigms of four Spanish verbs

	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

happens, it resembles the 3SBJV (/naθk/-), rather than the 3IND (/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 alternations. In weather verbs like *tronar* 'thunder' or *nevar* 'snow', only 3SG and nonfinite forms are regularly present in natural speech. These forms, however, are enough to establish whether alternation (compare infinitive *nevar* to 3SG *nieva*) is present in a 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 semantic values.

It seems, therefore, that morphological entities and productive implicative patterns do not need to have a morphosyntactically coherent description. Morphological affinities alone can also prompt language users to construct grammatical categories like the morphomes in Table 5.7. As expressed by Hockett (1987: 88), and as I quoted him in the introductory Chapter 1, sometimes 'it is the resonances that induce the grammatical structure' rather than the other way around.

6
Conclusions

This monograph has been the first to approach the concept of the morphome from a typological and cross-linguistic perspective. Chapter 1 briefly presented the phenomenon and our knowledge of it, clarified the terminology, and presented the overall goals of this book. To make the morphome a workable concept suited for a typological investigation, Chapter 2 dealt with definitional and diagnostic issues: how to distinguish morphomes from accidental homophonies, how to define an unnatural class, what is the role attributed to blocking or zeroes ... as well as other issues that may arise when deciding on the morphomicity of some structure: segmentability, the intra- or extra-paradigmatic domain of a pattern, its cross-linguistic recurrence, economy, independence from phonology, etc. The overall *modus operandi* has been to set a high bar for unnaturalness and systematicity, at the same time avoiding reference to meta-empirical factors (e.g. theoretical analyses and controversial processes, and units like blocking or zeroes) in the present definition of the phenomenon, thus remaining as close as possible to the surface data.

The diachronically oriented Chapter 3 explored the different ways in which morphomes can arise, change, and disappear from a language. **Sound change was found to be the most frequent source of morphomes**, at least of the kind analysed in this book. Sound change, however, has been found to be an internally heterogeneous route to morphomehood, as the locus and result of sound changes can differ in nontrivial ways. Another finding of Chapter 3 is that not only sound change but also every other process that can possibly result in a change to the forms in a paradigm (e.g. **grammaticalization, analogy, pattern interactions, maybe even borrowing) may become a source of morphomes under the right conditions.**

Chapter 4 constituted the core of the book. It presented a multivariate typological deconstruction of cross-morphomic variation. Morphomes in different languages have been found to vary along several different dimensions, among others their degree of unnaturalness (a.k.a. Morphosyntactic Coherence), their number of exponents, their generality across the lexicon, the number of word forms they span, and how informative they are. A synchronic database was presented (Section 4.2) where 120 morphomes from languages all over the world have been painstakingly described, presented in their comparative and diachronic context when possible, and quantified for the above-mentioned variables. An exploration of the data (Section 4.3) and variable correlations (Section 4.4) followed. Some of the most interesting findings are the cross-linguistic recurrence of some

The Typological Diversity of Morphomes. Borja Herce, Oxford University Press. © Borja Herce (2023).
DOI: 10.1093/oso/9780192864598.003.0006

person–number morphomic patterns, and the prevalence of low-unnaturalness morphomes 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 morphomes (which points to a tradeoff between lexical and grammatical informativity), the greater structuredness (i.e. near-naturalness) of larger morphomes (which points to some upper limit to how complex a morphome can get), and the tendency for morphomes not to be orthogonal to other formatives within the paradigm (which suggests a preference for a morphology-based rationale of some sort to their distribution).

Elaborating on the findings of Chapter 4, it has been found that, even when setting a high bar for morphomicity, **morphomes are present across the world's languages**. They have been found here in as many as 37 genetically independent stocks both large (e.g. Austronesian, Indo-European, Otomanguean, Sino-Tibetan) and small (e.g. isolates like Basque, Burmeso, Nivkh, and Páez). This suggests that the phenomenon cannot be dismissed lightly as an accidental quirk of a few languages (e.g. Romance), and has to be explored in detail. It deserves, therefore, the systematic cross-linguistic treatment that it has been missing so far.

Previous morphomic literature has highlighted **the importance of morphological predictability relations within the paradigm**, which seem to constitute the synchronic *raison d'être* of morphomes, as well as the source of their purported diachronic resilience and productivity. This has received additional confirmation in this monograph (see e.g. Section 5.2). 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 largely complete productive system on the basis of sparse incomplete evidence. Because of this, as previous literature has found (e.g. Maiden 2018b), pre-existing forms can serve as templates for the distribution of new formatives. This book has provided many clear examples (beyond the Romance ones most often discussed, see e.g. the discussion on Luxembourgish (Tables 4.3 and 4.4), Yakkha (Tables 4.7 and 4.8), and Svan (Section 4.2.2.13)) of the power of forms to act as niches or templates for other forms. Morphomes and unnatural implicative patterns, therefore, can constitute productive grammatical categories and steer morphological change.

That predictability must lie at the core of morphomes is thus clear. There is, however, a fundamental fact that morphomic literature has not engaged with so far, which is that predictability relations also exist outside morphomes/morphemes, i.e. in the absence of morphological identity. As shown by Herce (2020b), for example, the +g stem-augment in the L-morphome 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. 3 PL.FUT *ven-dr-emos*) and vice versa.

This perfect predictability has emerged analogically, and so it appears that systematic differences can also steer morphological change. It remains to be investigated in more detail whether predictable identities (i.e. morphomes/morphemes) and predictable contrasts are different in any empirically meaningful way.

Another property that previous morphomic literature has usually ascribed to morphomes is 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 get rid of them (by means of analogical changes) within a few generations. As far as I can tell, the identification of resilience as a characteristic property of morphomes had been based to date exclusively on the evidence of Romance, which is unfortunate given that, as shown in Figure 4.18, N, L, U, and PYTA are not representative of the phenomenon as a whole. This book has confirmed that resilience (over at least two millennia) is not a parochial feature of Romance morphomes. Comparable evidence has been found in research in various other language families, most notably East Kiranti (see Sections 4.2.2.1 and 4.2.2.2 and Herce 2021a), Saami (see Sections 4.2.3.10 and 4.2.3.11 and Herce 2020a), Chinantec (see Sections 4.2.5.5 and 4.2.5.6), and Nakh-Daghestanian (see Table 4.175).[1]

Morphomes are defined as systematic morphological identities that do not map onto syntactic or semantic natural classes. The present cross-linguistic research has shown that, beyond this definitionally shared property, morphomes can differ dramatically in most respects: in their syntagmatic location (in prefixes, stems, or suffixes), their morphological diversity (i.e. number of allomorphs), their confinement 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' and naturalness within the paradigm, etc. **This monograph has identified what exactly those dimensions are along which morphomes may be different, and has proposed**

[1] These findings are subject to some caveats and limitations. On the one hand, one has to take the so-called survivor(ship) bias into account (see e.g. Mangel and Samaniego 1984). Since this book focuses on robust existing morphomic patterns, and discusses only reconstructable diachronic trajectories, unstable morphomes and their characteristics must necessarily be underrepresented. Thus, whereas the evidence from Saami or East Kiranti has been extensively discussed in this book, the patterns in closely related Finnic and West Kiranti have barely been explored. The morphological affinities in the latter families, in contrast to the former, show a very notable variability from one language to another. This 'mess' invites less comparative and diachronic work in general, but must be associated with the greater instability of (some of) those morphomic patterns. A second caveat with respect to the diachronic stability of morphomes is more ontological in nature. Even looking at the patterns that did manage to survive more or less unchanged in a language or language family, it is difficult to say whether they are resilient. Stability and resilience are relative, not absolute concepts. Two millennia may be long in human timescales, but not in biological evolutionary timescales. The evolution of human language is likely to fall somewhere in between. An assessment of whether morphomes are resilient or not should involve a comparison with other linguistic traits, such as the lifespan of morphemes, ergativity, the phoneme /x/, and lexical roots. Future research could thus be aimed at systematically assessing the relative stability of morphomes compared to other traits in language (see e.g. Greenhill et al. 2017 for phylogenetic work in this spirit).

novel ways to operationalize and measure this variation in the most fine-grained way possible. Adopting methodologies like Canonical or Multivariate Typology, wide typological and comparative research is possible even on such idiosyncratic entities as morphomes. As already hinted at in Section 4.4, the variables surveyed in this book do not exhaust all variation. Among missing aspects, the token frequency of a morphome (e.g. operationalized as the combined usage frequency of a morphome's cells as a proportion of the total frequency of the lexeme) is likely to be a factor of the utmost importance but could not be included in this database, simply because frequency data is hard to find, being hardly ever reported in grammatical descriptions.

