

The Complexities of Morphology

Edited by

PETER ARKADIEV *and*
FRANCESCO GARDANI

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

© editorial matter and organization Peter Arkadiev and Francesco Gardani 2020
© the chapters their several authors 2020

The moral rights of the authors have been asserted

First Edition published in 2020

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, 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. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

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: 2020932944

ISBN 978-0-19-886128-7

Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

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>List of Figures and Tables</i>	vii
<i>List of Abbreviations</i>	xi
<i>The Contributors</i>	xvii
1. Introduction: Complexities in morphology <i>Peter Arkadiev and Francesco Gardani</i>	1
I. THE LANGUAGE-SPECIFIC PERSPECTIVE	
2. Irregularity, paradigmatic layers, and the complexity of inflection class systems: A study of Russian nouns <i>Jeff Parker and Andrea D. Sims</i>	23
3. Demorphologization and deepening complexity in Murrinhpatha <i>John Mansfield and Rachel Nordlinger</i>	52
4. Overabundance resulting from language contact: Complex cell-mates in Gurindji Kriol <i>Felicity Meakins and Sasha Wilmoth</i>	81
5. Derivation and the morphological complexity of three French-based creoles <i>Fabiola Henri, Gregory Stump, and Delphine Tribout</i>	105
6. Simplification and complexification in Wolof noun morphology and morphosyntax <i>Michele Loporcaro</i>	136
II. THE CROSSLINGUISTIC PERSPECTIVE	
7. Canonical complexity <i>Johanna Nichols</i>	163
8. The complexity of grammatical gender and language ecology <i>Francesca Di Garbo</i>	193
9. Morphological complexity, autonomy, and areality in western Amazonia <i>Adam J. R. Tallman and Patience Epps</i>	230

vi CONTENTS

III. THE ACQUISITIONAL PERSPECTIVE

- | | |
|---|-----|
| 10. Radical analyticity as a diagnostic of adult acquisition | 267 |
| <i>John H. McWhorter</i> | |
| 11. Different trajectories of morphological overspecification and
irregularity under imperfect language learning | 283 |
| <i>Aleksandrs Berdicevskis and Arturs Semenuks</i> | |
| 12. Where is morphological complexity? | 306 |
| <i>Marianne Mithun</i> | |

IV. DISCUSSION

- | | |
|---|-----|
| 13. Morphological complexity and the minimum description
length approach | 331 |
| <i>Östen Dahl</i> | |
| <i>References</i> | 345 |
| <i>Language Index</i> | 383 |
| <i>Subject Index</i> | 387 |

1

Introduction

Complexities in morphology

Peter Arkadiev and Francesco Gardani

1.1 Setting the scene

Morphological and, broadly, linguistic complexity has become a popular topic in linguistic typology and theorizing, as several recent publications testify to, such as McWhorter (2001, 2005, 2018); Kusters (2003); Dahl (2004); Hawkins (2004, 2014); Trudgill (2004a, 2011); Shosted (2006); Miestamo et al. (2008); Sampson et al. (2009); Dressler (2011); Kortmann & Szmrecsanyi (2012); Newmeyer & Preston (2014); Baerman et al. (2015b, 2017); Reintges (2015); Baechler & Seiler (2016); Mufwene et al. (2017); among many others. While this large body of work has contributed to significantly improving our understanding of morphological complexity, a number of key issues remain unsettled. They are of both theoretical and empirical nature and pertain to the domain of morphology and morphosyntax as well as to the ways language use and its socioecological conditions influence linguistic structure. Undoubtedly, the most pressing question is what morphological complexity actually is. There is no straightforward answer to this question, as we will see. The issue of how to define ‘morphological complexity’ is of central importance to us and will be treated in detail in the course of this Introduction and of the volume. To properly frame this central issue, however, we can anticipate that the notion of ‘complexity’ in morphological systems is often revealed and investigated through a set of relative measures that attempt to quantify the extent of morphology in a language, the predictability of the morphological system, and the pressures this places on processing and acquisition. The goal of the present volume is to build upon previous work on morphological complexity and to provide a crosslinguistic view on the key problems of its investigation seen from the perspective of a variety of current approaches.

In the heart of all discussions of linguistic complexity, and especially of morphological complexity, lies the idea that complexity itself is a parameter of crosslinguistic variation. The history of this line of thought (see Joseph & Newmeyer 2012 for an excellent overview) shows some non-trivial swings of the pendulum ranging from the pre-theoretical assumptions of the linguists and

2 PETER ARKADIEV AND FRANCESCO GARDANI

philosophers of the early nineteenth century about the ‘complex’ classic Indo-European languages as opposed to the ‘primitive’ languages of ‘uncivilized people’ to explicit statements that all languages are equally complex. The latter view, which is known under the label of ‘equicomplexity hypothesis’, takes into account obvious differences between languages in the mere degree of elaboration of different structural subdomains (such as, e.g., vowels vs. consonants or nominal vs. verbal morphology); it states that ‘these isolable properties may hang together in such a way that the total complexity of a language is approximately the same for all languages’ (Wells 1954: 104; see also Hockett 1958: 180). Such a position, which is still commonly held by linguists of different backgrounds and theoretical persuasions (see, again, Joseph & Newmeyer 2012: 348–9; and Miestamo 2017), has been challenged by others, who have shown that ‘complexity in one area of grammar [correlates] positively with complexity in another area’ (Sinnemäki 2014: 190).

With the development of contact linguistics and especially of pidgin and creole studies in the second half of the twentieth century, claims started being made that pidgins and creoles are structurally overall simpler than languages with a ‘regular’ sociolinguistic history (see, e.g., such work as Bickerton 1984; McWhorter 2001, 2005; Parkvall 2008; Bakker et al. 2011; Good 2012b, 2015), and, more generally, it has been claimed that linguistic complexity is subject to diachronic change and the effects of language contact (see Dahl 2004 and Trudgill 2011). As a matter of fact, statements to the effect that sociolinguistic parameters such as the number of speakers and degree of contact with other languages affect the complexity of linguistic (sub)systems go back as early as Jakobson (1929) and Trudgill (1983).

