PART III

MORPHOLOGICAL THEORY AND OTHER FIELDS

CHAPTER 21

MORPHOLOGICAL THEORY AND TYPOLOGY

PETER ARKADIEV
AND MARIAN KLAMER

21.1 Introduction

Morphology as the study of word structure is intimately related to both language description and linguistic theory. Both these enterprises should be informed by cross-linguistic variation in the domain of morphology, albeit for different reasons. The task of a field-worker or grammar-writer is to describe and interpret the morphological structure of an individual language as adequately as possible, including intricate details and idiosyncrasies. The task of a theoretical linguist, on the other hand, is to construct an empirically and explanatorily adequate model of language in general, or morphology in particular. Both descriptivists and theoreticians thus have to be aware of the range of morphological phenomena occurring in languages, and of the attested cross-linguistic diversity. In the ideal situation, they should also have access to information on the frequencies of certain cross-linguistic patterns, and on the genealogical, areal, and structural distributions of these patterns. The aim of morphological typology, as part of the broader linguistic typological enterprise, is to map the cross-linguistic variation and unity found in the domain of word structure, and to link this to other independently established typological generalizations.

The typological study of morphology faces several challenges, the most important of which is the very nature of the empirical domain. As Baerman and Corbett (2007: 115) put it, "[o]f all the aspects of language, morphology is the most language-specific and hence least generalizable. Indeed, even the very presence of a meaningful morphological component is language-specific". Given this, it is hard to make statements about morphology that are cross-linguistically valid. Even comparing morphological phenomena in different languages requires the typologist to carefully devise and cautiously apply analytical notions and methods. Comparative notions cannot be directly "borrowed" from descriptive studies of individual languages. Such commonly accepted notions as "root", "affix", "lexeme", "paradigm", and the very notion of "word" itself, have proven to be notoriously difficult

to define in a cross-linguistically valid way (see §21.2). The current state of research has to acknowledge the fundamental problem that none of these notions can be applied cross-linguistically to yield consistent results throughout.

Typology has often been associated with the quest for language universals. However, from the outset it has also been clear that the study of rare and unique patterns is as important as the study of cross-linguistically recurrent ones (see e.g. Plank no date; Wohlgemuth and Cysouw 2010). This is especially true for morphology, where many, perhaps most, of the attested patterns are rare, or obviously non-universal. However, cross-linguistically unique patterns can be and usually are revealing of the range of possibilities open for human language structures, and reflect—albeit in a paradoxical way—potentially universal patterns admittedly common to all languages. To give a striking example, the Australian language Kayardild (see Evans 1995; Round 2013) overtly marks clausal morphosyntactic features, such as case role, tense, and mood, on each word of a relevant constituent, cf. example (1), where the Instrumental case appears not only on the head of the noun phrase but on its Genitive modifier, too, while the Ablative and the Oblique suffixes mark past tense and epistemic modality, respectively.

(1) Kayardild (Tangkic, Northern Australia; Evans 1995: 115)[1]

a. *dangka-**karra-nguni*** *mijil-**nguni***
man-GEN-INS$_1$ net-INS$_1$
'with the man's net'.

b. *maku* *yalawu-jarra* *yakuri-**na*** *dangka-**karra-nguni-na*** *mijil-**nguni-na**.*
woman catch-PST$_2$ fish-ABL$_2$ man-GEN-INS$_1$-[illegible] net-INS$_1$-ABL$_2$
'The woman caught some fish with the man's net.'

c. *maku-**ntha*** *yalawu-jarra-**ntha*** *yakuri-**naa-ntha***
woman-OBL$_3$ catch-PST$_2$-OBL$_3$ fish-ABL$_2$-OBL$_3$
*dangka-**karra-nguni-naa-ntha*** *mijil-**nguni-naa-nth**.*
man-GEN-INS$_1$-ABL$_2$-OBL$_3$ net-INS$_1$-ABL$_2$-OBL$_3$
'The woman must have caught fish with the man's net.'

Except for the closest relatives of Kayardild, this phenomenon is not attested in any other language. This unique feature of Kayardild shows a logical and beautifully iconic mapping of the hierarchical structure of syntax on the morphological structure of words, which is largely obscured in other, less "exotic" languages. Unique patterns like this one might well turn out to be no less instructive for linguistic theory than cross-linguistically recurrent ones. Moreover, typological *rara* are crucial for morphological description, since morphology is precisely the domain where irregular, idiosyncratic, and unfamiliar phenomena are most expected to occur. All of these phenomena require accurate, detailed, and unbiased documentation.

The aim of the present chapter is to present a concise overview of the current state of typologically-oriented research in morphology, and to suggest ways in which morphological typology and theory can enrich each other. While we address both empirical and methodological issues, we refrain from discussing the technical details of any particular

[1] Glossing is slightly simplified; coindexation indicates "concord" relation between inflections.

theoretical framework. None of the current morphological theories is probably able to equally adequately account for the plethora of morphological phenomena attested in the world's languages, but most of them have contributed significantly to our understanding of many of these phenomena.

Morphology is "the grammar of words" (cf. Booij 2005d). In what follows, we first discuss the notion of "word" and the issues surrounding it in §21.2. The primary goal of morphological typology and theory is to analyze the ways in which languages establish relations between forms and meanings when they build words, and to discover the principles underlying the cross-linguistic variation in this domain. This relation between meaning and form in morphology is the topic of §21.3. Another important domain of morphological inquiry are the syntagmatic and paradigmatic relations between words and their components. In §§21.4 and 21.5, we briefly review empirical and theoretical issues relating to the syntagmatic and paradigmatic dimensions of cross-linguistic diversity in morphology.

21.2 The notion of "word" and its problems

As the notion "word" is central to morphology, its definition and identification are crucial both for morphological analysis and morphological typology. There are two relevant understandings of "word". On the syntagmatic axis, we have to distinguish wordforms from phrases and parts of words (i.e. morphemes), while on the paradigmatic axis we need to identify lexemes, that is, sets of wordforms sharing lexical meaning and differing in the values of inflectional features only. Both understandings of "word" create their own problems, which will be discussed in turn in §§21.2.1 and 21.2.2.

21.2.1 Is "wordform" a typologically valid concept?

Bloomfield (1933: 178) defined "word" as the "minimal free form". However, it has proven to be notoriously difficult to identify what precisely a "minimal free form" is, especially in languages that have no written tradition and are not used in formal education. Moreover, some languages have numerous lexical items denoting various events or activities of verbal communication, but lack a word for 'word', for example Kambera in (2).