After assembling a large enough sample of morphomes and finding ways of measuring different logically independent aspects about their form and distribution, we now have a better understanding of what morphomes tend to be like (see the general properties of morphomes in Section 4.3) and, by way of statistical analysis, what logically independent properties tend to occur together (see Section 4.4). This can provide insights into linguistic cognition and the properties of morphological architecture. Future research could seek experimental confirmation for the observations derived from Sections 4.3 and 4.4, for example, for the equi-informativity hypothesis ventured in Figure 4.23, or the preference for morphological elements to pack either all-lexical or all-grammatical information.

This research has thus spotted generalizations and ventured biases and diachronic pressures that might shape the synchronic properties of morphomes. This contrasts with most of the extant literature, which has tended to regard morphomes as accidental, unique, idiosyncratic structures that, because of their very nature, are largely incompatible with the extraction of meaningful cross-linguistic generalizations. Here it has been found, that, quite on the contrary, various regularities can be observed. In the domain of person–number agreement, for example, **some unnatural patterns (namely SG+3PL, 1SG+3, 2+1PL, SG+1PL, PL+1SG, PL+2SG, and PL+3SG) have been found to be recurrent** and are instantiated by three or more unrelated morphomes each. A cogent explanation of why these particular morphomes are more frequent than other logically possible combinations (e.g. 2+3SG, SG+2PL, 3SG+1PL) 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 3.1.1.3) how vague accidental splits between marked and less marked/zero values are often transformed by sound change into more robust morphomic splits. The token frequency of different values may also favour morphomic patterns where deviations from naturalness occur in more frequent cells (e.g. PL+3SG rather than SG+2PL).

Factors like these explain why some unnatural paradigmatic distributions are more frequent than others. Together with a naturalness bias, they also explain why **morphomes tend to span a geometrically contiguous (i.e. comparatively more natural) set of cells** (see all the recurrent person–number patterns above) rather than a discontinuous paradigmatic space (e.g. 1SG+2SG+3PL, 1PL+3SG). This agrees with some proposed cognitive biases in category learning (e.g. Pertsova 2011; Saldana et al. 2022), which render 'discontinuous' morphological affinities harder to acquire (consider also the so-called *ABA constraints, see e.g. Bobaljik and Sauerland 2018, and morphological models like that of McCreight and Chvany 1991). Due to a bias favouring more natural distributions, these patterns may arise more frequently during language change and/or enjoy a higher degree of stability once they have become established in the language. This demands that the importance of syntactic and semantic structure in morphology be acknowledged (see Section 5.2). Thus, **even in the realm of morphomes, morphosyntactic values and distinctions seem to constitute an important constraining factor**. This is something that other approaches to the phenomenon, with their focus on morphological autonomy, have frequently failed to appreciate.

The findings of this book, both incremental and novel, argue thus in favour of the view that **morphology cannot be reduced to either morphosyntactic values and their expression or to morphological resonances and the abstraction of exclusively morphological implicative patterns**. Both syntactic/semantic (Section 5.1) and morphological (Section 5.2) templates must be allowed to constitute active components of morphological architecture. Furthermore, their relative strength 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, PRS) morphological 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 from of the linguist but also, probably, from 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), learning concrete exponents as expressions of particular values (e.g. DAT, PL) seems to be the easiest analysis. By contrast, in a deeply morphomic system like many of those presented here (e.g. Daasanach, Chinantec, Murrinh-Patha, Ngkolmpu, Saami, Yagaria), autonomously morphological rules, and using forms to predict other forms (see Table 5.7) might be the best available solution. Sometimes, reference to both form and function is necessary to narrow down the paradigmatic distribution of one and the same formative. Consider, for example, the distribution of the Yakkha suffixes *-wa* and *-me* in Tables 3.36 and 3.37. As explained there, reference to the morphomic stem alternation pattern coextensive with them is unavoidable. At the same time, these are still present-tense suffixes, and are consequently found everywhere through the present and nowhere outside the present. Morphology-provided and feature value-provided templates can thus be used for one and the same exponent.

This account of how grammar works (i.e. one with templates provided by form and meaning for morphological elements) is cognitively realistic, and is grounded on abundant evidence on how *Homo sapiens* 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: 224), '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, morphological (i.e. acoustic or visual), along with various sorts of semantic and syntactic information, can all serve as the basis for language users to construct their linguistic categories. It may be the case that some kinds of evidence (e.g. morphosyntactic values like 'speaker', 'plural', or 'past') are more salient than others (e.g. morphological 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 database summary

morphome_ID	str_ constr	weak_ constr	repeated_ parad	number_ exponents	shared_ segments	word_ forms	informative	percent_ lexicon	MS_ coherence
Achumawi	0	1	0	16	1.8	7	0	0.9	0.333
Aguaruna	0	1	0	4	1.5	1	1	1	0.333
Aragonese	1	0	0	2	1.5	9	0	0.01	0.429
Asturian	2	1	0	2	1.5	3	0	0.003	0.333
Athpariya	2	2	0	2	3	3	0.5	1	0.333
Ayoreo	0	0	0	28	1.9	3	0	0.21	0.333
Bantawa1	1	0	1	5	1	8	0	0.055	0.333
Bantawa2	0	1	0	8	1	6	0.5	0.61	0.333
Bantawa3	1	1	1	5	1	5	0	0.11	0.333
Barai	1	0	1	2	1	1	1	1	0.333
Basque	0	0	0	3	2	4	0.5	0.001	0.333
Benabena	0	1	0	2	1	5	0	0.676	0.333
Biak	0	1	0	2	1.5	4	1	1	0.333
Bororo	0	0	0	3	1	4	0	0.35	0.333
Burmeso1	0	1	0	2	1	1	1	1	0.25
Burmeso2	0	1	0	2	1	1	1	1	0.25
Burushaski	0	1	0	3	1.2	1	1	1	0.333

Chinantec L.1	0	2	0	10	1.7	6	0	0.051	0.443
Chinantec L.2	0	0	0	8	1.75	14	0	0.033	0.382
Chinantec P.	0	1	0	11	1.9	8	0	0.042	0.443
Daju M.	1	0	1	5	1.4	1	1	1	0.333
Darma	1	1	1	2	1	1	1	1	0.333
Daasanach1	0	1	0	23	1.2	1	1	1	0.4
Daasanach2	0	0	0	21	2.1	1	1	1	0.397
Ekari1	2	0	1	2	1.5	3	0	1	0.333
Ekari2	2	0	1	2	1.5	5	0	1	0.365
English1	1	0	1	2	1.5	1	1	0.0001	0.333
English2	1	0	0	3	1	1	1	0.0003	0.333
French	0	1	0	7	1.3	6	0.5	0.07	0.333
Fur1	1	0	0	20	1.6	2	0.5	0.99	0.333
Fur2	0	0	0	14	1.5	5	0.5	0.99	0.333
Girawa	1	1	1	6	1.2	4	0	0.3	0.333
Greek	1	0	0	2	1	4	0	0.4	0.333
Icelandic	1	0	0	13	1.6	6	0.5	0.058	0.333
Iraqw1	0	1	0	5	1.2	2	0.5	1	0.466
Iraqw2	0	1	0	6	1.16	4	0	0.9	0.429
Irish	0	0	0	4	1.6	1	1	0.4	0

