1 *Introduction*

Inflectional morphology equips a language with the means to combine lexical and grammatical information. For example, the Aymara word *anunakataki* and the Polish word *psom* both mean 'for dogs', and in each language the notions of being a dog, plurality, and benefaction are expressed by a single word form. By contrast, in a language like Tok Pisin, each of these three notions is conveyed by a separate word: *long* (derived from English *belong*, expressing benefaction), *ol* (derived from English *all*, expressing plurality), and *dok* (derived from *dog*, the lexical information); these combine into the phrase *long ol dok*. As a medium of expression, the morphological forms of Aymara and Polish are doing the same job as the syntactic construction of Tok Pisin, and the only difference is how the information is packaged: as separate words (Tok Pisin), or as aspects of a single word (Aymara and Polish). But the inclusion of this information in the word form opens up a new dimension – the paradigm – absent in purely syntactic arrangements. And with this new dimension come new relationships and a host of complexities characteristic of morphological systems.

By way of illustration, consider the noun paradigms in Table 1.1, from Aymara, an Aymaran language spoken in Bolivia and neighbouring countries, and in Table 1.2, from the Slavonic language Polish. The Aymara paradigms differ only minimally from what we might construct, item-by-item, in syntax. We can identify a noun stem and a set of elements that follow, and then concatenate the two. Morphophonological quirks show us that this is not *just* syntactic concatenation: the accusative and comparative both require deletion of the stem-final vowel found in most of the rest of the paradigm. But aside from this essentially implementational detail, the organization of these paradigms is much the same as it would be if they were bare nouns paired with various different adpositions: in order to construct an ablative case form meaning 'from the house', we take the stem labelled 'house' (*uta*) and the suffix labelled ABLATIVE (*-ta*) in the right column, and put the two together.

Table 1.1 *Aymara noun paradigms*

	'house'	'money'	'dog'	suffixes
NOMINATIVE	uta	qullqi	anu	Ø
ACCUSATIVE	ut	qullq	an	Ø
GENITIVE	uta-na	qullqi-na	anu-na	na
ABLATIVE	uta-ta	qullqi-ta	anu-ta	ta
ALLATIVE	uta-ru	qullqi-ru	anu-ru	ru
INSTRUMENTAL	uta-mpi	qullqi-mpi	anu-mpi	mpi
BENEFACTIVE	uta-taki	qullqi-taki	anu-taki	taki
COMPARATIVE	ut-hama	qull-hama	an-hama	hama
INTERACTIVE	uta-pura	qullqi-pura	anu-pura	pura
LIMITATIVE	uta-kama	qullqi-kama	anu-kama	kama
PURPOSIVE	uta-lajku	qullqi-lajku	anu-lajku	lajku

Source: Coler 2015.

Table 1.2 *Polish noun paradigms (singular forms)*

	'fish'	'nose'	'cart'	suffixes		
NOMINATIVE	ryb-a	nos	wóz	a	Ø	Ø
ACCUSATIVE	ryb-ę	nos	wóz	ę	Ø	Ø
GENITIVE	ryb-y	nos-a	woz-u	y	a	u
DATIVE	rybi-e	nos-u	woz-u	e	u	u
LOCATIVE	rybi-e	nosi-e	wozi-e	e	e	e
INSTRUMENTAL	ryb-ą	nos-em	woz-em	ą	em	em

Now consider the Polish paradigms in Table 1.2. Here too we have three different nouns. But we cannot assemble these forms with the same ease as in Aymara. First, instead of there being a single set of suffixes that can be matched to any noun, there are different suffixes that occur with different nouns: for example, 'fish' takes a genitive singular suffix *-y*, whereas 'nose' takes *-a*. Simply labelling *-y* as genitive singular and *-a* as genitive singular will not tell us when to use which. Second, the way the grammatical functions themselves are packaged differs between the nouns: for example, the suffix *-u* is used for dative singular with 'nose', but for both genitive and dative with 'cart'. That means no single grammatical label will tell us when exactly to use *-u*.

The Polish word forms are so shot through with lexical idiosyncracies that it takes more work to say what is going on than with Aymara. Schematically, we can represent the contrasting descriptive tasks as in Table 1.3. Following the same schema as in Tables 1.1 and 1.2, we list the morphosyntactic values (like NOMINATIVE or DATIVE) as rows, and the lexemes as columns.

Table 1.3 *Simplicity vs. complexity*

a. Aymara-type system

	lexeme 1,2,3
value a	
value b	
value c	

b. Polish-type system

	lexeme 1	lexeme 2	lexeme 3
value a			↕
value b	↕		
value c			

For Aymara there is just one column, because all the lexemes behave in the same way, and we need only to fill in the blanks by means of rules. For Polish, we need to expand this model along the horizontal dimension, creating three columns to accommodate three patterns of inflection for three different classes of lexeme. And further, in the vertical dimension, we need to acknowledge that the dimensions of the cells vary across items, if we are to be true to the evidence provided by the forms themselves. This is what we mean by morphological complexity: on the assumption that the two systems in Table 1.3 are doing the same job, the Polish-type system is more complex, because we need more elements and more steps to describe it. We should, however, caution against overinterpreting our use of the term *complexity*. We take no stand here on its cognitive effects – that is, is it really complex for the language user? – or on the possibility that it may have some application above and beyond the grammatical functions that define the paradigm – for example, as an aid to memory. This approach to complexity is thus local and not global, in the terms discussed by Miestamo (2006, 2008), in that it is defined in relation to a particular subsystem. This is not because the broader systemic implications are not interesting or not relevant, but because the main aim of this volume is typological, to explore what kinds of configurations are actually found in languages. To that end, our terms of analysis are those that are generally used for inflectional systems, even for otherwise poorly described languages; this typically means an inventory of forms and the contexts they are found in.