(2) Kambera (Austronesian; Sumba, eastern Indonesia; Onvlee 1984; Klamer 1998)
hilu 'a verbal exchange; a language'
lí 'a sound, a story, an event, a tradition; to speak'
luluk 'a proverb, a speech'
langu 'a message, something that is being talked about, a situation'
pulung 'an advice, an order, a judgment, a gossip; to gossip'
kareuk 'to talk'
reu 'sound of talking'

Wordforms in different languages can only be identified using structural criteria, both phonological and morphosyntactic (see e.g. Dixon and Aikhenvald 2002a; Julien 2006). Most of these criteria are language-specific, and often they yield conflicting results even in the same language (Haspelmath 2011; van Gijn and Zúñiga 2014). It is necessary to keep in mind that phonological criteria (such as the assignment of primary stress, the tonal contour, or the domain of phonological phenomena like vowel harmony or sandhi) identify phonological words which do not always align with grammatical or morphosyntactic words (cf. Bickel and Nichols 2007: 172–4; Bickel and Zúñiga 2017).

The morphosyntactic word is the unit that pre-literate speakers most often associate with the term "word". It is the minimal response that speakers would give to a question like "what is the name for that [*pointing at object*] in your language?". It is usually also the smallest linguistic unit that can be subject to such syntactic operations as coordination, movement (e.g. in questions) or ellipsis. This is accounted for by the Principle of Lexical Integrity proposed in certain formal theories of grammar (e.g. Di Sciullo and Williams 1987; Spencer 2000a; Montermini Chapter 7, this volume); according to this, syntax cannot manipulate the internal structure of words.[2] The morphosyntactic word is also the unit that is the outcome of morphological word-formation processes, and the basic unit used by speakers to build more complex expressions (i.e. syntactic phrases). It is also the unit on which speakers typically apply self-repair when they are telling a story or having a conversation. For instance, when mispronouncing a word, a speaker's self-repair will often involve repeating the entire morphosyntactic word, rather than a part of it (cf. e.g. Wouk 2005; Podlesskaya 2015: 72–3; cf. Fox et al. 2017 for a typological study).

Phonological words can be preceded and/or followed by conscious and deliberate pauses and intonation breaks, while speakers seldom make such breaks in the middle of them. This does not mean that a natural text will not contain word-internal breaks or pauses; indeed, all natural texts contain hesitations, self-repairs, and false starts occurring in the middle of words. However, speakers are normally able to recognize these as "errors" when they listen to the recording, and they consider the utterance without an internal break or pause as the "correct" form.

Despite the theoretical and practical importance of the notion of morphosyntactic word, different diagnostics do not always converge. Well-known cases are the German, Dutch, and Hungarian separable verbal prefixes (see e.g. Ackerman and Webelhuth 1998: Ch. 10; Müller 2003; Zeller 2004 on German; Booij 1990, 2002b on Dutch; Ackerman 2003, Ackerman and Webelhuth 1997 on Hungarian), illustrated in (3). On the one hand, preverbs such as German *aus* 'out', *an* 'at', or *ein* 'into' (3a–d) form a tight semantic and syntactic unit with the verb following them, which is reflected in the orthography (3a),—a compound, as evidenced by the stress pattern of the preverb+verb complex, the ability of the preverb+verb complex to serve as an input to word-formational operations (German ***áus**gehen* 'go out' ~ ***Áus**gang* 'exit'), and the fact that many such combinations have idiomatic meanings and therefore must be listed in the lexicon as units. On the other

[2] However, see Baker (1988, 1996) for a model of syntax–morphology interaction apparently discounting lexical integrity, together with much work in the framework of Distributed Morphology (Siddiqi, Chapter 8 this volume). From a different perspective, Haspelmath (2011) also argues against lexical integrity as a universal principle of grammar.

hand, there is evidence that the preverb and the verb do not form a single phonological or morphosyntactic word even when adjacent, and moreover, the preverb can be detached from the verb and, in German and Dutch, be separated from it by long and syntactically complex strings of words; such free standing preverbs behave like autonomous words in that they are able to bear independent stress (3b), be focused (3c), and be coordinated (3d).

(3) German (Indo-European)
   a. *Er sagt, dass er uns ein Bier* ***áusgibt***.
      'He is saying that he is going to buy us a beer.' (Zeller 2004: 181)
   b. *Er* ***gibt*** *uns ein Bier* ***áus***.
      'He is buying us a beer.' (Zeller 2004: 181)
   c. *Ich* ***lache*** *dich nicht* ***áus****, sondern* ***án***.
      'I'm not laughing at you, I am smiling at you.' (Zeller 2004: 190)
   d. *Die Türen öffnen sich, Leute steigen* ***áus*** *und* ***éin***.
      'The doors open, the people are getting off and on.'[3]

Another issue relating to the notion of word refers to the level above the word. How can we distinguish morphologically complex words, e.g. compounds, from syntactic phrases (cf. Lieber and Štekauer 2009a)? Phrases and compounds can look quite similar because the latter often derive historically from the former. The wordhood of a compound in contrast to a multi-word phrase is often determined semantically: the meaning of a compound is typically not the sum of its parts, while the meaning of a phrase is typically regular and transparent (compositional). In addition, components of compounds usually show referential opacity, that is, they cannot on their own refer to discourse participants (see however Koptjevskaja-Tamm 2013 on an interesting case of compounds formed from personal names). In relation to this semantic compositionality, we see that parts of phrases can also be modified separately (*a very black board*), while this is not possible for the parts of a compound (**a* [*very black*]*board*). However, the semantic distinction between phrases and compounds is never categorical: languages with semantically irregular and non-transparent compounds also often have semantically regular and transparent ones, just as probably every language has phrases that are idiomatic (see e.g. Di Sciullo and Williams 1987 on the distinction between words and "listemes"). Again, phonological and morphosyntactic criteria have to be invoked in order to distinguish phrases from compounds. Thus, in English noun phrases main stress is claimed to be on the head (*a black bóard*), whereas nominal compounds have stress on the modifying element instead (*a bláckboard*), see, however, Giegerich (2009) against such a view; in German, adjectival modifiers in phrases must be inflected for gender, number, and case (e.g. *ein rot**er** Kohl* 'a red cabbage'), while this inflection does not appear in compounds (e.g. *Rotkohl* 'red cabbage'). In languages with noun incorporation, the incorporated nominal root may occur between the inflectional affixes and the root of the verb, and be subject to word-internal phonological processes, as in Chukchee, (4).