Continued

(*Continued*)

morphome_ID	str_ constr	weak_ constr	repeated_ parad	number_ exponents	shared_ segments	word_ forms	informative	percent_ lexicon	MS_ coherence
Italian1	1	0	0	5	1.4	4	0	0.09	0.4
Italian3	1	1	0	11	1.5	3	0.5	0.01	0.333
Jabuti	0	1	0	2	1	2	0	0.07	0.333
Jerung	0	0	0	7	1.6	5	0.5	0.2	0.333
Karamojong	2	1	1	2	1.5	1	1	1	0.333
Kele	0	0	0	6	1.6	9	0	0.04	0.333
Ket	2	0	0	2	1	1	1	0.7	0.333
Khaling1	1	1	0	4	1	3	0	0.15	0
Khaling2	0	1	0	4	1	5	0	0.07	0.333
Khinalug1	0	1	0	3	1	1	1	0.3	0
Khinalug2	0	1	0	2	1	1	1	0.25	0.25
Koasati	0	1	0	4	2.5	3	0	0.4	0.333
Koiari	1	0	1	2	1.5	2	1	1	0.333
Kosena	1	2	1	2	1.5	2	1	1	0.333
Luxembourgish	1	0	0	13	1.1	2	0.5	0.05	0.333
Maijiki	1	1	0	4	2	1	1	1	0.333
Malinaltepec M.	0	0	0	10	2.2	5	0.5	0.03	0.333
Maranunggu1	0	1	0	4	2	3	0	0.302	0.333
Maranunggu2	0	0	0	6	1.83	5	0	0.264	0.433
Mazatec C.1	0	1	0	13	1.5	2	0.5	0.32	0.333
Mazatec C.2	0	1	0	5	1.8	4	0.5	0.32	0.333
Mehri	2	1	0	3	1	2	0	0.5	0
Menggwa Dla	0	2	0	4	1.75	3	0	0.01	0.443
Mian	1	1	1	3	1.3	1	1	1	0.25

Murrinh-Patha1	0	1	0	12	1.4	2	1	1	0.333
Murrinh-Patha2	0	1	0	14	1.6	3	1	1	0.333
Nen	1	0	1	3	1	3	0.5	1	0.333
Ngkolmpu1	1	1	1	3	1.7	1	1	1	0
Ngkolmpu2	1	1	1	2	1	1	1	1	0.333
Nimboran1	0	0	0	30	1.4	9	1	0.8	0.317
Nimboran2	0	1	0	3	1.7	6	1	0.7	0.466
Nivkh	1	0	1	2	1	1	1	1	0.333
North Saami	0	0	0	2	2	1	1	1	0
Northern Akhvakh	2	0	1	6	1.5	4	1	1	0.333
Nuer1	0	0	0	11	1	5	0	0.75	0.333
Nuer2	0	0	0	11	1	4	0	0.12	0.333
Paez	1	1	1	3	3.7	1	1	1	0.333
Pite Saami1	0	1	0	5	1.3	3	0.5	0.9	0.333
Pite Saami2	0	1	0	3	2	5	0	0.2	0.333
Pite Saami3	0	0	0	5	1.3	8	0	0.8	0.397
Pite Saami4	0	0	0	4	1	3	0	0.1	0.267
Pite Saami5	0	0	0	4	1	9	0	0.1	0.407
Servigliano	1	1	0	6	1.3	2	0	0.09	0.443
Skolt Saami1	0	0	0	7	1	5	0	0.3	0.333

Continued

(*Continued*)

aamorphome_ID	str_ constr	weak_ constr	repeated_ parad	number_ exponents	shared_ segments	word_ forms	informative	percent_ lexicon	MS_ coherence
Skolt Saami2	0	0	0	7	1	7	0	0.3	0.4
Skolt Saami3	1	0	0	9	1.2	3	0	0.1	0
Sobei1	1	0	0	2	1	5	0.5	0.005	0.333
Sobei2	0	0	0	2	1	7	0.5	0.005	0.466
Spanish1	1	0	0	2	1.5	4	0	0.03	0.333
Spanish2	1	0	0	4	1.5	6	0	0.03	0.466
Sunwar	2	0	0	3	1	5	0	0.05	0.333
Svan	2	0	0	11	1.2	5	0.5	0.95	0.333
Tapiete	0	1	0	2	1	3	1	0.8	0.333
Thulung1	0	0	0	3	1	8	0.5	0.05	0.238
Thulung2	0	0	0	3	1	12	0.5	0.08	0.466
Tol1	0	1	0	5	1.5	4	0	0.65	0.333
Tol2	0	0	0	13	1.6	11	0.5	0.4	0.382
Tol3	0	0	0	16	1.7	7	0.5	0.15	0.388
Tol4	0	1	0	5	1.3	4	0.5	0.1	0.333
Tol5	0	0	0	2	1	10	0.5	0.004	0.429
Tol6	0	1	0	3	1	8	0.5	0.006	0.466
Tol7	0	0	0	3	1	4	0	0.002	0.333
Toposa	0	1	0	2	1	3	1	1	0.466
Turkana	0	1	0	2	1	3	1	1	0.443

Twi	1	0	0	2	2	2	1	1	0
Udmurt	0	1	0	2	1.5	6	0.5	1	0.466
Vitu	1	0	0	2	1	2	0.5	1	0.333
Vures	0	0	0	5	1	8	0	0.5	0.333
Wambisa1	0	0	0	2	1	1	1	1	0.333
Wambisa2	1	1	1	2	1	2	0	1	0.333
Wubuy1	1	0	0	7	2.43	4	0.5	0.95	0.429
Wubuy2	0	1	0	7	2.14	2	0.5	0.72	0.333
Wubuy3	0	1	0	10	2.3	2	0.5	0.97	0.333
Wubuy4	0	1	0	7	1.7	3	0.5	0.95	0.333
Wutung1	0	1	0	4	1.3	2	1	0.1	0
Wutung2	0	0	0	3	1.3	3	1	0.15	0.333
Wutung3	0	0	0	3	1.3	4	1	0.15	0.365
Wutung4	0	0	0	3	1	2	1	0.15	0.333
Yagaria	1	2	1	4	1	2	1	1	0
Yele	2	0	1	3	1.33	4	0.5	1	0.333
Yorno-So	1	0	1	6	1.5	4	0.5	1	0.333
Zapotec T.	0	0	0	4	1	16	0	0.08	0.333
Zapotec Y.	1	0	0	8	1	6	0	0.15	0.333

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 morphome occurs more than once in a single lexeme's paradigm) is a binary, yes/no variable but has been coded as 1 for 'yes' and 0 for 'no'. Informativity, in turn, 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: 109–49.

Ackerman, Farrell, James P. Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins and J. Blevins (eds), *Analogy in Grammar: Form and Acquisition*, 54–81. Oxford: Oxford University Press.

Ackerman, Farrell and Olivier Bonami. 2017. Systemic polyfunctionality and morphology–syntax interdependencies. In Nikolas Gisborne and Andrew Hippisley (eds), *Defaults in Morphological Theory*, 233–68. Oxford: Oxford University Press.

Ackerman, Farrell, and Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89,3: 429–64.

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, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90,2: 119–61.

Ali, Mohammed and Andrzej Zaborski. 1990. *Handbook of the Oromo Language*. Cracow: Polska Academia Nauk-Oddzial W. Krakowie.

Allen, Nicholas Justin. 1975. *Sketch of Thulung Grammar: With Three Texts and a Glossary*. Ithaca, NY: Cornell University East Asia.

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*. The Hague: Martinus Nijhoff.

Andersen, Torben. 1993. Vowel quality alternation in Dinka verb inflection. *Phonology* 10,1: 1–42.

Anderson, Judi Lynn. 1989 . *Comaltepec Chinantec Syntax*. Dallas, TX: 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, and 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. In Matthew Baerman (ed.), *The Oxford Handbook of Inflection*, 11–33. Oxford: Oxford University Press.

Arkadiev, Peter and Maria Klamer. 2018. Morphological theory and typology. In Jenny Audring and Francesca Masini (eds), *The Oxford Handbook of Morphological Theory*, 435–54. Oxford: Oxford University Press.

Arnott, David W. 1970. *The Nominal and Verbal System of Pula*. Oxford: Clarendon 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 and Miriam Voghera (eds), *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della società di linguistica italiana*, 39–52. Rome: Bulzoni.

Aronson, Howard I. 1991. Modern Georgian. In Alice C. Harris (ed.), *The Kartvelian Languages*, 219–312. New York: Delmar.

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: 539–51.

Baerman, Matthew. 2007b. Morphological reversals. *Journal of Linguistics* 43,1: 33–61.

Baerman, Matthew. 2012. Paradigmatic chaos in Nuer. *Language* 88,3: 467–94.

