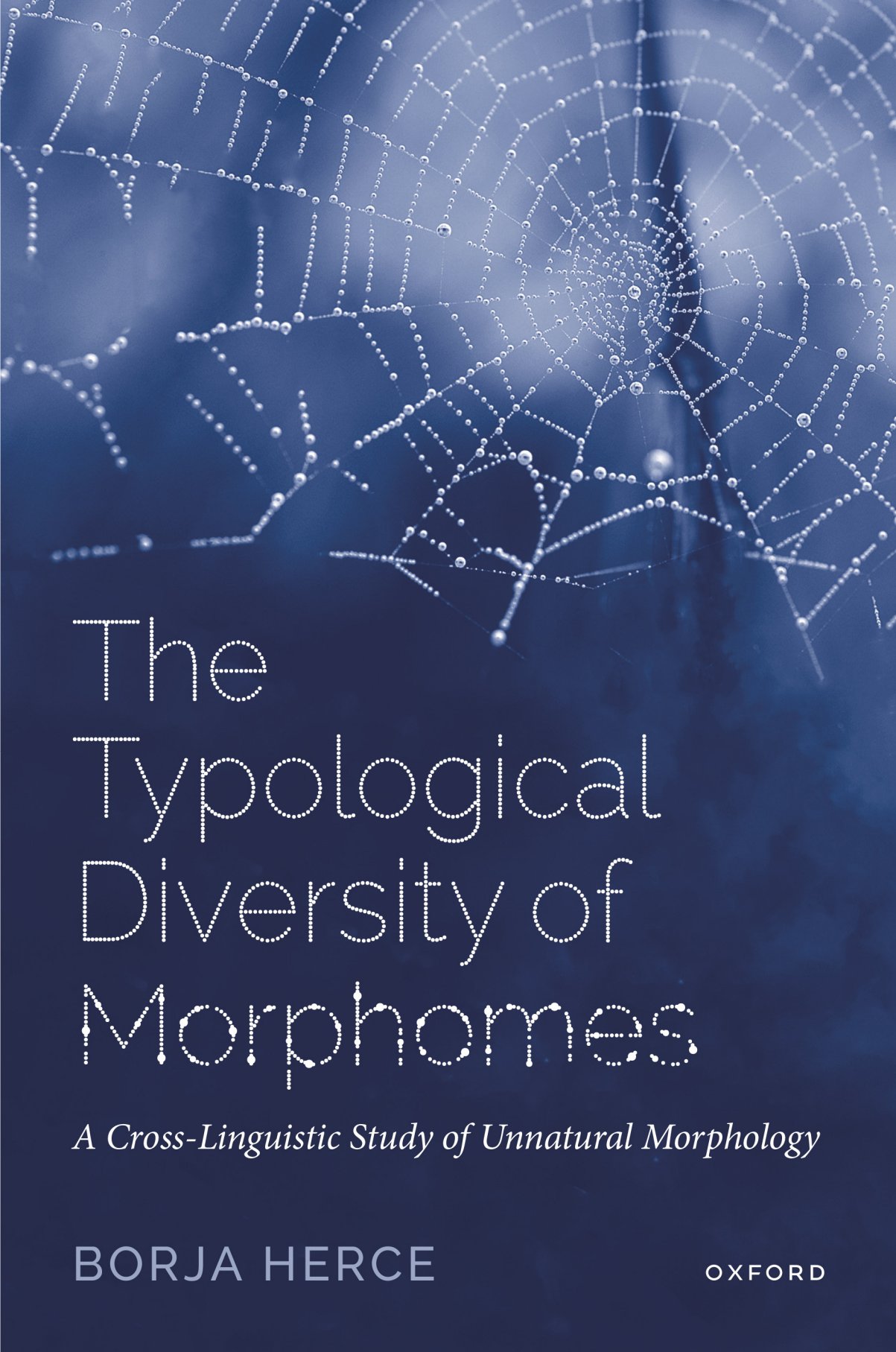




The Typological Diversity of Morphomes

A Cross-Linguistic Study of Unnatural Morphology

BORJA HERCE

OXFORD

The Typological Diversity of Morphemes

The Typological Diversity of Morphemes

*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

© Borja Herce 2023

The moral rights of the author have been asserted

Some rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or
transmitted, in any form or by any means, for commercial purposes, without the prior permission in writing
of Oxford University Press, or as expressly permitted by law, by licence or under terms agreed with the
appropriate reprographics rights organization.