[3] http://www.hna.de/kassel/hilfe-leichter-sprache-6100722.html, accessed 21 February 2016.

(4) Chukchee (Chukotko-Kamchatkan, Russia; Mithun 2000a: 916)
a. *gam-nan tə-ntəwatə-rkən **utkucʔ**-ən.*
1SG-ERG 1SG-set-PRS trap-ABS
'I am setting a trap.'
b. *gəm t-**otkocʔə**-ntəwatə-rkən*
1SG.ABS 1SG-trap-set-PRS
'I am trap-setting.'

However, morphosyntactic criteria like these cannot be usefully applied to languages that lack phrase-internal inflectional concord, or languages that have only suffixes and no prefixes. Thus, in Persian, idiomatic noun+verb combinations (5a) are on the surface indistinguishable from verb phrases with non-specific bare nouns (5b).

(5) Persian (Indo-European > Iranian, Iran; Megerdoomian 2012: 189)

a.	*kotæk xordæn*	lit. beating eat	'to be beaten'
	færib xordæn	lit. deception eat	'to be deceived'
	šekæst xordæn	lit. defeat eat	'to be defeated'
b.	*qæza xordæn*	lit. food eat	'to eat'
	xyar xordæn	lit. cucumber eat	'to eat cucumber'
	šam xordæn	lit. dinner eat	'to eat dinner'

Even in highly inflectional languages like Russian there is a continuum, illustrated in (6), where phrases formed in syntax occupy one end (6a), unequivocal compounds with linking elements occupy the other end (6e), and cases with doubtful status occur in between (6b–d) (cf. Benigni and Masini 2010; Masini 2009; see also Booij 2010a: ch. 7 on "phrasal names").

(6) Russian (personal knowledge of P.A.)[4]

a.	*želézn-aja*	*mísk-a*	syntactic phrase (adjective+noun)
	iron-NOM.SG.F	bowl-NOM.SG	
	'iron bowl'		
b.	*želézn-aja*	*dorógˈ-a*	phrasal name (adjective+noun)
	iron-NOM.SG.F	road-NOM.SG	
	'railway'		
c.	*krésl-o=kačálk-a*		doubly-inflected noun+noun compound
	armchair-NOM.SG=rocker-NOM.SG		
	'rocking chair'		
d.	*generàl=gubernátor*		compound without a linking element
	general=governor[NOM.SG]		
	'governor-general'		
e.	*svin-**o**-férm-a*		compound with a linking element
	pig-LNK-farm-NOM.SG		
	'pig farm'		

[4] The "=" sign stands in place of the orthographic hyphen, while the hyphen indicates morpheme boundaries; the acute and the grave signs mark primary and secondary stress, respectively.

Distinguishing between compounds and phrases is especially difficult in languages where syntactic operations apparently create morphologically complex words. Thus, in Adyghe, an adjectival modifier obligatorily forms a compound with its head noun, as illustrated in (7). The resulting phrase inflects as a single unit, and forms a single domain for stress and phonological alternations (Lander 2016, 2017). Some such compounds are idiomatic, but most are formed by general syntactic mechanisms in the course of speech.

(7) Adyghe (West-Caucasian > Circassian; Lander 2017: 84)
Ø-jə-zə-šolk-ʒ'ene-daxe-r
3SG.IO-POSS-one-silk-dress-beautiful-ABS
'one beautiful silk dress of hers'

Another problematic issue in the definition of wordforms is clitics, which show properties of both words and affixes (see Bickel and Nichols 2007: 174–80, and especially Spencer and Luís 2012a, 2012b for a comprehensive discussion and references). Phonologically, clitics are not free forms, as they must attach to a host with which they form a single prosodic domain. Morphologically, they often behave like affixes in displaying fixed order and various co-occurrence restrictions and idiosyncrasies. Syntactically, however, clitics and clitic clusters show more freedom than genuine affixes, which normally attach to hosts of a particular category. Clitics may attach to the edges of a syntactic phrase, or their position is structurally defined as following the first stressed word or first phrase of a sentence (so-called 'second-position' or 'Wackernagel' clitics, cf. Anderson 1993, 2005), as in Cupeño, (8).

(8) Cupeño (Uto-Aztecan > Northern, California; Hill 2006: 72)
*hani=**qwe=n=pe*** *ilily-i* *mamayew.*
EXHORT=PTCL=1SG=IRR coyote-OBJ help.HAB
'I wish I could help Coyote.'

Despite being notoriously difficult to define and identify typologically (Haspelmath 2015), clitics, and in particular second position clitics, are an important and widely attested phenomenon. The terms "clitic" and its derivatives like "clitic doubling" or "clitic left dislocation" should however be used with caution and be clearly defined in contrast to affixes and free-standing wordforms.

In sum, the concept "word" is not simple and not clear-cut: many criteria for wordhood are applied language-specifically; some yield conflicting results in a single language, and often words in a language take different positions on the continuum going from 'word' to 'phrase'. That "word" is not a category with robust boundaries is a problem for theories built around the idea that syntax and morphology are clearly distinct modules. Some eschew the problem by deeming the very notion "word" invalid, and the distinction between syntax and morphology irrelevant for linguistic theory (e.g., Haspelmath 2011). Instead, we believe that it is worthwhile to investigate the typological space generated by various wordhood properties in order to arrive at empirically grounded generalizations about combinations of such properties and their cross-linguistic patterns (cf. Bickel and Zúñiga 2017).

21.2.2 Inflection vs. derivation and the notion of "lexeme"

Orthogonal to the problem of the definition of the wordform is the issue of the delimitation of lexemes and, consequently, of inflectional paradigms. The notion "lexeme" is roughly equivalent to a lexical entry in a dictionary. A lexeme is, by definition, a set of wordforms distinguished solely by inflectional features and their exponents. Therefore, the delimitation of lexemes crucially hinges on the distinction between inflectional and derivational morphology, the latter creating new lexemes. Though apparently clear-cut in simple cases like *(to) walk ~ (she) walks ~ walked* (inflection) vs. *walk ~ walker* (derivation), the distinction between inflection and derivation has proven notoriously difficult to specify in an adequate and unproblematic way (Bybee 1985: ch. 4; Dressler 1989; Plank 1994; Laca 2001; Spencer 2013a). The common intuition that derivation feeds the lexicon, while inflection is relevant to syntax (cf. the "Split Morphology hypothesis", Anderson 1982; Perlmutter 1988; Scalise 1988; Bickel and Nichols 2007: 169–72) is demonstrably wrong. Derivation may have syntactic repercussions (e.g. in causativization or in nominalization), and some inflection is not directly relevant to syntax (cf. the distinction between "contextual" and "inherent" inflection introduced by Booij 1993, 1996 or between "early" vs. "late system morphemes" in Myers-Scotton 2002; these notions are not unproblematic themselves, see Spencer 2013a: 77–82).