Baerman, Matthew. 2013. Inflection class interactions. In Nabil Hathout, Fabio Montermini, and Jesse Tseng (eds), *Morphology in Toulouse: Selected Proceedings of Décembrettes 7*, 11–33. Munich: Lincom.

Baerman, Matthew, and Dunstan Brown. 2013. Syncretism in verbal person/number marking. In Matthew S. Dryer and Martin Haspelmath (eds), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2005. *The Syntax–Morphology Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett (eds) 2017. *Morphological Complexity*. Cambridge: Cambridge University Press.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2017. External typology of inflection classes. In Matthew Baerman, Dunstan Brown, and Greville G. Corbett (eds), *Morphological Complexity*, 8–33. Cambridge: Cambridge University Press.

Baerman, Matthew, and Greville G. Corbett. 2007. Linguistic typology: morphology. *Linguistic Typology* 11,1: 115–17.

Baerman, Matthew and Greville G. Corbett. 2013. Person by other means. In Dik Bakker and Martin Haspelmath (eds), *Languages across Boundaries: Studies in Memory of Anna Siewierska*, 1–14. Berlin: de Gruyter

Baerman, Matthew, Greville G. Corbett, and Dunstan Brown (eds) 2010. *Defective Paradigms: Missing Forms and What They Tell Us*. Oxford: Oxford 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, and Jochen Trommer. 2012. Paradigm learning and subanalysis complexity. In *Proceedings of the 48th Annual Meeting of the Chicago Linguistic Society*, 45–59.

Barcos, Miguel Ángel. 2007. *El aragonés ansotano. Estudio lingüístico de Ansó y Fago*. Zaragoza: Institución Fernando el Católico.

Baudouin de Courtenay, Jan. 1972 [1895]. An attempt at a theory of phonetic alternations. In Edward Stankiewicz (ed.), *A Baudouin de Courtenay Anthology*, 144–215. Bloomington: Indiana University Press

Bauer, Laurie, Rochelle Lieber, and 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: 363–6.

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, NY: SUNY Press.

Beaton, Arthur Charles. 1968. *A Grammar of the Fur Language*. Sudan Research Unit, Faculty of Arts, University of Khartoum.

Bermúdez-Otero, Ricardo and Ana R. Luís. 2016. A view of the morphome debate. In Ana R. Luís and Ricardo Bermúdez-Otero (eds), *The Morphome 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*. 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 Bril (ed.), *Clause-Hierarchy and Clause-Linking: The Syntax and Pragmatics Interface*, 51–101. Amsterdam: John Benjamins.

Bickel, Balthasar, and 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, and 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*, 26–27 May 2002. Nijmegen: Max Planck Institute for Psycholinguistics.

Bickel, Balthasar, and Johanna Nichols. 2005. Inclusive–exclusive as person vs. number categories worldwide. In Michael Cysouw and Elena Filimonova (eds), *Clusivity: Typology and Case Studies of the Inclusive/Exclusive Distinction*, 49–72. Amsterdam: John Benjamins.

Bickel, Balthasar, and Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, 169–240. Cambridge: Cambridge University Press.

Bittner, Dagmar. 1995. Affixhomonymie in der Naturlichkeitstheorie. In Norbert Boretzky et al. (eds), *Natürlichkeitstheorie und Sprachwandel. Beitrage zum internationalen Symposium über Natürlichkeitstheorie und Sprachwandel an der Universität Maribor, 1993*, 123–44. Bochum: Brockmeyer.

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler, and Morten H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113,39: 10818–23.

Blevins, James P. 1995. Syncretism and paradigmatic opposition. *Linguistics and Philosophy* 18: 113–52.

Blevins, James P. 2016. *Word and Paradigm Morphology*. Oxford: Oxford University Press.

Blevins, James P., Farrell Ackerman, Robert Malouf, and Michael Ramscar. 2016. Morphology as an adaptive discriminative system. In Daniel Siddiqi and Heidi Harley (eds), *Morphological Metatheory*, 271–302. Amsterdam: John Benjamins.

Blevins, James P., and Juliette Blevins (eds) 2009. *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.

Blevins, James P., Petar Milin, and Michael Ramscar. 2017. The Zipfian Paradigm Cell Filling Problem. In F. Kiefer, J. P. Blevins, and H. Bartos (eds), *Perspectives on Morphological Structure: Data and Analyses*, 139–58. Leiden: Brill.

Bloomfield, Leonard. 1926. A set of postulates for the science of language. *Language* 2,3: 153–64.

Bloomfield, Leonard. 1943. Meaning. *Monatshefte für deutschen Unterricht* 35,3/4: 101–106.

Bobaljik, Jonathan David, and Uli Sauerland. 2018. ABA and the combinatorics of morphological features. *Glossa* 3,1: 1–34.

Bonami, Olivier. 2017. Predictability in inflection and word formation. Talk delivered at ParadigMo 2017.

Bonami, Olivier, and Pollet Samvelian. 2009. Inflectional periphrasis in Persian. HPSG 2009 Conference. Stanford, CA: CSLI. DOI:10.21248/hpsg.2009.2

Bonami, Olivier, and Gregory T. Stump. 2016. Paradigm function morphology. In Andrew Hippisley and Gregory T. Stump (eds), *Cambridge Handbook of Morp*hology, 449–81. Cambridge: Cambridge University Press.

Booij, Geert. 2016. Morphology: The structure of words. In Keith Allan (ed.), *The Routledge Handbook of Linguistics*, 104–17. Abingdon: Routledge.

Booij, Geert, and 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, Université Paris VII.

Boyé, Gilles, and 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, and Gauvain Schalchli. 2016. The status of paradigms. In A. Hippisley and G. Stump (eds), *The Cambridge Handbook of Morphology*, 206–34. Cambridge: Cambridge University Press.

Braune, Wilhelm, and Ingo Reiffenstein. 2004. *Althochdeutsche Grammatik I. Laut- und Formenlehre*. Tübingen: Niemeyer.

Brown, Dunstan, and Marina Chumakina. 2013. What there might be and what there is: An introduction to Canonical Typology. In Dunstan Brown, Marina Chumakina, and Greville G. Corbett (eds), *Canonical Morphology and Syntax*, 1–19. Oxford: Oxford University Press.

Burzio, Luigi. 2001. Zero derivations. *Linguistic Inquiry* 32,4: 658–77.

Butler, Inez M. 1980. *Gramática zapoteca. Zapoteco de Yatzachi el Bajo*. México: Instituto Linguístico 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. PhD thesis, University of Tromsø.

Camilli, Amerindo. 1929. Il dialetto di Servigliano (Ascoli Piceno). *Archivum Romanicum* 13: 220–71.

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, and Giuliana Giusti. 2001. Semi-lexical motion verbs in Romance and Germanic. In Norbert Corver and 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*, 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 70,4: 737–88.

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. and Richard N. Ringler (eds) 1971. *Bright's Old English Grammar and Reader*. Fort Worth, TX: Harcourt Brace Jovanovich.

Chumakina, Marina, and Greville G. Corbett. 2015. Gender–number marking in Archi: Small is complex. In Matthew Baerman, Dunstan Brown, and Greville G. Corbett (eds), *Understanding and Measuring Morphological Complexity*, 93–116. Oxford: Oxford University Press.

Ciucci, Luca. 2016. *Inflectional Morphology in the Zamucoan Languages*. Asunción: CEADUC.

Ciucci, Luca, and 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., and Linda L. Criswell. 1993. *El idioma koreguaje (tucano occidental)*. Dallas, TX: SIL.

Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges, and David S. Rood (eds), *Linguistic Diversity and Language Theories*, 25–49. Amsterdam: John Benjamins.

Corbett, Greville G. 2009. Canonical inflectional classes. In *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 1–11.

Corbett, Greville G. 2016. Morphomic splits. In Ana R. Luís and Ricardo Bermúdez-Otero (eds), *The Morphome Debate: Diagnosing and Analysing Morphomic Patterns*, 64–88. Oxford: Oxford University Press.

Corbett, Greville G., and Sebastian Fedden. 2016. Canonical gender. *Journal of Linguistics* 52,3: 495–531.

Coupé, Christophe, Yoon Mi Oh, Dan Dediu, and 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*. Leiden: 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. http://www.deniscreissels.fr/public/Creissels-conj.disj.pdf

Crowell, Janet I. 1983. Gramática pedagógica Borôro. Brasília: Summer Institute of Linguistics.