This is an open access publication, available online and distributed under the terms of a Creative Commons
Attribution – Non Commercial – No Derivatives 4.0 International licence (CC BY-NC-ND 4.0), a copy of
which is available at <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Enquiries concerning reproduction outside the scope of this licence should be sent to the Rights
Department, Oxford University Press, at the address above

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

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Contents

<i>Preface</i>	x
<i>Acknowledgement</i>	xi
<i>List of abbreviations</i>	xii
1. Introduction	1
1.1 Initial approximation and goals	1
1.2 History	3
1.3 Terminology	6
1.4 A working definition	9
2. Issues in morpheme identification	13
2.1 Systematic vs accidental	13
2.1.1 Assessing systematicity	14
2.1.1.1 Feature conflict resolution	14
2.1.1.2 Co-occurrence restrictions	15
2.1.1.3 Morphosyntactic evidence for systematicity	16
2.1.1.4 Diachronic evidence for systematicity	17
2.1.1.5 Allomorphic or morphophonological evidence for systematicity	18
2.1.2 On the empirical status of homophony and polysemy	19
2.1.2.1 Semantically motivated split	22
2.1.2.2 Formally motivated split	24
2.2 Natural vs unnatural	25
2.3 Maximal domain	34
2.4 Independence from phonology	41
2.5 Stem spaces	47
2.6 Cross-linguistic recurrence	49
2.7 Blocking	51
2.8 Stem vs affix	54
2.9 Segmentability	56
2.10 Morphological zeroes	58
2.11 Economy	59
2.12 Difficult cases	64
2.12.1 The problem of the 1PL	64
2.12.2 Syntactically licensed morphemes	66
2.12.3 Gender or morpheme?	69
2.13 What (else) can be morphomic?	75
2.13.1 Syncretism/feature sensitivity	75
2.13.2 Heterocclisis	77
2.13.3 Overabundance and defectiveness	78

2.13.4	Morphomicity in derivation	80
2.13.5	Morphomicity in syntagmatics	81
3.	Morphomes in diachrony	83
3.1	The emergence of morphemes	83
3.1.1	Sound change	83
3.1.1.1	Morphological result of sound change	84
3.1.1.2	Paradigmatic locus of sound change	87
3.1.1.3	Zero as a source of morphemes via sound change	88
3.1.2	Semantic drift	90
3.1.3	Analogy	92
3.1.3.1	Morphosyntactically motivated analogy	92
3.1.3.2	Analogy motivated by form	94
3.1.3.3	A note on the motivation of analogy	96
3.1.4	Pattern interactions	97
3.1.5	Grammaticalization	99
3.1.6	Borrowing	102
3.1.7	Mixed origins	104
3.2	Loss and change of morphomic structures	105
3.2.1	Loss of productivity and gradual erosion	106
3.2.2	Loss of morphosyntactic categories	108
3.2.3	Sound change	109
3.2.4	Analogy	110
3.2.4.1	Change into a natural class	110
3.2.4.2	Change into a different unnatural class	112
3.2.5	Mixed causes	113
3.3	Discussion	113
3.4	Conclusion	117
4.	Morphomes in synchrony	120
4.1	Criteria for inclusion in the morphome database	120
4.1.1	Unmistakably unnatural paradigmatic distribution	121
4.1.2	Unmistakably systematic formal identity	124
4.1.3	Other requirements	127
4.1.4	Some excluded morphemes	129
4.2	A Cross-linguistic database of morphemes	133
4.2.1	Africa	135
4.2.1.1	Daasanach	135
4.2.1.2	Daju, Mongo	137
4.2.1.3	Fur	138
4.2.1.4	Iraqw	139
4.2.1.5	Karamojong	140
4.2.1.6	Nuer	140
4.2.1.7	Turkana and Toposa	142

4.2.1.8	Twi	143
4.2.1.9	Yorno-So	145
4.2.2	Asia	145
4.2.2.1	Athpariya	145
4.2.2.2	Bantawa	146
4.2.2.3	Burushaski	149
4.2.2.4	Darma	149
4.2.2.5	Jerung	150
4.2.2.6	Ket	150
4.2.2.7	Khaling	151
4.2.2.8	Khinalugh	152
4.2.2.9	Mehri	153
4.2.2.10	Nivkh	154
4.2.2.11	Northern Akhvakh	155
4.2.2.12	Sunwar	155
4.2.2.13	Svan	156
4.2.2.14	Thulung	158
4.2.2.15	Udmurt	159
4.2.3	Europe	160
4.2.3.1	Aragonese	160
4.2.3.2	Basque	161
4.2.3.3	English	162
4.2.3.4	French	163
4.2.3.5	Greek	163
4.2.3.6	Icelandic	164
4.2.3.7	Irish	165
4.2.3.8	Italian and Servigliano	166
4.2.3.9	Luxembourgish	167
4.2.3.10	North Saami	168
4.2.3.11	Pite Saami	169
4.2.3.12	Skolt Saami	171
4.2.3.13	Spanish and Asturian	172
4.2.4	Australasia	174
4.2.4.1	Barai	174
4.2.4.2	Benabena	175
4.2.4.3	Biak	175
4.2.4.4	Burmeso	176
4.2.4.5	Ekari	177
4.2.4.6	Girawa	177
4.2.4.7	Kele	179
4.2.4.8	Koiari	179
4.2.4.9	Kosena	180
4.2.4.10	Maranunggu	181
4.2.4.11	Menggwa Dla	182
4.2.4.12	Mian	183