In most recent discussions of inflection and derivation—in both descriptions of individual languages and typological studies—they are regarded as two poles on a continuum structured by a set of features (Dressler 1989; Plank 1994; Nau 2001; Haspelmath and Sims 2010: ch. 5; Corbett 2010; Spencer 2013a). In Table 21.1 we list some of the familiar features (cf. Haspelmath 2002: 70–7; Booij 2006: 655–9; Kroeger 2006: 70–7; Brown and Hippisley 2012: 37).

These features are useful as heuristics to place particular morphological processes on the continuum between prototypical inflection and prototypical derivation (with different uses of the same morpheme often occupying different positions on the scale, see e.g. Say 2005 on Russian reflexive verbs). However, morphological typology and morphological theory should ask the empirical question whether these two traditionally recognized clusters of properties are the only ones attested in languages. The answer is in the negative (see Spencer 2013a for a recent comprehensive and convincing discussion).

Thus, Bauer (2004b) proposes a six-way classification of morphological processes, setting valency-changing, class-changing, and evaluative formations aside from other kinds of derivational morphology as being regular and in some sence paradigmatic, and in opposition to inflectional morphology, which does not create new lexemes. This latter criterion of new lexeme creation, in our view, is problematic not only because it obviously involves circularity, but also on purely empirical grounds. In languages with highly productive and compositional valency- or class-changing operations, it is hardly feasible to treat all such cases as distinct lexemes (cf. Spencer 2013a: 42–3). For example, in Adyghe there are about a dozen applicative prefixes which add an object to the valency frame of the verb (Smeets 1992; Lander and Letuchiy 2017), cf. (9a) with a benefactive applicative and (9b) with a comitative one. Not only do these applicatives occur farther from the root than certain markers of contextual inflection such as prefixes cross-referencing the agent (9b), but their occurrence is sometimes obligatory and often fully semantically transparent, so postulating separate lexemes is not a viable descriptive option.

Table 21.1. Features of prototypical inflection and derivation

Parameter	Inflection	Derivation
Function	Does not change syntactic category of a word	May change syntactic category of a word
Meaning	Often has purely grammatical meaning	Tends to have lexical semantic content, i.e. meanings similar to the meanings of independent words
Regularity	Is often semantically regular	May have unpredictable semantic content
Syntacticdeterminism	Is often syntactically determined	Does not require a specific syntactic environment
Obligatoriness	Function is obligatory	Function is not obligatory
Productivity	Is highly productive	Often applies only to certain words, or classes of words
Paradigmaticity	Is often organized in paradigms	Is often not organized in paradigms
Fusion	Can be marked by portmanteau morphemes	Is rarely marked by portmanteau morphemes
Recursivity	Is marked only once in the same word	May apply twice in the same word
Position	Occurs in a peripheral position near the edges of a word	Occurs in a central position close to the root

(9) Adyghe (examples from narratives, Yu. Lander, p.c.)

a. *weš'x* *q-a-**f**-je-š'xə-r-ep*.
rain DIR-3PL.IO-**BEN**-IO-rain-DYN-NEG
'it does not rain for them'

b. *zə-qə-b-**d**-jə-ʔetə-š't*
RFL.ABS-DIR-2SG.IO-**COM**-3SG.ERG-raise-FUT
'it will go up together with you'

Another typologically important notion has been proposed by de Reuse (2009), who singles out "Productive Non-inflectional Concatenation" (PNC) as a special kind of morphology distinct from inflection and derivation and sharing many features with syntax, see Table 21.2 and example (10). PNC is especially characteristic of polysynthetic languages such as those of the Eskimo-Aleut or Abkhaz-Adyghe families, but is also attested, though rarely, in familiar European languages (e.g. the English productive and potentially recursive prefix *anti-*, de Reuse 2009: 28).

(10) Central Siberian Yupik Eskimo (Eskimo-Aleut, Alaska and Chukotka; de Reuse 2009: 23)
*negh-**yaghtugh-yug-uma-yagh-pet**-aa*
eat-go.to-want.to-PST-FRUSTR-INFRN-IND.3SG>3SG
'It turns out s/he wanted to go eat it, but . . . '

Table 21.2. Productive noninflectional concatenation

	Inflection	(Nonproductive) derivation	PNC	Syntax
Productivity	yes	no	yes	yes
Recursivity	no	no	yes	yes
Necessarily concatenative	no	no	yes	yes
Variable order possible	no	no	yes	yes
Interaction with syntax	yes	no	yes	yes
Category change	no	yes	yes	yes

Source: de Reuse (2009: 22).

In conclusion, the traditional notions of inflection and derivation are associated with a large number of empirical and conceptual problems, and both morphological theory and typology should address these problems in order to arrive at a cross-linguistically informed and unbiased set of concepts and distinctions, which will most probably yield a multidimensional space rather than a binary opposition (cf. again Spencer 2013a: ch. 3).

21.3 The relation between meaning and form in morphology

Morphology is the relation between meaning and form in the structure of words, cf. the title of Bybee (1985). The primary goals of morphological typology and theory are thus to determine the ways languages connect meaning and form, and to discover the principles underlying the cross-linguistic variation found in this domain.