Crowley, Terry. 1998. *An Erromangan (Sye) Grammar*. Honolulu: University of Hawai'i Press.

Crysmann, Berthold, and Olivier Bonami. 2016. Variable morphotactics in Information-Based Morphology. *Journal of Linguistics* 52,2: 311–74.

Csúcs, Sándor. 1988. Die wotjakische Sprache. In Denis Sinor (ed.), *The Uralic Languages: Description, History and Foreign Influences* 1, 131–46. 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*, 213–30. Amsterdam: John Benjamins.

Davis, Colin. 2019. Mismatched suppletion in Azeri as morphology/phonology competition. In *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic* 4,1: 29–40.

de Angulo, Jaime, and Lucy Shepard Freeland. 1930. The Achumawi language. *International Journal of American Linguistics* 6,2: 77–120.

de Sousa, Hilario. 2006. The Menngwa Dla language of New Guinea. PhD dissertation, University of Sidney.

Dennis, Ronald K. 1992. *Conjugación del verbo en idioma tol (jicaque)*, vol. 8. Honduras: Instituto Hondureño de Antropología e Historia.

Derksen, Rick. 2007. *Etymological Dictionary of the Slavic Inherited Lexicon*. Leiden: Brill.

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. Berlin: 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.

Döhler, Christian. 2018. *A Grammar of Komnzo*. Berlin: Language Science Press.

Donohue, Mark. 2001. Animacy, class and gender in Burmeso. In Andrew Pawley, Malcolm Ross, and Darrell Tryon (eds), *The Boy from Bundaberg: Studies in Melanesian Linguistics in Honour of Tom Dutton*, 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*, 405–18. 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* The Hague: Martinus Nijhoff.

Dutton, Thomas Edward. 1975. A Koita grammar sketch and vocabulary. *Pacific Linguistics* 29: 281–412.

Dutton, Thomas Edward. 1996. *Koiari*. Munich: Lincom.

Dutton, Thomas Edward. 2003. *A Dictionary of Koiari, Papua New Guinea, with Grammar Notes*. Canberra: Australian National University.

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. 2016. On the distribution of stem alternants: Separation and its limits. In Ana R. Luís and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 276–305. Oxford: Oxford University Press.

Esher, Louise. 2013. Future and conditional in Occitan: A non-canonical morphome. In Silvio Cruschina, Martin Maiden, and John Charles Smith (eds), *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*, 95–115. Oxford: Oxford University Press.

Esher, Louise. 2015. Morphomes 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*, 543–75. Oxford: Oxford University Press.

Evans, Nicholas, I. Wayan Arka, Matthew Carroll, Christian Döhler, Eri Kashima, Emil Mittag, Kyla Quinn, Jeff Siegel, Philip Tama, and 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.

Fast, Gerhard, and 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. 2015. *A Grammar of Skolt Saami*. Helsinki: Suomalais-Ugrilainen Seura.

Feist, Timothy, and Enrique L. Palancar. 2015. Oto-Manguean inflectional class database. University of Surrey. http://dx.doi.org/10.15126/SMG.28/1

Feist, Timothy Richard. 2011. A grammar of Skolt Saami. PhD dissertation, University of Manchester.

Fertig, David. 2016. Mechanisms of paradigm levelling and the role of universal preferences in morphophonological change. *Diachronica* 33,4: 423–60.

Floyd, Simeon, Elisabeth Norcliffe, and Lila San Roque (eds) 2018. *Egophoricity*. Amsterdam: 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, CA: SIL.

Forker, Diana. 2013. *A Grammar of Hinuq*. Berlin: de Gruyter.

François, Alexandre. 2001. Contraintes de structures et liberté dans l'organisation du discours: Une description du Mwotlap, langue océanienne du Vanuatu. PhD dissertation, Université Paris Sorbonne.

François, Alexandre. 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 44: 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*, 103–26. 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*, 147–60. London: Continuum.

Gamkrelidze, Thomas V. 1966. A typology of Common Kartvelian. *Language* 42,1: 69–83.

Gasaway, Eileen, and Heather Sims. 1977. *Girawa Grammar*. Ukarumpa: SIL.

Gause, Georgij F. 1934. *The Struggle for Existence*. Baltimore, MD: Williams and Wilkins.

Georg, Stefan. 2007. *A Descriptive Grammar of Ket (Yenisei-Ostyak)*, Part 1: *Introduction, Phonology and Morphology*. The Hague: Brill/Global Oriental.

Giraudo, Hélène, Serena Dal Maso, and Sabrina Piccinin. 2016. The role of stem frequency in morphological processing. *Mediterranean Morphology Meetings*, 10.

Gönczöl, Ramona. 2007. *Romanian: An Essential Grammar*. Abingdon: Routledge.

González, Hebe Alicia. 2005. A grammar of Tapiete (Tupi-guarani). PhD dissertation, University of Pittsburgh.

Goodman, Noah D., Joshua B. Tenenbaum, Jacob Feldman, and Thomas L. Griffiths. 2008. A rational analysis of rule-based concept learning. *Cognitive Science* 32,1: 108–54.

Gordon, Matthew K. 2009. A fortition-based approach to Balto-Fennic-Sámi consonant gradation. *Folia Linguistica Historica* 18,1–2: 49–79.

Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, and Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences* 114,42: E8822–E8829.

Gruzdeva, Ekaterina. 1998. *Nivkh*. Munich: Lincom.

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, and T. Bures (eds), *Papers on Phonology and Morphology*, 275–88. Cambridge, MA: MIT Working Papers in Linguistics.

Hansson, Gunnar Oláfur. 2007. Productive syncretism in Saami inflectional morphology. In Ida Toivonen and Diane Nelson (eds), *Saami Linguistics*, 91–136. Amsterdam: John Benjamins.

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 Henri Cohen and Claire Lefebvre (eds), *Handbook of Categorization in Cognitive Science*, 21–43. 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: 169–80.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *New Psychology of Language* 2: 1–30.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86,3: 663–87.

Haspelmath, Martin. 2020. The morph as a minimal linguistic form. *Morphology* 30,2: 117–34.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57: 605–33.

Haspelmath, Martin, and 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: 535–53.

Healey, Alan, Ambrose Isoroembo, and Martin Chittleborough. 1969. Preliminary notes on Orokaiva grammar. *Pacific Linguistics. Series A. Occasional Papers* 18: 33–64.

Heath, Jeffrey. 1984. *Functional Grammar of Nunggubuyu*. Canberra: Australian Institute of Aboriginal Studies.

Heath, Jeffrey. 2011. A Grammar of Togo Kan. Available at www.dogonlanguages.org.

Heath, Jeffrey. 2014. A Grammar of Yorno So. Available at www.dogonlanguages.org.

Heine, Bernd, and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Henderson, James E. 1995. *Phonology and Grammar of Yele, Papua New Guinea*. Canberra: Pacific Linguistics.

Hengeveld, Kees, and J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A Typologically-Based Theory of Language Structure*. Oxford: Oxford University Press.

Herce, Borja. 2018. 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*, 105–28. Newcastle upon Tyne: Cambridge Scholars.

Herce, Borja. 2019a. Morphome interactions. *Morphology* 29,1: 109–32.

Herce, Borja. 2019b. Deconstructing (ir)regularity. *Studies in Language* 43,1: 44–91.

Herce, Borja. 2020a. On morphemes and morphomes: Exploring the distinction. *Word Structure* 13,1: 45–68.

Herce, Borja. 2020b. 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. *Morphology* 30: 91–115.

Herce, Borja. 2021a. Stem alternations in Kiranti and their implications for the morphology–phonology interface. *Journal of Linguistics* 57: 321–63.

Herce, Borja. 2021b. Stress and stem allomorphy in the Romance perfectum: Emergence, typology, and motivations of a symbiotic relation. *Linguistics* 60.4: 1103–47.

Herce, Borja, Carmen Saldana, John Mansfield, and Balthasar Bickel. Forthcoming. Positional splits in person–number agreement paradigms reflect a naturalness gradient: Typological and experimental evidence. *Glossa*.

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.

Herrity, Peter. 2000. Slovene. *A Comprehensive Grammar*. Abingdon: Routledge.