Once it had been recognized that morphological complexity is a parameter of crosslinguistic variation, the urge arose to develop non-impressionistic and cross-linguistically applicable ways of measuring and quantifying the degree of morphological complexity of individual languages. The most important proponent of this line of thought is certainly Greenberg (1954), who developed a methodology of quantitative measurement of different types of morphological structure, the most famous of which is the ‘synthetic index’ (p. 185), that is, morpheme-to-word¹ ratio in a sample of texts, which arranges languages into a continuum spanning from radically isolating to polysynthetic. This simple metric, however, is clearly insufficient for the assessment of morphological complexity, since morphology is much more than mere arrangement of morphemes into words. As a simple illustration, consider the case-number paradigms of Turkish (Lewis 2001: 28) and Lithuanian (P.A.’s own knowledge) nouns in Table 1.1.

Both Turkish and Lithuanian have two number and six case values, yielding twelve word forms. However, while in Turkish case and number are expressed

¹ ‘Word’ is intended as ‘word form’.

Table 1.1. Case paradigm of Turkish *ev* ‘house’ and Lithuanian *miestas* ‘city’

	SG	PL		SG	PL
NOM	<i>ev</i>	<i>ev-ler</i>	NOM	<i>miest-as</i>	<i>miest-ai</i>
ACC	<i>ev-i</i>	<i>ev-ler-i</i>	ACC	<i>miest-q</i>	<i>miest-us</i>
GEN	<i>ev-in</i>	<i>ev-ler-in</i>	GEN	<i>miest-o</i>	<i>miest-ų</i>
DAT	<i>ev-e</i>	<i>ev-ler-e</i>	DAT	<i>miest-ui</i>	<i>miest-ams</i>
LOC	<i>ev-de</i>	<i>ev-ler-de</i>	LOC	<i>miest-e</i>	<i>miest-uose</i>
ABL	<i>ev-den</i>	<i>ev-ler-den</i>	INS	<i>miest-u</i>	<i>miest-ais</i>

separately by dedicated suffixes in a compositional way, Lithuanian has cumulative (fused) exponence of both features. Under Greenberg’s morpheme-per-word ratio, Turkish nominal word forms are more complex than Lithuanian ones just because Turkish may have three (and in fact much more) morphemes per nominal word form (e.g., *ev-ler-de* house-PL-LOC), while Lithuanian has only two (*miest-uose* city-LOC.PL). However, if we consider the total number of different affixes occurring in the given paradigms, we find that Turkish with its six overt affixes is actually simpler than Lithuanian with its twelve affixes (see, e.g., Plank 1986 for an early attempt to assess the complexity of morphological systems in such terms). Things become even more complicated if we go beyond Table 1.1 and consider the existence of at least five arbitrary inflectional classes of nouns in Lithuanian intersected by four partly arbitrary accentual classes, also called ‘accentual paradigms’ (a.p.), in Table 1.2 (from Arkadiev et al. 2015: 16; ‘hard’ and ‘soft’ refers to subdeclensions with non-palatalized and palatalized stem-final consonant, respectively; for more details on Lithuanian declension classes, see Ambrazas et al. 2006: 107–33).

This example suggests that along with morphological complexity on the syntagmatic axis (something that can be measured by the morpheme-to-word ratio) there exists morphological complexity on the paradigmatic axis, the two being logically and empirically independent of one another. Thus understood, morphological complexity becomes a composite notion and does not admit of such simple measurement as syntagmatic complexity (see more on this issue below), therefore an unbiased and non-reductionist crosslinguistic empirical investigation of morphological complexity itself becomes a fairly complex problem.²

All in all, it seems to us that the most urgent still unsolved issues in morphological complexity can be captured in terms of the following questions:

² In this connection, Haspelmath (2009) has shown that parameters traditionally attributed to ‘flexion’, as opposed to ‘agglutination’, such as cumulation, stem allomorphy, and affix allomorphy, are logically and empirically independent of each other.

Table 1.2. Sample paradigms of Lithuanian nouns

SG	I hard 'man' (M)		I soft 'horse' (M)		II hard 'day' (F)		II soft 'bee' (F)		III hard 'son' (M)		IV (soft) 'night' (F)	
	I a.p.	III a.p.	I soft 'horse' (M) III a.p.	IV a.p.	II hard 'day' (F) IV a.p.	II soft 'bee' (F) II a.p.	III hard 'son' (M) III a.p.	IV (soft) 'night' (F) IV a.p.				
SG	NOM	vyras	arklys	dienà	bitè	sūmūs	naktis					
	GEN	vyro	arklio	dienòs	bitès	sūnaūs	naktiès					
	DAT	vyrui	arkliui	dienai	bitei	sūmui	nàkčiai					
	ACC	vyrą	arklį	dieną	bitę	sūmų	nàktį					
	INS	vyrui	arkliu	dienà	bitė	sūnumi	naktimi					
	LOC	vyre	arklyje	dienoje	biteje	sūmujè	naktyje					
	VOC	vyre	arkly	diena	bite	sūnaù	naktiè					
	NOM	vyrui	arkliai	dienos	bitės	sūnūs	naktys					
PL	GEN	vyrų	arklių	dienų	bičių	sūnų	naktų					
	DAT	vyrams	arkliams	dienoms	bitėms	sūnums	naktims					
	ACC	vyrus	arklius	dienàs	bitès	sūnus	naktis					
	INS	vyrAIS	arkliais	dienomis	bitėmis	sūnumis	naktimis					
	LOC	vyruose	arkliuose	dienosè	bitėse	sūmuosè	naktysè					

1. The hypothesis that morphology and syntax represent distinctly different, but interdependent types of grammatical organization has been challenged by scholars such as Haspelmath (2011), claiming that the divide between morphology and syntax is not clear-cut and hence irrelevant for typology. Given this, are there theoretical and methodological tools suitable to define morphological complexity and if yes, which ones?
2. If we, however, accept the hypothesis that the morphology vs. syntax divide is crosslinguistically and theoretically valid (see Arkadiev & Klamer 2019; Arkadiev 2020)—a view which we espouse—can we arrive at a uniform notion of morphological complexity given the diversity of morphological phenomena?
3. In direct connection to the former question, can we arrive at a single and straightforward measure of complexity that applies to languages that display radically different morphological encoding strategies?
4. What is the role of sociolinguistic, psycholinguistic, and diachronic factors in affecting morphological complexity?