There are two important dimensions of morphological variation in relating meaning to form (apart from the variation in the morphologically encoded meanings themselves), cf. Anderson (2015a: 13). The first dimension is how morphological meanings are expressed and how such expressions are organized with respect to each other (morphological exponence and morphotactics). The second is how expressions with the same meaning may vary in context (allomorphy). Both of these dimensions have figured prominently in the classic morphological typology since at least Friedrich von Schlegel and Wilhelm von Humboldt (cf. Rousseau 2001). They are reflected in the traditional typological classification of languages into "isolating", "agglutinating", and "flexive" types, using criteria such as cumulative vs. separatist exponence of morphological features, fusion between stems and affixes, and presence of phonologically opaque alternations of stems and affixes (for an overview see Plungian 2001). As with any "holistic" approach to typology, this classic typology has proven to be inadequate because languages rarely behave uniformly with respect to the different criteria (Plank 1999; Haspelmath 2008). Instead of few discrete classes we must again assume a multidimensional typological space that is yet to be fully investigated (for earlier proposals in this vein see e.g. Sapir 1921 and Alpatov 1985; the latter is discussed in English by Testelets 2001: 309–10).

Table 21.3. Deviations from biuniqueness according to Carstairs (1987)

	many meanings ~ one form	many forms ~ one meaning
syntagmatic axis	cumulation	extended exponence
paradigmatic axis	syncretism	allomorphy

Table 21.4. Deviations from biuniqueness in Russian nominals

	'brother'		'mother'	
	Singular	Plural	Singular	Plural
Nominative	*brát*	*brát'-j-a*	*mát'*	*máter-i*
Accusative	*brát-a*	*brát'-j-ev*	*mát'*	*mater'-éj*
Genitive	*brát-a*	*brát'-j-ev*	*mát'er'-i*	*mater'-éj*
Locative	*brát'-e*	*brát'-j-ax*	*mát'er'-i*	*mater'-áx*
Dative	*brát-u*	*brát'-j-am*	*mát'er'-i*	*mater'-ám*
Instrumental	*brát-om*	*brát'-j-am'i*	*mát'er'-ju*	*mater'-ám'i*

Note: For the sake of consistency, palatalized consonants are marked by ' throughout, including cases of automatic palatalization not reflected in the orthography.

A useful starting point for studying the meaning–form relations in morphology is the idealized model that assumes a biunique mapping between meaning and form, with each morphological feature or 'meaning' expressed by only one form, and each form expressing only one such 'meaning' (cf. Dressler 1987: 111). Most languages display certain deviations from this ideal, and the cross-linguistic investigation of such deviations is one of the primary concerns of morphological typology. A classification of such deviations has been proposed by Carstairs (1987: 12–18), see Table 21.3. (See also Carstairs-McCarthy 1994, 1998, 2001, 2002b, 2010.)

Table 21.4 with a subset of the Russian nominal declension illustrates all four types of deviations from biuniqueness identified by Carstairs. The expression of case and number values in Russian is cumulative and often syncretic (thus, in 'brother' AccSg = GenSg, in 'mother' NomSg = AccSg, GenSg = LocSg = DatSg = NomPl, and in both nouns AccPl = GenPl). The plural subparadigm of 'brother' involves extended (or multiple) exponence of number, since the plural is expressed both by the suffix *-j-* and by cumulative case-number endings. Finally, there are numerous instances of allomorphy of both stems and affixes, the latter clearly showing the distinction between two inflection classes.

Another point of departure for the typological investigation of morphological phenomena is the "canonical inflection" model proposed by Corbett (2005) and further refined in Corbett (2007a, 2007b; see also Bond, Chapter 20 this volume), which can be viewed as an extension of Carstairs' classification, see Table 21.5.

Table 21.5. Corbett's "canonical inflection" and deviations from it

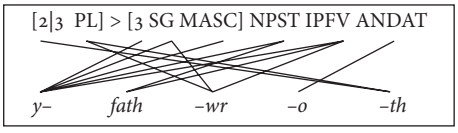
	comparison across cells of a lexeme		comparison across lexemes	
	"canon"	deviation	"canon"	deviation
composition/structure	same	fused exponence periphrasis	same	defectiveness overdifferentiation anti-periphrasis
lexical material	same	stem alternations suppletion	different	homonymy
inflectional material	different	syncretism inflectability	same	inflection classes heteroclisis deponency

Most of these phenomena have been investigated from a cross-linguistic perspective by the Surrey Morphology Group (see http://www.smg.surrey.ac.uk/projects/), cf. Brown et al. (2012), Chumakina and Corbett (2013) on periphrasis; Corbett (2007a), Corbett et al. (2005) on suppletion; Baerman, Brown, and Corbett (2005), Baerman and Brown (2005a, 2005b) on syncretism; Corbett (2009), Baerman (2012, 2014) on inflection classes; Baerman et al. (2007) on deponency, Baerman, Corbett, and Brown (2010) on defectiveness, and many others; a similar perspective with some non-trivial extensions is provided in Stump (2016a); cf. also Harris (2017) for a typological study of multiple exponence. Though most of these phenomena have usually been considered by typologists and theoretical linguists as "exceptions" and "irregularities", their cross-linguistic study has proven to be not only possible, but fruitful and instructive by showing what types of mismatch between meaning and form are possible in morphological systems, how they interact with each other and with syntax, and what kind of motivations may underlie them.

One of the extreme cases of form–meaning mismatch in morphology is the so-called "distributed exponence" (Caballero and Harris 2012: 170–1). In this type of mismatch, the grammatical interpretation of a wordform is constructed through the unification of the meanings of several morphemes, each of which is underspecified with respect to particular feature values. Perhaps the most striking examples of this kind of morphological organization come from the Yam family of New Guinea (Evans 2012, 2015). In Yam languages, the morphological features of participant person and number, aspect and tense rarely have dedicated exponents, but are inferred from particular combinations of affixes and stem allomorphs, each associated with several distinct feature values. An illustration is the Komnzo verbal form presented in Figure 21.1, where four of the morphemes (including the lexical stem *fath-*) combined in the word map to various feature values in complex ways.

Another dimension of morphological diversity is the type of exponence that languages employ (cf. Trommer 2012). Concatenative or linear exponence by means of prefixes and suffixes, as well as reduplication,[5] is the most common type of morphological expression cross-linguistically. However, various kinds of non-concatenative morphology also abound

[5] In the sample of Rubino (2005) there are five times as many languages with reduplication (311) as languages without (56).



'They hold him away.'