Hockett, Charles F. 1947. Problems of morphemic analysis. *Language* 23,4: 321–43.

Hockett, Charles F. 1987. *Refurbishing Our Foundations*. Amsterdam: John Benjamins.

Holmes, Philip, and Ian Hinchliffe. 2003. *Swedish: A Comprehensive Grammar*. Abingdon: Routledge.

Holt, Dennis. 1999. *Tol (Jicaque)*. Munich: Lincom.

Holton, David, Peter Mackridge, and Irene Philippaki-Warburton. 2012. *Greek: A Comprehensive Grammar of the Modern Language*. Abingdon: 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: 271–95.

Jacques, Guillaume. 2017. A reconstruction of Proto-Kiranti verb roots. *Folia Linguistica* 51: 177–215.

Jacques, Guillaume, Aimee Lahaussois, Boyd Michailovsky, and 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.

Jakobi, Angelika. 1990. *A Fur Grammar: Phonology, Morphophonology, and Morphology*. Hamburg: Helmut Buske.

Jakobson, Roman. 1971 [1936]. Beitrag zur allgemeinen Kasuslehre. Gesamtbedeutungen der russischen Kasus. In *Selected Writings* vol. 2, 23–71. The Hague: Mouton.

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, 179–97.

Janda, Richard D., and María Sandoval. 1984. *'Elsewhere' in Morphology*. Bloomington: Indiana University Linguistics Club.

Jensen, John Thayer. 1990. *Morphology: Word Structure in Generative Grammar*. Amsterdam: John Benjamins.

Johnson, Orville E., and Stephen H. Levinsohn. 1990. *Gramática secoya*. Dallas, TX: 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, and Paula Jones. 1991. *Barasano Syntax*. Summer Institute of Linguistics and University of Texas at Arlington.

Jörg, Christina. 1989. *Isländische Konjugationstabellen—Icelandic Conjugation Tables—Tableaux de Conjugaison Islandaise—Beygingatöflur Islenskra Sagna*. Hamburg: Buske.

Jung, Ingrid. 1989. Grammatik des Paez. Ein Abriss. PhD dissertation, Universität Osnabrück.

Kahn, Lily, and Riitta-Liisa Valijärvi. 2017. *North Sámi: An Essential Grammar*. Abingdon: 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, and John Charles Smith (eds), *The Boundaries of Pure Morphology*, 181–208. Oxford: Oxford University Press.

Khalilova, Zaira. 2009. A grammar of Khwarshi. PhD dissertation, LOT.

Kibrik, Aleksandr E. 1994. Khinalug. *The Indigenous Languages of the Caucasus* 4: 367–406.

Kibrik, Aleksandr E., Sandro V. Kodzasov, and Irina A. Muravyova. 2004. Alutor grammar. In Megumi Kurebito (ed.), *Language and Folklore of the Alutor People*. ELPR Endangered Languages of the Pacific Rim. A2-042. Research Institute for Languages and Cultures of Asia and Africa, 203–337.

Kießling, Roland. 1994. *Eine Grammatik des Burunge*. Cologne: Köppe.

Killian, Don 2015. Topics in Uduk phonology and morphosyntax. PhD dissertation, University of Helsinki.

Kimball, Geoffrey D. 1985. A descriptive grammar of Koasati. PhD dissertation, 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, and 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: 10681–6.

Klein, Devorah E., and Gregory L. Murphy. 2001. The representation of polysemous words. *Journal of Memory and Language* 45,2: 259–82.

Koontz-Garboden, Andrew. 2016. Thoughts on diagnosing morphomicity: A case study from Ulwa. In Ana R. Luís and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 309–40. Oxford: Oxford University Press.

Kornfilt, Jaklin. 2013. *Turkish*. Abingdon: Routledge.

Kossmann, Maarten. 2015. Contact-induced change. In Matthew Baerman (ed.), *The Oxford Handbook of Inflection*, 251–272. Oxford: Oxford University Press.

Krishan, Shree. 2001. A sketch of Chaudangsi grammar. In Yasuhiko Nagano and Randy J. Lapolla (eds), *New Research on Zhangzhung and Related Himalayan Languages*, 401–48. Osaka: National Museum of Ethnology.

Kutsch Lojenga, Constance. 1994. *Ngiti: A Central-Sudanic Language of Zaire*. Cologne: Köppe.

Lahaussois, Aimée. 2002. Aspects of the grammar of Thulung Rai: An endangered Himalayan language. PhD dissertation, University of California Berkeley.

Leiss, Elisabeth. 1997. Synkretismus und Natürlichkeit. *Folia Linguistica* 31: 133–60.

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, and Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In LSTL 51: Morphology in Toulouse, 133–53. Munich: Lincom.

Lockwood, William Burley. 1977. *An Introduction to Modern Faroese*. Copenhagen: Munksgaard.

Loporcaro, Michele. 2012. Stems, endings and inflectional classes in Logudorese verb morphology. *Lingue e linguaggio* 11,1: 5–34.

Loporcaro, Michele. 2013. Morphomes in Sardinian verb inflection. In Silvio Cruschina, Martin Maiden, and John Charles Smith (eds), *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*, 137–60. Oxford: Oxford University Press.

Loporcaro, Michele, Nadja Kägi, and Francesco Gardani. 2018. Morfomi sommersi in Pantesco o dell'arte di arrangiarsi in morfologia. In Annamaria Chilà and Alessandro De Angelis (eds), *Capitoli di morfosintassi delle varietà romanze d'Italia. Teoria e dati empirici*, 273–305. Palermo: Centro di studi filologici e linguistici siciliani.

Loving, Richard. 1973. An outline of Awa grammatical structures. In Howard McKaughan (ed.), *The Languages of the Eastern Family of the East New Guinea Highland Stock*, 65–87. Seattle: University of Washington Press.

Luís, Ana R., and Ricardo Bermúdez-Otero (eds) 2016. *The Morphome Debate*. Oxford: Oxford University Press.

Lupardus, Karen Jacque. 1982. The language of the Alabama Indians. PhD dissertation, University of Kansas.

Lynch, John. 2000. *A Grammar of Anejom*. Canberra: Pacific Linguistics.

Lynch, John, Malcolm Ross, and Terry Crowley (eds) 2002. *The Oceanic Languages*. Abingdon: Routledge.

McCreight, Katherine, and Catherine V. Chvany. 1991. Geometric representation of paradigms in a modular theory of grammar. In Frans Plank (ed.), *Paradigms: The Economy of Inflection*, 91–112. Berlin: de Gruyter.

McPherson, Laura. 2013. *A Grammar of Tommo So*. Berlin: de Gruyter.

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* 78,4: 441–64.

Maiden, Martin. 2005. Morphological autonomy and diachrony. *Yearbook of Morphology 2004*, 137–75.

Maiden, Martin. 2011a. Morphological persistence. In Martin Maiden, John Charles Smith, and Adam Ledgeway (eds), *Cambridge History of the Romance Languages*, 155–215. Cambridge: Cambridge University Press.

Maiden, Martin. 2011b. Morphomes and 'stress-conditioned allomorphy' in Romansh. In Maria Goldbach, Marc-Olivier Hinzelin, Martin Maiden, and 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, and Adam Ledgeway (eds), *Cambridge History of the Romance Languages*, 216–67. Cambridge: Cambridge University Press.

Maiden, Martin. 2012. A paradox? The morphological history of the Romance present. In S. Gaglia and M.-O. Hinzelin (eds), *Inflection and Word Formation in Romance Languages*, 27–54. Amsterdam: John Benjamins.

Maiden, Martin. 2013. The Latin 'third stem' and its Romance descendants. *Diachronica* 30,4: 492–530.

Maiden, Martin. 2016. Some lessons from history: Morphomes in diachrony. In Ana R. Luís and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 33–63. Oxford: Oxford University Press.

Maiden, Martin. 2017. Romansh allomorphy (again!). In Claire Bowern, 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* 53,2: 217–60.

Maiden, Martin. 2018b. *The Romance Verb: Morphomic Structure and Diachrony*. Oxford: Oxford University Press.

Maiden, Maiden, and Paul O'Neill. 2010. Morphomic defectiveness. In Matthew Baerman, Greville G. Corbett, and Dunstan Brown (eds), *Defective Paradigms: Missing Forms and What They Tell Us*, 103–24. London: British Academy.