4.2.4.13	Murrinh-Patha	184
4.2.4.14	Ngkolmpu and Nen	185
4.2.4.15	Nimboran	186
4.2.4.16	Sobei	188
4.2.4.17	Vitu	188
4.2.4.18	Vurës	188
4.2.4.19	Wubuy	189
4.2.4.20	Wutung	191
4.2.4.21	Yagaria	193
4.2.4.22	Yele	195
4.2.5	America	195
4.2.5.1	Achumawi	195
4.2.5.2	Aguaruna	196
4.2.5.3	Ayoreo	197
4.2.5.4	Bororo	198
4.2.5.5	Chinantec, Lealao	199
4.2.5.6	Chinantec, Palantla	200
4.2.5.7	Jabuti	201
4.2.5.8	Koasati	201
4.2.5.9	Maijiki	202
4.2.5.10	Mazatec	203
4.2.5.11	Me'phaa	203
4.2.5.12	Páez	205
4.2.5.13	Tapieté	205
4.2.5.14	Tol	206
4.2.5.15	Wambisa	208
4.2.5.16	Zapotec, Yatzachi, and Texmelucan	210
4.3	Measuring cross-linguistic variation in morphemes	211
4.3.1	External morphosyntactic constraints	213
4.3.2	Word-form recurrence	215
4.3.3	Paradigmatic recurrence	217
4.3.4	Cross-lexemic recurrence	218
4.3.5	Number of exponents	221
4.3.6	Shared form	222
4.3.7	Informativity	224
4.3.8	Morphosyntactic coherence	226
4.3.9	Morpheme paradigm size and others	228
4.3.10	Locus of marking	232
4.3.11	Cross-linguistic morpheme recurrence	233
4.3.11.1	Recurrent values	233
4.3.11.2	Recurrent patterns	237
4.4	Statistical analysis	240
4.5	Discussion	249
4.5.1	A naturalness bias in morphemes?	249
4.5.2	The nature of the stem-affix distinction	252
4.6	Conclusion	256

5. Implications: Features and forms	259
5.1 The importance of features	259
5.2 The importance of form	263
6. Conclusions	267
<i>Appendix: Morphome database summary</i>	273
<i>References</i>	280
<i>Language index</i>	299
<i>Index of topics</i>	302

Preface

This book constitutes the first typologically oriented monograph on morphemes, 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 morphemes, 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 morphemes 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 morphemes, 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 morphemes 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 morphemes). 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 morphemes, 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, morphemes 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)

<i>caber</i> 'fit' illustrating the L-morphome				<i>poder</i> '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.

against a wider cross-linguistic range of data.² However, and in stark contrast to the wealth of research on Romance morphemes, 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, morphemes, to be truly morphemes, 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 regrettably 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 morphemes, 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 morphemes? What patterns are common and which are infrequent and why? Synchronic data will be complemented with diachronic insights to inform us about: what are the most frequent sources and outcomes of morphemes? What role do frequency or morphosyntactic features play in their evolution?

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

The answers to these questions and the outcomes of this research will be 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 ‘morpheme’, these structures are undoubtedly underreported and underdescribed in descriptive grammars. A cross-linguistic exploration and typologization of morphemes, 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 ‘morpheme’ and the adjective ‘morphemic’ 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 morphemes, 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 morpheme is Janda (1982). There it was argued, for example, that 'morphological homophony in languages is too extensive and too widespread to be due to chance' (Janda 1982: 185) and also that 'a language's system of inflectional and derivational morphology is more highly valued if the same formative appears in more than one word-formation rule' (Janda 1982: 190). To account for the facts, Janda advocated for autonomous morphology and 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 ‘morpheme’ 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 morpheme³ 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 exist⁴ 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)

	<i>venir</i> 'come'		<i>nacer</i> 'be born'		<i>caber</i> '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_{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 morphemes (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 morphomhood (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 morphemes (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.

Issues in morpheme 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/, /ɪz/). 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 haplogy (Stemberger 1981), but can be shown to have grammatical import (consider the ungrammaticality of **the kings 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 morpheme it is obviously an unwarranted aprioristic assumption to always regard as grammatically uninteresting all morphological identities which lack cross-linguistic generality. If some

of those patterns, like the morphemes 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 morphemes 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 inapplicable¹ to most morphemes, 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/-muş/-muş/-müŝ* 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	<i>adam-ı</i>	<i>adam-lar-ı</i>
Possessee PL	<i>adam-lar-ı</i>	<i>adam-lar-ı</i>

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*. Morphemes 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 renowned examples of morphemes (e.g. the Romance L- and N-morphemes) 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 *-rii*, 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 *-kã* and *-ã* to the oblique singular).

As illustrated in the above cases, although perfect morphosyntactic determination is definitionally impossible in morphemes, 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 morphemes 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 morpheme 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: *conduzco* ‘I drive’, *conduje* ‘I drove’, *conducción* ‘driving’. It is hard to believe that every verb ending in *-ducir* (and only those in *-ducir*) is independently and by chance subject to these same operations.

The alternative explanation is that speakers do posit, on the basis of form alone, a grammatical unit at some level despite the lack of shared semantic content.³ Kayardild’s (mero)morphemes (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 paradigm cells of a single lexeme, as in the best-known Romance morphemes (see Table 2.2).

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

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

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 unit 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*.