These problems constitute the main research questions of this volume, which aims to tackle them in a principled way, by presenting a collection of original research papers on different aspects of morphological complexity. This introductory chapter is meant to outline the field and take the reader through the volume, and it is organized as follows: section 1.2 pursues the question of the scope of ‘morphological complexity’; section 1.3 surveys several conceptions and methodological approaches to morphological complexity distinguishing between two main types: formal approaches (section 1.3.1) and psycholinguistic approaches (section 1.3.2). Section 1.4 presents the structure of the volume and summarizes the contributions to it.

1.2 What is complex?

In all discussion on morphological complexity, a question hangs in the air. Is morphology complex in its own right? This question is partly rhetorical, maybe trivial, but still central, as it concerns the theoretical demarcation of the object of investigation. The widespread expression ‘morphological complexity’ has at least two readings. It can refer to the overall contribution of morphology to complexity in grammar or it can mean complexity inside morphology.

The first reading, viz. morphology as a source of complexity for the overall language system, would be justified by the fact that languages can do (almost) entirely without morphology and that ‘a language can persist for a long time with little or no morphology’ (Aronoff 2015: 282). In this vein, Carstairs-McCarthy (2010: ch. 2) and Anderson (2015a: 12–13) conceive of morphology as a

6 PETER ARKADIEV AND FRANCESCO GARDANI

redundant architectural quirk added to the logically necessary systems of syntax and phonology, and Aronoff goes so far to declare: ‘morphology is inherently unnatural. It’s a disease, a pathology of language’ (Aronoff 1998: 413). Such a view apparently entails that languages without morphology (e.g., Yoruba) are less complex than languages with at least a little morphology (e.g., Tok Pisin). This type of morphological complexity could then be paraphrased as ‘complexity induced by morphology’. The assumption that morphology *per se* is a complication resonates with the terminological use of ‘morphological complexity’ to define the property of words having an internal morphological structure, being, so to say, morphologically complex, as we find in some authors concerned with word recognition (e.g., Fiorentino & Poeppel 2007; Bozic & Marslen-Wilson 2010), sign linguistics (Zwitserslood 2003), and rarely word formation (Hay 2003). Clearly, in this usage, complexity means the presence of internal structure, and claiming that a formally complex (i.e., composite) word is in itself complex, as opposed to a simplex word, amounts to saying that morphology as such is complexity. That would imply that morphology makes the language system more complex—an observation that is relative to other components of a language’s grammar. Adopting the concept of ‘effective complexity’ by Gell-Mann (1995), Moscoso del Prado Martín (2011) performs a corpus-based measure of the inflectional complexity of six European languages and claims that there is a ‘strong degree of mutual dependence between morphological and syntactic information.’ As he shows, when information on word order is explicitly factored in, the apparent gradation in complexity across languages, as calculated on the basis of the number of inflected forms per word, disappears. He arrives at the conclusion that ‘inflectional morphology serves a role in reduction of uncertainty, simplifying the description of the whole grammar’ (p. 3528). Whether or not this be the case, this question—although of great importance also for cognitive approaches to complexity—is not within the scope of the present book. Rather, we are concerned with the second reading of morphological complexity, that is, complexity inside morphology.

Taking an inner-morphological perspective, we focus on which morphological phenomena can be considered complex or more complex than others and look at different degrees of complexity *within* morphology. Some authors have swiftly found an answer to this question, by identifying the core of morphological complexity in phenomena currently running under the heading of autonomous (or ‘pure’) morphology—including morphological entities and processes that are not extramorphologically motivated in a straightforward way, such as, for example, inflectional classes, allomorphy, patterns of syncretism, suppletion, etc. (Aronoff 1994; Maiden et al. 2011; Cruschina et al. 2013). For example, Baerman et al. (2015b: 4) consider morphological complexity as ‘the additional structure that cannot readily be reduced to syntax or phonology’. This extra layer of purely morphological structure, such as inflection classes in the Lithuanian example in

section 1.1, may attain an astonishing degree of gratuitous complexity, whereas the mere presence of (possibly elaborate) transparent and regular affixal expression of grammatical meaning, such as exemplified by Turkish, is of least relevance for the study of morphological complexity (see also a discussion of different aspects of complexity in the polysynthetic languages, traditionally assumed to be the hallmark of morphological complexity, by Dahl 2017 and Sadock 2017).

Of course, the decision to only focus on autonomous morphology has a great methodological advantage, as it provides a clear answer to the question we formulated in section 1.1, concerning the problematic demarcation of morphology and syntax. However, while we acknowledge that phenomena of pure morphology (‘morphology by itself’) do increase the complexity of morphology as a whole because they have no external motivation, morphology by itself, as it has been theorized, only includes inflection. This would imply that only inflection counts as the locus of complexity and it is a matter of fact that most of the literature published on this topic is exclusively devoted to inflection (see Baerman et al. 2015a, 2017; Baechler 2017). Definitions of morphological complexity (in quantitative terms) such as the number of morphosyntactic features that a language has and the morphological means that are used to realize these features (see below) conform to this view, for morphosyntactic features are typically realized by inflection.