FIGURE 21.1. Distributed exponence in Komnzo (Yam, Papua New Guinea)
Source: Döhler (2016: 209, Fig, 5.4). Reproduced with permission

Table 21.6. Non-concatenative exponence in Dinka nouns

	'ground'	'house'	'fire'
Absolutive	*pìɲ*	*ɯ̀ɔ̀t*	*mà̰ac*
Oblique	*pîɲ̂*	*ɯ̂ɔ̂t*	*mâ̰ac*
1st construct state	*pìɲ*	*ɯ̀ɔ̀n*	*mà̰aɲ*
2nd construct state	*pyɛ̀ɛ̀ɲ*	*ɯ̀ɔ̀ɔn*	*mà̰aɲ*
Allative	*pìɲ*	*ɯ̀ɔ́t*	*mɛ̂ɛɛc*
Inessive-ablative	*pɪ́ɪ́ɪɲ*	*ɯ̀ɔ̀t*	*mɛ́ɛɛc*

Source: Andersen (2002: 29).

in the languages of the world. These include infixation, vocalic and consonantal alternations, truncation, as well as non-segmental exponence such as stress and tone changes, and combinations thereof. Probably the best-known and most widely studied case of non-concatenative exponence is the Semitic root-and-pattern morphology (McCarthy 1981; Arad and Shlonsky 2005, among many others). However, perhaps the most striking case of non-concatenative morphology comes from the Western Nilotic language Dinka (Andersen 1993, 1994, 2002). Dinka words are largely monosyllabic, but the language has considerably elaborate morphological paradigms. Affixal exponence is almost absent in Dinka, and most morphological properties are expressed by means of alternations in vowel length, consonant and vowel quality, voice quality, and tone, cf. Table 21.6.

Such exuberant non-concatenative morphology is instructive for descriptive linguists, who must be aware that investigating the morphology of a language may require sophisticated phonetic and prosodic analysis. It also presents challenges for morphological theories which assume linear morphological exponence to be the default case (e.g. Bye and Svenonius 2012) or regard affixal exponence as fundamentally distinct from stem alternations (e.g. Carstairs-McCarthy 2002b, 2010). Non-concatenative morphology is also said to be a hallmark of sign languages, see e.g. Aronoff et al. (2005), Aronoff, Meir, and Sandler (2005) and Napoli (Chapter 30 this volume).

Orthogonal to type of exponence is the locus of marking, that is, the distinction between head-marking and dependent-marking introduced by Nichols (1986), cf. Bickel and Nichols (2005a; 2007: 193–7). Perhaps most importantly, this morphological property of 'locus', whose values are unevenly distributed across language families and linguistic areas, has been shown to correlate cross-linguistically with other typological variables such as basic word order and morphosyntactic alignment (Nichols 1992).

In sum, studying the relation between meaning and form in morphology has been a central issue in morphological research, and has led to a number of different typological classifications. While classic holistic classifications have been proven to be inadequate, more useful approaches have studied meaning–form relations in morphology as departures from a biunique mapping between meaning and form, or as having more or less canonical properties. Other dimensions of morphological typology are constituted by the locus and type of morphological exponence, and here it is worth emphasizing that although concatenative exponence and dependent-marking are prominent in the more familiar European languages, non-concatenative expression and especially head-marking are widely attested in the world's languages and thus have to be accounted for by any theory of morphology.

21.4 Sytagmatic dimensions of morphological typology

One of the traditional fields of morphological inquiry concerns the syntagmatic relations between the components of complex words. In this field, affix ordering has featured prominently, starting perhaps with Greenberg's (1963/1966) Universals #28 concerning the mutual order of inflectional and derivational affixes and #39 concerning the mutual order of case and number affixes (see Baker 1985; Bybee 1985; Muysken 1986; Stump 1997, 2006a; Cinque 1999; Mithun 2000b; Paster 2009; Manova and Aronoff 2010; Spencer 2013a: 219–49; Manova 2015; for a general overview see Rice 2009).

Among the universal principles explaining cross-linguistic tendencies in affix ordering, Baker's (1985) Mirror Principle—couched in the generative framework—and Bybee's (1985) Principle of Relevance—from an expressly functionalist perspective—both reflect the observation that if a language has words hosting more than one affix in sequence, the relative ordering of the affixes is largely steered by semantics. In many languages this is manifested in verbal affixes occurring in the order "(verbal root)-aspect-tense-mood-person" (Bybee 1985: 34–5). This order corresponds both to the meanings' decreasing degree of "relevance" to the meaning of the root and their widening semantic scope (Bybee's "generality").

The much more fine-grained hierarchy of affixal positions proposed in the generative framework by Cinque (1999) largely reflects the same observation. Moreover, in many languages affixes may admit variable order depending on their mutual scope, as in Adyghe (11) where the habilitive ('can') and similative ('seem/pretend') suffixes can be permutated in accordance with their mutual scope.

(11) Adyghe (Korotkova and Lander 2010: 305, 306)

a. *waŝʷe-m ẑʷaʁe qə-tje-s-xə-ŝʷə-ŝʷe.*
sky-OBL star DIR-LOC-1SG.ERG-take-HBL-SML
'It seems that I can take a star from the sky.' (similative > habilitive)

b. *waŝʷe-m ẑʷaʁe qə-tje-s-xə-ŝʷe-ŝʷə.*
sky-OBL star DIR-LOC-1SG.ERG-take-SML-HBL
'I can pretend as if I am taking a star from the sky.' (habilitive > similative)

Table 21.7. The Bininj Gun-Wok verb structure

−12	Tense	
−11	Subject	obligatory "pronominal zone"
−10	Object	
−9	Directional	
−8	Aspect	
−7	Miscellaneous I	
−6	Benefactive	
−5	Miscellaneous II	optional zone
−4	Generic incorporated nominal	
−3	Body part incorporated nominal	
−2	Numerospatial	
−1	Comitative	
E	Embedded verb stem	
0	Stem	
+1	Reflexive/Reciprocal	
+2	Tense-Aspect-Mood	obligatory "conjugation zone"
+3	Case	

Source: Evans (2003a: 318–19).

Table 21.8. Layered vs. template morphology

Diagnostics	Layered morphology	Template morphology
Zero morphemes (significant absence)	No	Yes
Zero derivation	Yes	No
Monodeterminacy (one root, one head)	Yes	No
Only adjacent morphemes may influence each other	Yes	No
Morphemes cannot be sensitive to more peripheral morphemes	Yes	No
Usually encodes at most one argument	Yes	No
Scope-determined position	Yes	No

Source: Stump (2006a: 561); Bickel and Nichols (2007: 214)

However, in many other languages affixes occur in a rigid order hardly amenable to a transparent synchronic motivation in terms of scope, cf. Table 21.7 showing the organization of the verbal word in Bininj Gun-Wok (Gunwinyguan, Northern Australia).