Maiden, Martin, and Cecilia Robustelli. 2014. *A Reference Grammar of Modern Italian*. Abingdon: Routledge.

Maiden, Martin, John Charles Smith, Silvio Cruschina, Marc-Olivier Hinzelin, and 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*. 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: 304–55.

Mangel, Marc, and Francisco Samaniego. 1984. Abraham Wald's work on aircraft survivability. *Journal of the American Statistical Association* 79,386: 259–67.

Mansfield, John, Sabine Stoll, and Balthasar Bickel. 2020. Category clustering: A probabilistic bias in the morphology of verbal agreement marking. *Language* 96, 2: 255–93.

Mansfield, John. 2021. The word as a unit of internal predictability. *Linguistics* 59,6: 1427–72.

Mansfield, John. n.d. Murrinhpatha finite verb inflection paradigms. 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, and Nino Amiridze (eds), *Borrowed Morphology*, 47–80. Berlin: de Gruyter.

Matthews, Peter H. 1965. The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics* 1,2: 139–71.

Matthews, Peter H. 1991. *Morphology*. Cambridge: Cambridge University Press.

Meier-Brügger, Michael. 2013. *Indo-European Linguistics*. Berlin: de Gruyter.

Mel'čuk, Igor A. 2002. Towards a formal concept 'zero linguistic sign': Applications in typology. In Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer, and Maria D. Voeikova (eds), *Morphology 2000*, 241–58. Amsterdam: John Benjamins.

Merlan, Francesca C. 1994. *A Grammar of Wardaman*. Berlin: de Gruyter

Merrifield, William R. 1968. Palantla Chinantec grammar. PhD 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). In *Linguistics of the Sino-Tibetan Area: The State of the Art*, 363–75. Canberra: Pacific Studies.

Michailovsky, Boyd. 1988. *La Langue Hayu*. Paris: CNRS.

Mielke, Jeff. 2008. *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

Milizia, Paolo. 2015. Patterns of syncretism and paradigm complexity: The case of old and middle Indic declension. In Matthew Baerman, Dunstan Brown, and Greville G. Corbett (eds), *Understanding and Measuring Morphological Complexity*, 167–84. Oxford: Oxford University Press.

Montermini, Fabio, and Olivier Bonami. 2013. Stem spaces and predictability in verbal inflection. *Lingue e linguaggio* 12,2: 171–90.

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: 431–42.

Mosonyi, Esteban Emilio, and 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*: 229–71. 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*. Cologne: Köppe.

Nagore Lain, Francho. 1986. *El aragonés de Panticosa: Gramática*. Huesca: Instituto de Estudios Altoaragoneses.

Naït-Zerrad, Kamal. 1994. *Manuel de conjugaison kabyle. 6000 verbes, 176 conjugaisons: amyag di tmaziyt (le verbe en berbère)*. Paris: L'Harmattan.

Nedjalkov, Vladimir P., and Galina A. Otaina. 2013. *A Syntax of the Nivkh Language: The Amur Dialect*. Amsterdam: John Benjamins.

Nevins, Andrew, Cilene Rodrigues, and Kevin Tang. 2015. The rise and fall of the L-shaped morphome: Diachronic and experimental studies. *Probus* 27,1: 101–55.

Newmark, Leonard, Philip Hubbard, and Peter Prifti. 1982. *Standard Albanian: A Reference Grammar for Students*. Stanford, CA: Stanford University Press.
Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86,3: 688–95.
Nichols, Johanna. 2011. *Ingush Grammar*. Berkeley, CA: University of California Press.
Nikolaeva, Irina. 1999. *Ostyak*. Munich: Lincom.
Noonan, Michael. 2011. *A Grammar of Lango*. Berlin: 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*, 491–520. Oxford: Oxford University Press.
Noreen, Adolf. 1904. *Altschwedische Grammatik*. Tubingen: Niemeyer.
Novelli, Bruno. 1985. *A Grammar of the Karimojong Language*. Berlin: Dietrich Reimer.
Olson, Michael L. 1973. *Barai Sentence Structure and Embedding*. Santa Ana, CA: Summer Institute of Linguistics.
O'Neill, Paul. 2011. The Ibero-Romance verb: Allomorphy and the notion of the morphome. PhD dissertation, Oxford University.
O'Neill, Paul. 2013. The morphome and morphosyntactic/semantic features. In Silvio Cruschina, Martin Maiden, and John Charles Smith (eds), *The Boundaries of Pure Morphology*, 221–46. Oxford: Oxford University Press.
O'Neill, Paul. 2014. The morphome 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.
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*, 53–64.
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 systèmatique 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 (Iiniá 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: 365–84.
Panman, Otto. 1982. Homonymy and polysemy. *Lingua* 58,1–2: 105–36.
Parlier, James. 1964. Managalasi verb inflection. *Te Reo* 7: 28–35.
Pato, Enrique, and 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.
Peña, Jaime. 2016. A grammar of Wampis. PhD dissertation, University of Oregon.
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: 225–66.

Pertsova, Katya. 2014. Logical complexity in morphological learning: Effects of structure and null/overt affixation on learning paradigms. *Annual Meeting of the Berkeley Linguistics Society* 38: 401–13.

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, and Elena Filimonova. 2000. The Universals Archive: A brief introduction for prospective users. *STUF—Language Typology and Universals* 53,1: 109–23.

Pylkkänen, Liina, Rodolfo Llinás, and 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*. 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. 1976. *A Grammar of the Icelandic or Old Norse Tongue*. Amsterdam: John Benjamins.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Real Academia Española. n.d. CORPES XXI: Corpus del Español del Siglo XXI. http://www.rae.es/recursos/banco-de-datos/corpes-xxi (accessed Dec. 2016).

Reh, Mechthild. 1996. *Anywa Language: Description and Internal Reconstruction*. Cologne: Köppe.

Reid, Tatiana. 2010. Aspects of phonetics, phonology and morphophonology of Thok Reel. MSc dissertation, University of Edinburgh.

Reid, Tatiana. 2019. The phonology and morphology of the Nuer verb. PhD dissertation, University of Surrey.

Rensch, Calvin R. 1989. *An Etymological Dictionary of the Chinantec Languages*. Dallas, TX: Summer Institute of Linguistics.

Rießler, Michael. 2022. Kildin Saami. In Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik (eds), *The Oxford Guide to the Uralic Languages*, 216–39. Oxford: Oxford University Press.

Roberts, John R. 1987. *Amele*. London: Croom Helm.

Ronjat, Jules. 1930. *Grammaire historique des parlers provençaux modernes*. Montpellier: Société des langues romanes.

Ross, Malcolm. 1980. Some elements of Vanimo, a New Guinea tone language. In M. Boxwell et al. (eds), *Papers in New Guinea Linguistics* No. 20. A-56: 77–110. Pacific Linguistics, Australian National University. DOI:10.15144/PL-A56.77

Ross, Malcolm. 2001. Kele. In John Lynch, Malcolm Ross, and Terry Crowley (eds), *The Oceanic Languages*, 123–47. Abingdon: Routledge.

Round, Erich, and Greville G. Corbett. 2020. Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24,3: 489–525.

Round, Erich R. 2011. Morphomes as a level of representation capture unity of exponence across the inflection–derivation divide. *Linguistica* 51: 217–30.

Round, Erich R. 2015. Rhizomorphomes, meromorphomes, and metamorphomes. In Matthew Baerman, Dunstan Brown, and 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 and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 228–47. Oxford: Oxford University Press.

Round, Erich R., and 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.

Rupp, James. 1996. *Diccionario chinanteco de San Juan Lealao, Oaxaca*. Tucson, AZ: SIL.

Rupp, James E. 1989. *Lealao Chinantec syntax*. Dallas, TX: 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.

Saeed, John. 1999. *Somali*. Amsterdam: John Benjamins.

Saldana, Carmen, Borja Herce, and Balthasar Bickel. 2022. More or less unnatural: Semantic similarity shapes the learnability and cross-linguistic distribution of syncretism in morphological paradigms. Open Mind.

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, and Jonathan Bobaljik. 2013. Syncretism distribution modeling: Accidental homophony as a random event. In *Proceedings of GLOW in Asia IX 2012*, 31–53.