As a matter of fact, work on the complexity of word formation processes is virtually missing in the literature, the only two exceptions known to us being a one-paragraph section in Nichols et al. (2006: 101–3) and Stump (2017: 70), each. Therefore, there is no study investigating whether inflection or word formation differ in their degree of complexity along one or another parameter. As Franz Rainer (personal communication, 2017) observes, ‘a great number of asymmetries emerge between word formation and inflection with respect to different dimensions of complexity’, such as the number of elements in the system, number of affixes in a word, or the complexity of allomorphy, among others. However, he notices, ‘in the literature on the inflection-derivation divide (cf. Štekauer 2015), complexity has not been identified up to now as a possible dimension along which these two subcomponents of morphology might differ’. Lack of work on this specific topic might be due to multiple reasons: first, the boundaries between inflection and word formation are often fuzzy; second, word formation, with lexical enrichment as its central function and all its corollaries (e.g., importance of encyclopedia, semantic drift), is less neat and less automatic than inflection and more difficult to grasp (see Kusters 2003: 14–16); third—and crucially—the generally adopted metrics of morphological complexity (see section 1.3) mostly focus on formal criteria, thus lumping together categories of inflection and those of word formation under the general heading of *morphological* complexity. As we will see in more detail below, research in particular by Dahl (2004, 2009) and Trudgill (2009, 2011) has identified three major ingredients of synchronic

morphological complexity, which seem to apply to both inflection and word formation: (a) irregularity (e.g., allomorphy); (b) morphosemantic and morphotactic opacity (such a fusion of formatives, cumulative or portmanteau formatives, suppletion and non-linear suprasegmental feature realizations); and (c) syntagmatic redundancy (e.g., pleonastic affixation, see Gardani 2015).

1.3 How many complexities?

As we have seen in section 1.1, the linguistic literature on complexity is abundant, not least because ‘[h]ow to measure morphological complexity is itself an issue of some complexity’ (Nichols 1992: 64). As Miestamo (2017: 229) has appropriately noticed, complexity refers either to ‘something that is rich in internal composition (i.e. contains many parts as well as multiple and intricate connections between them), or to something that is difficult to do or to understand.’ In the first case, complexity is an objective property of a linguistic system and therefore labeled ‘objective complexity’ (Dahl 2004: 2) or ‘absolute complexity’ (Miestamo 2008) or ‘formal complexity’ (Stump 2017); in the second case, complexity is conceived as cost/difficulty that a given linguistic system or structure causes to language users and labeled ‘relative complexity’ (Miestamo 2008, 2017) or ‘psycholinguistic complexity’ (Stump 2017). In the following, we will adopt Stump’s terminology.

1.3.1 Formal morphological complexity

Formal complexity can be subsumed under the following general definition of complexity provided by the philosopher Nicholas Rescher: ‘Complexity is first and foremost a matter of the number and variety of an item’s constituent elements and of the elaborateness of their interrelational structure, be it organizational or operational’ (Rescher 1998: 1). In linguistics, we identify three principal directions in research on formal complexity, in terms of how it is conceptualized and measured: (1) quantitative approaches; (2) qualitative approaches; and (3) information-theoretic approaches.

Quantitative approaches conceive complexity in terms of the number of elements of which a given morphological entity consists, mainly inventory size and string length, or alternatively, the length of the rules necessary to describe a form. This quantitatively construed type of complexity, dubbed ‘enumerative complexity’ by Ackerman & Malouf (2013), is detectable both syntagmatically and paradigmatically. On the syntagmatic axis, it can be the before-mentioned average number of morphemes per word form (Greenberg 1954, 1960) or the maximal number of inflectionally expressed categories per verb (Bickel & Nichols 2005); this type corresponds to Rescher’s *constitutional complexity*, viz. the ‘[n]umber of

constituent elements or components’ (Rescher 1998: 9). On the paradigmatic axis, enumerative complexity relates to the number of distinct inflectional classes for a given part-of-speech (i.e., allomorphy) or the number of cells in a paradigm corresponding to the realizations of different values of a given morphological feature (e.g., case); this type of complexity corresponds to Rescher’s *taxonomical complexity*, the ‘[v]ariety of constituent elements, i.e., number of different kinds of components in their physical configuration’ (Rescher 1998: 9). Up to fairly recent times, only enumerative complexity had featured prominently in the literature, especially in typologically oriented research; for example, it is only this kind of complexity that is represented in WALS (Haspelmath et al. 2005; Dryer & Haspelmath 2013), certainly due to practical reasons. In this respect, it is worth mentioning several works specifically addressing the issue of enumerative paradigmatic complexity, such as Rhodes (1987) on the different morphological makeup of large and small paradigms and a whole series of works by Carstairs-McCarthy, whose aim was to find constraints on enumerative complexity of inflectional classes in terms of the number of affixal allomorphs and their properties (see Carstairs 1983; Carstairs-McCarthy 1994, 1998, 2010). Another type of quantitative measure concerns not the number of the elements composing a morphologically complex form but rather the (minimum) size (or length) of the rules required to describe and generate such a form. This type of qualitative approach, often referred to as Kolmogorov complexity, resonates with the Rescher’s concepts of both *descriptive complexity* (the ‘[l]ength of the account that must be given to provide an adequate description of the system at issue’) and *generative complexity* (the ‘[l]ength of the set of instructions that must be given to provide a recipe for producing the system at issue’, Rescher 1998: 9) (cf. Dahl’s ‘minimum description length’, Chapter 13, this volume).

Qualitative approaches conceive complexity in terms of identifying those morphological patterns/elements that are complex or more complex than others. Proponents of qualitative approaches need to stipulate an unmarked, complexity-neutral ideal—a canon, often conceived as an isomorphic relation of content to form—upon which to construe hierarchies of complexity in terms of degrees of deviation from it. Most notably, work by Corbett (e.g., 2007, 2015) has propagated the notion of non-canonicity (both in inflection and derivation), which can be defined as any deviation from properties such as transparency, regularity, and form-function biuniqueness, as is manifested, for example, in non-phonological allomorphy of affixes and stems (Baerman et al. 2017: 100–7), overabundance (Thornton 2019), multiple (extended) exponence (Harris 2017), syncretism (Baerman et al. 2005), defectiveness (Baerman et al. 2010), and poly-functionality (Stump 2016: 228–51), let alone more dramatic deviations such as suppletion (Stump 2006a; Corbett 2007) or deponency (Baerman et al. 2007). Early discussions of non-canonicity and its possible interactions with enumerative complexity can be found in Plank (1986) and Carstairs (1987) in addition to