The widespread occurrence of conventionalized affix orders has led researchers to postulate two types of morphological organization referred to as "layered morphology" vs. "template morphology" (Simpson and Withgott 1986; Stump 2006a; Bickel and Nichols 2007: 214–20; Good 2016). The prototypical differences between these are presented in Table 21.8, see Stump (2006a) for more details and examples.

Both layered morphology and template morphology are idealized concepts rather than concrete language types, since most languages with complex morphology present a mixture of both kinds of ordering. Thus, in the abovementioned Adyghe, the suffixes appear to be organized in a layered system, while prefixes follow a more or less rigid template, cf. Korotkova and Lander (2010: 302), with scope-based rearrangements being nevertheless possible for some prefixes as well, see Lander (2016: 3519).

The question of ordering of morphological exponents is relevant not only for affixes and clitics (on the latter, see Simpson and Withgott 1986; Spencer and Luís 2012a: 112–26; for a description of a complex clitic system in an individual language, see e.g. Klamer 1997 on Kambera), but for non-concatenative morphology as well. For instance, the non-linear morphology of Dinka is organized into a layered structure of successively applying operations, as shown in example (12).

(12) Dinka (Western Nilotic, Ethiopia, Andersen 2002: 29)

root = plural	*lɛ̀c*	'teeth'
root+singular	*lɛ̰̂ec*	voice quality shift, vowel lengthening
root+singular+construct state 1	*lɛ̰̂eɲ*	nasal replacement
root+singular+construct state 1+ +construct state 2	*lɛ̰̂ɛɲ*	vowel lowering

Besides morpheme ordering, the worldwide distribution of prefixation vs. suffixation has received much attention. It is received wisdom that suffixes are more common cross-linguistically than prefixes (Dryer 2005a), and explanations for this preference on the basis of psycholinguistics (Hall 1988; Hawkins and Cutler 1988) and prosody (Himmelmann 2014) have been proposed. It has also been shown that different morphological categories prefer suffixal exponence to differing degrees (cf. Bybee, Pagliuca, and Perkins 1990; Bakker and Siewierska 1996; Dryer 2005b, 2005c, 2005d), which implies that choice of exponence is motivated not only by ease of processing.

Another aspect which has gained prominence in typology relates to the quantification and cross-linguistic comparison of syntagmatic morphological complexity. Starting from the classic morpheme-to-word ratio proposed by Greenberg (1954), this field of inquiry has been extended by Nichols (1992, 2009), who considers such parameters as sum of head-marking and dependent-marking constructions or the number of inflectional categories expressed in the verb (Bickel and Nichols 2005b). Such an approach to morphological complexity is, however, fairly limited in that it disregards the paradigmatic aspects of morphology, to which we will now turn.

21.5 Paradigmatic dimensions of morphological typology

Morphological paradigms have been prominent in traditional and pedagogical grammar since Antiquity, and have become an object of theoretical inquiry in work such as Matthews (1972) and Anderson (1992). These authors have advocated the so-called Word-and-Paradigm

models of morphology (see also the typologically-oriented work of Plank 1986 and the contributions to Plank 1991; for more details see Blevins, 2016; Blevins, Ackerman, and Malouf Chapter 13, this volume, and Stump Chapter 14, this volume). Though paradigms are looked at with skepticism by some generative morphologists (cf. e.g. Bobaljik 2007), such phenomena as syncretism, suppletion, inflection classes, deponency, etc. cannot be neglected by any theory of morphology aiming at empirical and cross-linguistic adequacy (cf. e.g. Ackerman, Blevins, and Malouf 2009 or Stump 2016a). It is precisely the paradigmatic dimension of morphology, in particular such phenomena as "morphomic" (opaque) allomorphy and inflection classes (Aronoff 1994; Carstairs-McCarthy 2010), that has been called "autonomous morphology" (cf. Maiden 2005, Cruschina, Maiden, and Smith 2013). These features of morphology are claimed to be irreducible to other components of grammar (cf. Stump 2016a) and to constitute one of the core domains of linguistic complexity (cf. Dahl 2004; Baerman, Brown, and Corbett 2015b).

The broad typological investigation of various aspects of paradigmatic morphology, in particular of deviations from the "canonical inflection" model, have been mainly carried out by the Surrey Morphology Group (see §21.3). Besides that, such work as Cysouw (2003) on the paradigmatics of verbal person marking and Veselinova (2003, 2005a, 2005b) on verbal suppletion, deserve attention. The latter work, based on a large cross-linguistic sample, shows that even such an apparently irregular phenomenon as suppletion is subject to systematic typological generalizations, promising fruitful insights in other related domains as well. Akin to the topic of suppletion is the study of stem alternations (Blevins 2003; Aronoff 2012; Spencer 2012). While this topic has received most attention in Romance linguistics (see first of all the work by Martin Maiden), it is certainly an important typological issue (Carstairs 1987: ch. 6; Stump 2001: ch. 6, 2016a: chs. 5, 11; Carstairs-McCarthy 2010: ch. 6). Bybee (1985: 92) and Veselinova (2003) have claimed that cross-linguistically suppletive stems tend to cut morphological paradigms along such major inflectional distinctions as singular vs. plural number, perfective vs. imperfective aspect or past vs. non-past tense. On the other hand, the work by Maiden (2005) and Carstairs-McCarthy (2010) has suggested that even "morphomic" stem alternations (including suppletion), not associated with any coherent set of morphosyntactic properties, play an important role in grammars and are not fully arbitrary, as evidenced for example by their diachronic stability.

Another currently prominent line of inquiry concerns inflection classes. Starting in the 1980s with the question of the possible limits on the number of inflection classes (Carstairs 1983, 1987: chs. 3, 7; Carstairs-McCarthy 1994, 2010: ch. 5), this field has substantially expanded its empirical database in the recent work by Blevins (2004), Stump (2006b), Stump and Finkel (2013), Finkel and Stump (2007), Baerman (2012, 2014, 2016). In particular, it has been shown that the fairly restrictive principles of paradigmatic economy proposed by Carstairs-McCarthy (1994, 2010) seem to be violated by languages with exuberant inflection class systems like Nuer (Western Nilotic, South Sudan) or Seri (isolate, Mexico), cf. Table 21.9, showing how just two Nuer affixes can create a large number of inflectional classes (only a small subset of actual Nuer declensions is shown in the table) when the distribution of these affixes is not tied to particular morphosyntactic values.