Saura Rami, José Antonio. 2003. *Elementos de fonética y morfosintaxis benasquesas*. Zaragoza: Gara d'Edizions–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, and Gilles Boyé. 2018. Paradigms and syncretism in derivation: The case of ethnics in French. *Lingue e linguaggio* 17,2: 197–216.

Schanen, François. 2004. *Parlons luxembourgeois. Langue et culture linguistique d'un petit pays au coeur de l'Europe*. Paris: L'Harmattan.

Schulze, Wolfgang. 1997. *Tsakhur*. Munich: Lincom.

Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91,3: 511–32.

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 and Randy J. LaPolla (eds), *New Research on Zhangzhung and Related Himalayan Languages*, 271–321. Osaka: National Museum of Ethnology.

Sharma, Suhnu Ram. 2001b. A sketch of Rongpo grammar. In Yasuhino Nagano and Randy J. LaPolla (eds), *New Research on Zhangzhung and Related Himalayan Languages*. Osaka: National Museum of Ethnology.

Shepard, Roger. N., Carl. I. Hovland, and Herbert M. Jenkins. 1961. Learning and memorization of classifications. *Psychological Monographs: General and Applied* 75(13), 1–42.

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, and Kenny Smith. 2015. Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive Science* 39,1: 212–26.

Silzer, Peter James. 1983. Ambai: An Austronesian language of Irian Jaya, Indonesia. PhD dissertation, Australian National University.

Slobin, Dan I., and 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 morphome as a gradient phenomenon: Evidence from Romance. In Silvio Cruschina, Martin Maiden, and John Charles Smith (eds), *The Boundaries of Pure Morphology*, 247–61. Oxford: Oxford University Press.

Smye, Graham. 2004. *A Grammar of Anufo*. Accra: Institute of African Studies, University of Ghana.

Söderholm, Eira. 2017. *Kvensk Grammatikk*. Oslo: Cappelen Damm Akademisk.

So-Hartmann, Helga. 2009. *A Descriptive Grammar of Daai Chin*. Berkeley, CA: Sino-Tibetan Etymological Dictionary and Thesaurus Project.

Speck, Charles H. 1984. The phonology of the Texmelucan Zapotec verb. *International Journal of American Linguistics* 50,2: 139–64.

Spencer, Andrew. 2016. Stems, the morphome and meaning-bearing inflection. In Ana Luís and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 207–27. Oxford: Oxford University Press.

Spencer, Katharine. 2008. Kwomtari grammar essentials. In Murray Honsberger, Carol Honsberger, and Ian Tupper (eds), *Kwomtari Phonology and Grammar Essentials*, 53–183. Ukarumpa: SIL-PNG.

Spike, Matthew. 2020. Fifty shades of grue: Indeterminate categories and induction in and out of the language sciences. *Linguistic Typology* 24,3: 465–88.

Stemberger, Joseph Paul. 1981. Morphological haplology. *Language* 57: 791–817.

Steriade, Donca. 2016. The morphome vs. similarity-based syncretism: Latin t-stem derivatives. In Ana R. Luís and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 112–72. Oxford: Oxford University Press.

Sterner, Joyce K. 1987. Sobei verb morphology reanalyzed to reflect POC studies. *Oceanic Linguistics* 26,1/2: 30–54.

Storme, Benjamin. 2021. Implicational generalizations in morphological syncretism: The role of communicative biases. *Journal of Linguistics* 58,2: 1–41. doi:10.1017/S0022226721000207

Stump, Gregory. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Stump, Gregory. 2006. Heteroclisis and paradigm linkage. *Language* 82,2: 279–322.

Stump, Gregory. 2014. Polyfunctionality and inflectional economy. *LiLT (Linguistic Issues in Language Technology)* 11: 73–93.

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 and Ricardo Bermúdez-Otero (eds), *The Morphome Debate*, 175–206. Oxford: Oxford University Press.

Stump, Gregory; and Raphael A. Finkel. 2013. *Morphological Typology: From Word to Paradigm*. Cambridge: Cambridge University Press.

Suárez, Jorge A. 1983. *La Lengua Tlapaneca de Malinaltepec*. México: Instituto de Investigaciones Filológicas, Universidad Nacional Autónoma de México.

Tallman, Adam J. R., and Sandra Auderset. 2022. Measuring and assessing indeterminacy and variation in the morphology–syntax distinction. *Linguistic Typology*. https://doi.org/10.1515/lingty-2021-0041

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: 154–62.

Tosco, Mauro. 2001. *The Dhaasanac Language*. Cologne: Köppe.

Trommer, Jochen. 2016. A postsyntactic morphome cookbook. In Daniel Siddiqi and Heidi Harley (eds), *Morphological Metatheory*, 59–93. Amsterdam: John Benjamins.

Trommer, Jochen, and Sebastian Bank. 2017. Inflectional learning as local optimization. *Morphology* 27,3: 383–422.

Tryon, Darrell T. 1970. *An Introduction to Maranungku (Northern Australia)*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Tryon, Darrell T. 1974. *Daly Family Languages*. Canberra: Pacific Linguistics.

Tuite, Kevin. 1994. Aorist and pseudo-aorist for Svan atelic verbs. In H. I. Aronson (ed.), *Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*, 319–40. Chicago: Chicago Linguistic Society.

Tuite, Kevin. 1995. Svan and its speakers. MS. http://mapageweb.umontreal.ca/tuitekj/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 word classes in Javanese. *Lingua* 3: 322–54.

van den Berg, Helma. 1995. *A Grammar of Hunzib (with Texts and Lexicon)*. Munich: Lincom.

van den Berg, René, and Peter Bachet. 2006. *Vitu Grammar Sketch*. Ukarumpa: SIL.

van den Heuvel, Wilco. 2006. *Biak: Description of an Austronesian Language of Papua*. Utrecht: LOT.

van der Voort, Hein. 2007. Proto-Jabutí. Um primeiro passo na reconstrução da língua ancestral dos Arikapú e Djeoromitxí. *Boletim do Museu Paraense Emílio Goeldi (Ciências humanas)* 2,2: 133–68.

Velie, Daniel, and 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*, 345–459. Canberra: Australian National University.

Waag, Christine. 2010. *The Fur Verb and its Context*. Cologne: 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, 1–183.

West, Birdie. 1980. *Gramática popular del Tucano*. Bogotá: Instituto Lingüístico de Verano.

West, Dorothy. 1973. *Wojokeso Sentence, Paragraph, and Discourse Analysis*. Canberra: Australian National University.

Wheeler, Alva. 1970. Grammar of the Siona Language, Colombia, South America. PhD dissertation, University of California at Berkeley.

Wichmann, Søren, and Eric W. Holman. 2009. *Assessing Temporal Stability for Linguistic Typological Features*. Munich: Lincom Europa.

Wilbur, Joshua. 2014. *A Grammar of Pite Saami*. Berlin: Language Science Press.

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*. Ukarumpa: SIL.

Winkler, Eberhard. 2001. *Udmurt*. Munich: Lincom.

Witzlack-Makarevich, Alena, Taras Zakharko, Lennart Bierkandt, Fernando Zúñiga, and Balthasar Bickel. 2016. Decomposing hierarchical alignment: Co-arguments as conditions on alignment and the limits of referential hierarchies as explanations in verb agreement. *Linguistics* 54,3: 531–61.

Van der Wilden, Jaap. 1976. Simplicity and detail in Kemtuk predication. *Irian* 5,2: 59–84.

Wonderly, William L., Lorna F. Gibson, and 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*, 121–59. Oxford: Oxford University Press.

Wurzel, Wolfgang U. 1980. Ways of morphologizing phonological rules. *Historical Morphology* 62: 443–62.

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. Norman, OK: Summer Institute of Linguistics.

Zipf, George Kingsley. 1935. *The Psycho-biology of Language*. Boston, MA: Houghton Mifflin

Žirkov, Lev I. 1955. Lakskij jazyk. Fonetika i morfologija [*The Lak Language: Phonetics and Morphology*]. Moscow: AN SSSR.

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*. Padua: Unipress.

Zwicky, Arnold M. 1991. Systematic versus accidental phonological identity. In Frans Plank (ed.), *Paradigms: The Economy of Inflection*, 113–31. Berlin: de Gruyter.

Language index

Note: entries in bold denote languages present in the database

Index of topics