works already mentioned, while recently, Johanna Nichols (2009) has hinted at a possible metric of morphological complexity related to non-canonicity (a proposal she fully develops in Chapter 7, this volume). Most studies of non-canonical phenomena in morphology have focused on the paradigmatic axis; however, nothing *per se* precludes the application of this notion to syntagmatic phenomena, such as combinatorics and mutual order of affixes (here comes to mind the distinction between semantically driven layered organization of morphology vs. opaque templatic morphology; see Stump 2006b, Good 2016), concatenative vs. non-concatenative exponence, morphophonological transparency vs. opacity and other issues belonging to the domain of morphotactics. It remains an empirical as well as a conceptual question, though, which kind of morphotactic organization should be considered ‘canonical’ and ‘less complex’. For instance, in languages where affix order directly reflects semantics, it is usually possible to permute certain affixes depending on their mutual scope (Rice 2011; Mithun 2016); whether such deviations from fixed ordering constitute additional complexity is not at all obvious.

While teleologically different, also Natural Morphology (Dressler et al. 1987; Dressler & Kilani-Schoch 2016; Dressler 2019) is centered on the idea of deviation from a core.³ Aiming at accounting for morphological preferences based on extralinguistic motivations, it theorizes a semiotically derived notion of naturalness, defined as the immediate, most unmarked, cognitively easiest, and thus universally preferred option. Conversely, naturalness-defining criteria determine deviation from the (most) natural option. This framework makes clear that other factors come to play a role in the conception and interpretation of morphological complexity, such as, for example, transparency vs. opacity of forms or morphotactic rules. As Hengeveld & Leufkens (2018: 141) observe, ‘languages may be complex, yet transparent, or simple, yet opaque’. To take the concrete case, the Turkish vs. Lithuanian data in Table 1.1 show that Turkish morphology is more complex in the sense that a single word form may potentially contain a high number of morphemes. At the same time, however, it is transparent in that every morpheme corresponds to one fixed meaning, while Lithuanian morphology is more opaque. In the framework of Natural Morphology, Dressler (2011) views unnaturalness as a source of complexity and morphological complexity as the sum of all morphological categories, rules, and inflectional classes of a language, including both productive and unproductive patterns. Distinguishing between productive and unproductive patterns, he considers morphological complexity a hyperonym of morphological richness, which is conceived only in terms of productive patterns (Dressler 2003: 47; see also Dressler, Kononenko, et al.

³ Note that, while qualitatively oriented, both Natural Morphology and Canonical Typology are implicitly able to quantify degrees of complexity, computing the degree of deviation from the natural core or canon, respectively.

2019). This distinction between active and static parts of morphology, is, in our view, not only of crucial importance with respect to psycholinguistic approaches to complexity but also foundational of approaches focused on predictability, as we will see below.

Finally, information-theoretic approaches play down the role of combinatorics and construe morphological complexity in terms of predictability and entropy. Their development is intimately related to word-and-paradigm models of morphology, which consider inflectional systems as networks of implicative relations holding between fully-inflected word forms. Consequently, they aim to understand to what extent the choice of exponence for a given cell is predictable from any other information available to the speaker, with complexity being in an obvious inverse relation to predictability (cf. Finkel & Stump 2007, 2009; Stump & Finkel 2013). Ackerman & Malouf (2013) propose the term ‘integrative complexity’, based on the notion of entropy as ‘a measure of the reliability of guessing unknown forms on the basis of known ones’, that is, a measure of predictability. They move from the intuition that ‘speakers must generalize beyond their direct and limited experience of particular words’ (p.436) and posit a ‘Low Entropy Conjecture’: morphological systems, such as paradigms, in which conditional entropy among related word forms is low, are more efficient, as they ‘permit these crucial inferences to be made easily’ (p. 436) (cf. ‘Paradigm Structure Conditions’ of Wurzel 1989).⁴ In other words, complexity derives from opaque intraparadigmatic relations, for opacity hampers the predictability and predictiveness among word forms in a lexeme’s paradigm. The ‘Low Entropy Conjecture’ is supported by recent studies on inflection class systems clearly violating the enumerative complexity-based constraints of the kind proposed by Carstairs-McCarthy (see Baerman 2012, 2016; Sims 2015).⁵

The approaches to formal morphological complexity surveyed thus far share the potential to seize the degree of complexity. However, some typological studies have pursued the topic without a focus on metrics. One line of investigation, for example, has concerned the relation of (certain aspects of) morphological complexity to any other typological parameters such as phonological systems (Shosted 2006; Fenk-Oczlon & Fenk 2008, 2014), word order (e.g., Sinnemäki 2008; Bentz & Christiansen 2013), among others. Other studies have focused on the differential elaboration of nominal and verbal morphology (e.g., Nichols 1986, 1992; Mithun 1988; Kibrik 2012). In this domain, there are still more open questions than established answers, partly because of the lack of consensus as regards the

⁴ Also morphomic stem distributions have been interpreted in terms of predictive relations by Blevins (2016b: 123), a view partly criticized by Maiden (2018: 23–4).

⁵ It is likely that a conception of complexity based on entropy applies better to inflection than word formation because inter-word relations are generally much more complex in inflectional than in derivational paradigms.

definition of the relevant aspects of complexity and the adequate ways of its measurement.