A new line of analysis of inflection class systems, which seems very promising from both a theoretical and a typological perspective, applies the insights of information theory. This type of work asks the question about the mutual predictiveness of particular forms in the paradigm (e.g. the typology of "principal part" systems proposed by Finkel

Table 21.9. Some Nuer inflection classes

	'milk'	'kind of tree'	'potato'	'hair'
NOM SG	*cak*	*kɛ̈c*	*tac*	*nhim*
GEN SG	*caak*	*kɛ̈c-kä*	*tac-kä*	*nhi̱m*
LOC SG	*caak*	*kɛ̈c-kä*	*tac*	*nhi̱m-kä*
NOM PL	*ca̱k*	*kɛɛc*	*tac-ni*	*nhiäm*
GEN PL	*ca̱k*	*kɛɛc-ni*	*tac-ni*	*nhiäm-ni*
LOC PL	*ca̱k-ni*	*kɛɛc*	*tac-ni*	*nhiäm-ni*

Source: Baerman (2012: 468).

and Stump 2007) and quantitatively compares inflection class systems in terms of entropy (Ackerman and Malouf 2013), taking into account such extramorphological parameters as type and token frequency of particular inflection classes. This line of inquiry requires a close collaboration between typologists, morphologists, and computational linguists (cf. Walther 2013). The entropy-based approach to morphological paradigms has also proven useful for the analysis of defectiveness, apparently an irregular quirk *par excellence*, see Sims (2015) for a view of defectiveness as a phenomenon amenable to systematic generalizations.

Alongside inflection classes, which constitute a prime example of lexically determined allomorphy, natural languages abound in phonologically and grammatically conditioned allomorphy of stems and affixes. Phonologically conditioned allomorphy is a relatively well-understood phenomenon, see for example Paster (2006), Nevins (2012). However, less is known about the types of grammatically conditioned allomorphy and the constraints on it, see for example Carstairs-McCarthy (2001), Bonet and Harbour (2012). In addition, it has been argued that allomorphy can be sensitive to the lexical semantics of the stem in principled ways. For instance, Aristar (1997) has shown that longer allomorphs of case markers tend to appear on nominals whose inherent meaning is not directly compatible with the function of the case. This promising topic has not yet received the attention it deserves, though cf. Arkadiev (2017) for a typological study of the allomorphy of ergative case.

Last but not least, morphological entities are often polysemous or polyfunctional. Indeed, the polyfunctionality of inflectional (and, more marginally, derivational) elements has received most attention in linguistic typology, see Haspelmath (2003) and Evans (2011) for overviews,[6] as well as numerous contributions to Rainer et al. (2014) and Müller et al. (2015). Cross-linguistic investigations have discovered recurrent patterns of polysemy of many morphological categories ('grams') and some of these have been linked to diachronic paths of grammaticalization and semantic development (e.g. Bybee, Perkins, and Pagliuca 1994), thus revealing systematic correspondences between aspects of morphological form and linguistic meaning.

In sum, the paradigmatic dimension in morphology, which has been prominent in traditional grammar but largely neglected in early morphological theorizing and cross-linguistic comparison, is currently enjoying a revival of interest from both theoretical

[6] The most comprehensive typological overview of grammatical polysemy is perhaps Plungian (2011), existing only in Russian and in Croatian and Lithuanian translations.

and typological perspectives. This multifaceted field of inquiry requires sophisticated methodology (including quantitative measures and computational modeling) and promises important insights into the structure and development of morphological systems and morphological complexity (cf. e.g. Nichols to appear).

21.6 Conclusions

Despite some notable achievements, morphological typology is still in a state of development. In our view, the major challenge for both morphological theory and morphological typology is to find a good balance between analytic and conceptual depth, on the one hand, and breadth of empirical coverage, on the other. While most of the non-trivial theoretical insights in morphology are based on data from a limited set of languages (fortunately, also including non-European ones), large-scale cross-linguistic studies of morphology have rarely gone beyond somewhat superficial observations (Harris 2017 being a notable and welcome exception). A balance between theory and typology can only be achieved by the joint efforts of typologists, theoreticians, and descriptive linguists.

Morphological typology, morphological theory, and descriptive and documentary linguistics mutually enrich each other in many respects. If linguists describing individual languages are aware of the analytical notions, methodological insights, and problematic issues of current morphological theory and typology (such as the multidimensional rather than binary nature of traditional distinctions word vs. affix, inflection vs. derivation, or agglutination vs. flexion), they will produce more sophisticated and empirically adequate descriptions. In turn, such descriptions will feed both theory and typology.

Advances in theoretical and typological research go hand in hand with new trends in descriptive and documentary linguistics. Current theorizing and cross-linguistic comparison require access not only to good grammatical descriptions, but also to dictionaries explicitly indicating such morphological information as inflection class membership, stem alternations and suppletion, or defectiveness. Theoreticians and typologists also need access to morphologically annotated corpora. With respect to this last point it should be mentioned that different types of morphological organization pose different problems for tasks like tokenization (linked to the definition of wordform), lemmatization (related to the inflection/derivation divide), and tagging, see e.g. Arkhangeliskiy and Lander (2015). Their solution can only be reached through collaboration between theoreticians, computational linguists, and typologists.

Morphological typology is indispensable for morphological theory, as typology is a testing ground for analytical models and hypotheses. Here the goals of the two enterprises—still conceived of by some as fundamentally distinct—largely converge. Morphology, which by its very nature is neither present in all languages nor cross-linguistically uniform, hardly admits overarching universal generalizations and much more readily provides answers to the "what's where why" type of question (Bickel 2007: 239) usually asked by typologists. At the same time, theoretical conclusions can only be valid when they are based on an understanding of the kinds of morphology (including exponence, morphotactics, allomorphy, and paradigmatic structure) found in certain language families and linguistic areas, as well as on an account of the ways morphological systems diachronically develop

through inheritance or contact, cf. Gardani (2008), Johanson and Robbeets (eds. 2012), Gardani, Arkadiev, and Amiridze (2015). Morphological theory needs morphological typology just like typology profits from theory, while good morphological descriptions have to be cross-linguistically and theoretically informed.

Acknowledgments

We are grateful to Jenny Audring, Geert Booij, Francesca Masini, Gabriele Schwiertz, and an anonymous reviewer for comments and corrections. All faults and shortcoming are ours.