Still another line of research is concerned with the relation between morphological complexity and sociolinguistic typology. In section 1.1, we already mentioned the idea that pidgins and creoles are in general less complex than languages with a long history and uninterrupted transmission. More generally, in recent work (e.g., Trudgill 1997, 2009, 2011, 2017; Kusters 2003, 2008; McWhorter 2007, 2008; Lupyán & Dale 2010; Bentz & Winter 2013; Bentz et al. 2015; Bentz 2016), claims have been advanced that the overall degree of complexity as well as certain particular types of grammatical complexity correlate with such socioecological conditions of language use as high vs. low degree of contact, number of adult learners, size and geographic expansion of the speaker population, and some others (see also Tinitis 2014 for a behavioural experiment with a miniature artificial language). Significantly, most of such studies have focused on simplification caused by language contact (see Dorian 1978; McWhorter 2001; among many others), emphasizing that morphological complexity requires long-term periods of socioecological stability to develop (Dahl 2004). Nevertheless, studies exist showing that certain types of language contact (e.g., those involving stable childhood multilingualism) can contribute to preserve complex patterns (Trudgill 2011; Mithun 2015) and even result in increase rather than loss of morphological complexity due to borrowing and contact-induced grammaticalization (see Vanhove 2001; Aikhenvald 2002, 2003a; de Groot 2008; Loporcaro 2018; Loporcaro et al. forthcoming). Also processes of language genesis brought about by language contact do not necessarily come along with morphological simplification. In a study on the rapid birth of a new mixed language in Australia, Gurindji Kriol, from the admixture of Gurindji and Kriol, Meakins et al. (2019) demonstrate that there was no preferential adoption into Gurindji Kriol of less complex variants and that, in fact, complex Kriol variants were more likely to be adopted than simpler Gurindji equivalents. Given that Gurindji Kriol is the primary language of the younger generation in the Gurindji community, Meakins et al. interpret these results in light of the fact that the acquisition of morphology in morphologically complex languages is less challenging for children than for adults (cf. also Miestamo 2008). The issue of ease vs. difficulty of processing in language acquisition leads us over to the second main type of morphological complexity introduced in section 1.3, viz. psycholinguistic morphological complexity.

1.3.2 Psycholinguistic morphological complexity

As we have seen in the previous section, also Natural Morphology and Ackerman & Malouf's (2013) integrative complexity appeal to ease in processing and

production, as a key to the interpretation of what is complex in morphology. These models build a bridge to the second type of approach to morphological complexity, psycholinguistic morphological complexity, that focuses on the cost/difficulty that a given linguistic system or structure causes to language users, that is, computational effort. Psycholinguistic approaches to morphological complexity assume that the degree of ease vs. cost of a morphological pattern in processing and production correlates with its degree of complexity. This line of research draws evidence from three areas of study: adult processing, L1 and L2 acquisition, and the performance of artificial automatic learning.

One line of investigation within this field has developed around the equation of complexity with low parsability (Stump 2017). In this respect, the debate on the balance between memory retrieval and online computation in language production is particularly relevant. In the context of the debate on lexical access and specifically of the so called English past-tense debate (for references, cf. Ambridge & Lieven 2011: 169–87), Pinker & Prince (1988) argued for a ‘dual-route’ model that could account for both irregular forms (*feell/felt*), which are memorized as wholes in the mental lexicon, and an online rule of default responsible for morphemic concatenation (*walk/walked*) (see also Gardani et al. 2019: 24–7). At the same time, it was observed that regular forms with high frequency can also be stored in the mental lexicon (Alegre & Gordon 1999a: 56). However, the fact that both morphologically less complex (i.e., highly parsable) and morphologically complex (i.e., low parsable) word forms can be lexically stored leads to concluding that complexity *qua* parsability does not correlate with processing cost. The role of frequency in lexical access has been stressed by nobody else as vigorously as by Joan Bybee (1985, 1995, 2007). Consequently, the conception of complexity focusing on system complexity, in which irregularity is viewed as an ingredient of complexity, is incompatible with the results of studies on processing complexity, which have shown that irregularity does not *per se* constitute an obstacle for the language user, as it can be defeated by frequency.

Studies in language acquisition, too, do not necessarily support the hypothesis that psycholinguistic complexity and formal complexity coincide. For example, in a crosslinguistic study on the relationship between the morphological complexity of child-directed speech and the speed of morphological acquisition in children, Xanthos et al. (2011) found a strong positive correlation between inflectional complexity of the input and the speed of acquisition. This result seems to suggest that the more morphology in the input, the easier the morphology is to acquire. According to Kelly et al. (2014), formal complexity such as heavy synthesis in polysynthetic languages is not a challenge for L1 acquisition if the templatic sequence in which formatives are used is regular, and Allen (2017) also reports longitudinal studies showing that Inuit children acquire elaborate derivational and inflectional morphology early and with ease. (See also Stoll et al. 2017, on the acquisition of verb morphology in polysynthetic Chintang.) Other acquisitional

studies construe formal complexity not as *constitutional complexity* but as *descriptive complexity*. For example, in a crosslinguistic study on the emergence and early development of synthetic compounds, Dressler, Sommer-Lolei, et al. (2019) provide evidence that synthetic compounds (i.e., compounds in which the head is derived from a verb and the non-head is an argument of this verb) such as German *Nussknacker* ‘nutcracker’ are acquired later than comparable three-constituent compounds. They interpret this later acquisition as a sign of higher complexity: equating the degree of complexity with the number of rules involved, synthetic compounds, which are derived by both a rule of compounding and a rule of derivation, are more complex than words derived either only by compounding or only by derivation rules.

Besides that, numerous studies, both typological and experimental (e.g., Wray & Grace 2007; Lindström 2008; Trudgill 2011; Bentz et al. 2015; Bentz & Berdicevskis 2016; Atkinson et al. 2018), show that morphological complexity, while being an obstacle to L2 acquisition in adults and hence subject to erosion, regularization, and loss in those situations of language contact that involve massive adult acquisition, does not, in fact, constitute a severe challenge for L1 acquisition in children. Moreover, Lupyan & Dale (2010) have hypothesized that infants, in fact, benefit from the increased redundancy brought about by morphological complexity in languages used in small groups.

Psycholinguistic approaches to morphological complexity have attracted criticisms mainly of two sorts. One problem is that the perception of ease or, conversely, difficulty, might vary among language users, and therefore might not be an objective metric; the other problem is that ‘psycholinguistic background research on the processing cost and learning difficulty of a given grammatical phenomenon’ might not be enough (Miestamo 2017: 232). As a matter of fact, the correlation between ‘our intuitive notion of morphological complexity and actual evidence of the pace of acquisition of more or less complex inflectional systems in child language’ (Marzi et al. 2018) seems to be poor. In order to solve at least the objectivity issue, recent research in morphological complexity has expanded into the field of neurobiologically inspired computational models of processing and learning. In one such study, Marzi et al. (2018) have focused on the performance of recurrent self-organizing neural networks trained to learn languages, in order to understand how degrees of inflectional complexity affect word processing strategies. They found a significant systematic correlation between regularity and predictability of verb forms and interpret the evidence ‘as the result of a balancing act between two potentially competing communicative requirements’, viz. recognition (leading to a maximally contrastive system) and production (leading to maximally predictable forms).

1.4 About this volume

In section 1.1, we identified four issues we deem among the most urgent to solve in research on morphological complexity. In order to tackle these issues in a principled way, we convened a dedicated workshop ‘Morphological Complexity: Empirical and Cross-Linguistic Approaches’ at the 48th Societas Linguistica Europaea (SLE) meeting in Leiden in 2015. The present volume is a collection of original research papers consisting in equal measure of papers delivered at the workshop and of invited contributions. (Each chapter was subject to a threefold reviewing process consisting of an anonymous external reviewing, a non-anonymous internal review performed by a fellow contributor, and comments by the editors.) The volume features: (a) various theoretical, methodological, and typological perspectives on morphological complexity (from ‘classic’ morphological description to experimental and information-theoretic approaches); (b) both detailed investigations of individual languages and wider crosslinguistic studies; (c) synchronic and diachronic analyses; (d) a broad coverage of topics including structural and sociolinguistic issues, such as the development of morphological complexity under different sociohistorical conditions (prominently, language contact); (e) empirical evidence drawn from languages from all continents and belonging to a number of typologically diverse language families. Unfortunately, the volume does not cover the complexity of word formation and the complexity of sign language morphology. We hope that future research will take care of these issues.

The volume, introduced by the present chapter, consists of three parts organized according to the chapters’ main focus and scope, and is closed by a discussion in Chapter 13 by Östen Dahl on the volume’s contributions and on the minimum description length approach. Part I includes five chapters dealing with issues of morphological complexity from a language-specific perspective. **Jeff Parker** and **Andrea Sims**’s Chapter 2, ‘Irregularity, paradigmatic layers, and the complexity of inflection class systems: A study of Russian nouns’ follow Stump & Finkel’s (2013: 55) definition of complexity of an inflection class system as ‘the extent to which the system inhibits motivated inferences about a lexeme’s full paradigm of realized cells [...]’. Using data from Russian, the authors explore the implications of gradient (ir)regularity for measuring and comparing the complexity of inflection class systems. They find that some, but not all, less regular inflectional patterns significantly increase the complexity of the system, but that the increased complexity is mitigated by structural and distributional properties of the inflectional system. In Chapter 3, ‘Demorphologization and deepening complexity in Murrinhpatha’, **John Mansfield** and **Rachel Nordlinger** investigate diachronic changes in the complexity of verb inflection in Murrinhpatha, a polysynthetic non-Pama-Nyungan language of northern Australia, which displays a high level of

complexity in terms of unpredictable analogical relations in inflectional exponence. The authors demonstrate that recent changes in inflection allomorphy blur the boundaries of stem and affix, resulting in gradual demorphologization and increasingly unpredictable exponence. **Felicity Meakins** and **Sasha Wilmoth**'s Chapter 4, 'Overabundance resulting from language contact: Complex cell-mates in Gurindji Kriol' examines the development of overabundance (see above) in the subject-marking system of Gurindji Kriol, an Australian mixed language. By means of generalized linear mixed models, which probabilistically measure the use vs. non-use of a feature, the authors interpret the insurgence of overabundance as an instance of complexification, providing a counterexample to the commonly held view that contact always results in reduction of morphological complexity. In Chapter 5, 'Derivation and the morphological complexity of three French-based creoles', **Fabiola Henri**, **Gregory Stump**, and **Delphine Tribout** take a fresh look at a controversial assumption in creole research, namely the widespread claim of poverty of creole morphology (see references in section 1.1). Analysing deverbal nominalizations via conversion in Mauritian, Guadeloupean, and Haitian, and assessing the integrative complexity of the respective morphological systems' derivational relations, the authors demonstrate that the complexity of the derivational relations in these creoles attains the same degree as those of the lexifier, French. Finally, in Chapter 6, 'Simplification and complexification in Wolof noun morphology and morphosyntax', **Michele Loporcaro** explores the diachronic dynamics of morphological complexity in the nominal morphology and morphosyntax of Wolof, an Atlantic language of Senegal. Loporcaro shows that, while changes such as the emergence of inflectional irregularities produced a local increase in complexity in noun and determiner morphology, overall the morphology of Wolof is less complex than that of closely related Atlantic languages. Loporcaro provides an explanation of the simplifying tendencies in sociolinguistic terms, referring to the correlation between simplification and prestige in the Wolof speech community. Here, speaking correctly is associated with low-caste in rural settings, while linguistic prestige is achieved through language mixing, extensive borrowing, and, crucially, the simplification, via paradigmatic leveling, of inherited alternations impacting on both the morphology and the morphosyntax of the language.

Part II consists of three chapters approaching morphological complexity from a crosslinguistic perspective. **Johanna Nichols**'s Chapter 7, 'Canonical complexity' considers not size but non-transparency the locus of morphological complexity and adopts the notion of (non-)canonicity to define crosslinguistically comparable variables, capture non-transparency, and restrict the comparanda to a manageable sample. **Francesca Di Garbo**'s Chapter 8, 'The complexity of grammatical gender and language ecology' is a crosslinguistic investigation of the evolution of gender agreement patterns, which are viewed as an instance of morphological complexity, and its ties to sociohistorical factors. Analysing a sample of thirty-six languages in

a qualitative fashion, the author is able to establish association between multiple patterns of change, such as loss, reduction, emergence, and expansion of gender, on the one hand, and various sociohistorical situations, ranging from demographic structure (population size) to language policies and language attitudes, on the other. In Chapter 9, 'Morphological complexity, autonomy, and areality in western Amazonia', **Adam Tallman** and **Pattie Epps** investigate the relationship between morphological complexity and areality-building processes across Amazonia. The authors observe (a) morphological proliferation in four domains (nominal classification, tense, evidentiality, and valency-adjusting mechanisms) across unrelated western Amazonian languages; (b) high system complexity across these domains; and (c) a link between complexity and language contact. They conclude that factors often associated with morphological complexity are in fact not necessarily morphological, as a large percentage of bound morphemes in these languages display ambiguity between morphology and syntax.

The three chapters in Part III address the problem of morphological complexity from an acquisitional perspective. In Chapter 10, 'Radical analyticity as a diagnostic of adult acquisition', **John McWhorter** proposes that languages can become radically analytic, that is, completely or near-completely void of inflectional morphology, only via incomplete acquisition. He draws evidence from West Africa and Southeast Asia and shows that the relevant languages score more like creoles than like older languages. In McWhorter's view, second-language acquisition decisively reduces grammatical complexity (in terms of bound inflection) to a degree that ordinary language change cannot. The author suggests that radical analyticity can be treated as evidence that such second-language acquisition occurred in the history of the language, and thus, synchronic morphological complexity can serve as a clue to the past of a language, in the absence of historical documentation. Also Chapter 11, 'Different trajectories of morphological overspecification and irregularity under imperfect language learning' by **Aleksandrs Berdicevskis** and **Arturs Semenuks** deals with imperfect language learning, partly supporting McWhorter's conclusion. By reference to the editors' fourth question (see section 1.1), the authors investigate how morphological complexity is related to socioecological parameters. They run an iterated artificial language learning experiment, tracing the change of two facets of complexity: overspecification and irregularity. They find that the presence of imperfect learners in a transmission chain leads to a much stronger decrease in morphological overspecification. Overspecification, however, is not usually fully eliminated, and its partial decrease often leads to increased irregularity, thus making languages simpler in one respect, but more complex in another. Additionally, higher irregularity decreases learnability, and this effect is stronger for imperfect learners compared to normal learners. Thus, the relationships between these two facets of morphological complexity and language learnability have their own complexities. Finally, **Marianne Mithun's** Chapter 12, 'Where is morphological complexity?' is firmly anchored in the debate on the

psycholinguistic reality of complexity. Examining the speech of native speakers of two North American languages influenced to varying degrees by contact with English, Mithun observes that even native speakers with limited proficiency produce morphological structures that are highly complex for the analyst, with large numbers of morphemes per word, fusion, and irregularity. She argues that the distinction between what linguists consider complex and what speakers find difficult (or easy) to acquire or preserve, is not surprising if one takes the view that morphology in these languages is not processed and learned online, but rather in chunks.

As we said, **Östen Dahl** closes the volume by critically reviewing the volume's chapters and seeing how the concepts of morphological complexity applied therein relate to the 'minimum description length approach'.

Turning now to the four research questions (section 1.1) the contributors to this volume focused on, we observe that (question 1) it is possible to define morphological complexity, even though the demarcation between morphology and syntax is in many cases fuzzy (see Tallman & Epps, Chapter 9, this volume). At the same time, however, we observe that different authors provide and apply different definitions, also within this volume. Seemingly, the very existence of multiple definitions of morphological (and morphosyntactic) complexity is related not only to the collocation of a specific linguistic feature along the grammar continuum (from pure morphology to morphosyntax), but also to the diversity of phenomena and types of complexity. This observation leads us to answer question 2, namely whether it is possible to arrive at a uniform notion of morphological complexity. We concur with Dahl (Chapter 13, this volume), that a set of shared notions and standard works that everybody refers to has not yet been reached. Thus our answer to question 2 is no, and the motivation for it is that the linguistic facts are so multifarious and diverse that not one, but many different complexities can be detected (whence the plural in this chapter's title).

Then we asked (question 3) whether it is possible to arrive at a crosslinguistically applicable and theoretically founded measure of morphological complexity. Berdicevskis et al. (2018) have recently pointed to the absence of a gold standard. We, too, have observed that there exists neither a commonly accepted definition of morphological complexity nor a uniform measure thereof. Admittedly, the growing understanding of the multifaceted nature of morphological complexity is much in line with the multivariate nature of typological comparison. So, perhaps we asked the wrong question. Probably, the quest for a unique measure is an epistemological fallacy. Once we have acknowledged that there is not one morphological complexity, but many morphological complexities, we should identify *a set* of complementary specific measures to apply crosslinguistically. Then, the only reasonable typological approach to morphological complexity is to break it down into individual variables (if necessary, each with its quantitative measure) and then look for mutual correlations between such variables or for their connections with other parameters of crosslinguistic variation. Of course, cumulative

measures such as the one developed by Nichols (Chapter 7, this volume) are also possible, but they are not holistic, either, and in many cases are based on a significant reduction of empirical data.

In conclusion (question 4), we wanted to investigate the role of such extra-morphological factors as diachronic development and (in)stability, susceptibility to loss vs. spread in situations of language contact, and, generally, of sociolinguistic and socioecological parameters, in affecting morphological complexity. As several chapters in this volume have demonstrated, in spite of at times diverging results, the study of the correlation between morphological complexity and extralinguistic factors such as the role of language contact or speakers' sociolinguistic attitudes, is fruitful and promising.

Of course, the answers we have provided here are *per force* partial and by far not definitive, as much more case studies and comparative evidence are necessary to get to a reliable picture of such complex phenomena as morphological complexities. We hope that future research will pursue these pathways.

Acknowledgements

The volume's editors wish to thank the authors, the external reviewers, and our editors at OUP. The support of the Swiss National Science Foundation (SNF CRSIII_160739) is gratefully acknowledged. Besides that, we thank Aleksandrs Berdicevskis, Wolfgang Dressler, Michele Loporcaro, and Franz Rainer for their insightful comments on a preliminary version of this introductory chapter.