Complexity in this sense is not just an interesting property of a subset of morphological systems. It is, more generally, what sets morphology apart from other linguistic components. It is restricted to paradigmatic structures: the whole idea that morphosyntactic values can be discussed separately from their means and patterns of exponence presupposes a paradigmatic arrangement of facts. And paradigms are characteristically morphological, and in particular,

inflectional.[1] In effect, morphological complexity is the defining property of morphology as an autonomous linguistic component. Our aim here is to confront the full richness of morphological complexity, accepting that these patterns take on a morphological life of their own, worth exploring for their elegance and interest, and especially since they are arguably the most purely linguistic part of language.

This puts us in a somewhat awkward position, which we freely admit. On the one hand, we have identified our topic in terms of inflectional patterns, but at the same time, we have defined these patterns negatively, as what is left over after morphosyntax has been subtracted. Equally, most observers would ascribe inflectional patterns that coincide semantic or phonological regularities to those very modules. Morphological structure is the explanation of last resort, invoked when all other means of analysis have been exhausted. As a heuristic, this is only right and proper. Syntax, semantics, phonology – these are aspects of every language, and no description, whatever framework it is embedded in, can get by without making reference to them. But inflection, let alone what we have described as morphological complexity, is not found in every language, and so it understandably assumes a subordinate role in our thinking. This sets the bar rather high for identifying inflectional structure as a linguistic phenomenon in its own right. First, it must meet the paradigmatic criteria illustrated in Table 1.3. But even then, so one argument goes, the surface patterns could be accidental, and not in any way officially sanctioned, in the same way that homophony and synonymy could be seen as quirks of the lexicon. Thus, a second criterion is that the patterns give some evidence of systematicity, for example, through diachronic persistence. As forcefully argued by Maiden (1992, 2005), morphological patterns whose origins lie in a chance association of disparate elements can nevertheless be propagated across generations and serve as a driver of morphological change, drawing new items into their sphere of influence. Maiden's examples are from Romance, and we supplement these with comparable examples from Finnic languages in Section 2.2. In the light of the sometimes contentious status of morphology within language, such case studies are crucial to demonstrating that somewhere, somehow, purely morphological relationships call the shots, at the expense, seemingly, of all considerations of meaning or function. However, it is not often that we can find such convincing cases, where the diachronic record is solid and where we

[1] We note that the notion of derivational paradigms has some currency at the moment. We make no claims here about the distinction between inflection and derivation – paradigmaticity is just one property among many that some use to make a distinction. Anything paradigmatic falls under our purview.

can distinguish productive patterns from mere inertia (if there is even a difference). Does that make autonomous morphological structure a real but rare phenomenon, like three degrees of vowel length or paucal number? Possibly, provided one accepts the primacy of non-morphological aspects of language. But we would like to suggest here that that is a hasty conclusion. As we hope to show in the ensuing chapters, languages abound in the sorts of paradigmatic deviations illustrated in Table 1.3. They are easily overlooked if one's focus is a morphosyntactic description, but these peculiarities are nonetheless crucial to understanding or producing an utterance in any language that contains them.

We are therefore taking a novel and somewhat eclectic approach to the topic. We are not trying to advance a particular theoretical claim, and so have aimed to maintain a presentation free from polemic. Our definition of morphological complexity is fluid in that it is dependent on what one assumes is already provided by other aspects of the linguistic system. But these other aspects – for example, syntax, semantics, phonology – are largely intangibles, as indeed is morphological structure itself. Our focus is therefore on what is most solidly observable – on forms and their distribution. The goal is to offer a typology of morphological complexity, gathering its various manifestations under one rubric. Armed with this, readers are, of course, free to decide where this fits in their own conceptions of grammar, we hope with an enhanced awareness of the formal and typolological richness of inflectional systems.

In concrete terms, this emerges as a study of inflection classes. This means not just variation in the shape of inflectional marking across different sets of lexemes, as traditionally understood, but also variation in the paradigmatic distribution of inflectional marking, as seen in the Polish example in Table 1.2, where the morphosyntactic value of the suffixes *-a*, *-e* and *-u* varies from paradigm to paradigm. One issue we have chosen largely to skirt around here is the conflation of values within the paradigm, otherwise known as syncretism. This is, partly because we have explored it before (Baerman, Brown, & Corbett 2005), and partly because it would be a distraction from our main goal, given the wealth of competing analyses of feature structure whose plausibility may be dependent on theory-specific assumptions. Distributional variation in the feature values of (what may be supposed to be) identical morphological formatives is a phenomenon which *may* overlap with that of syncretism (e.g., Polish *-u* is dative with 'nose' but genitive/dative with 'cart'), but it need not (e.g., *-a* is nominative with 'fish' and genitive with 'nose').

The volume has the following structure. Chapter 2 surveys the external typology of inflection classes. Although they are most clearly and uncontroversially manifested through affix allomorphy (as in the Polish examples

just illustrated), there is every reason to extend the notion to other types of inflectional exponence, as well as to more abstract properties of inflectional structure, such as affix position, variations in the paradigmatic distribution of affixes, or indeed the very presence versus absence of inflectional marking. Chapter 3 looks from the inside out, as it were, in terms of the morphosyntactic features that are being marked. Although in theory any feature is prone to have its expression split into different inflection classes, there are hints that the way this is manifested can be affected by the semantics of the feature itself. Chapter 4 shifts to the central question of inflection class assignment. Canonically speaking, inflection classes are completely arbitrary, unconnected with any other grammatical component. But often matters are not so clear, with morphological behaviour predictable to some degree from external factors, such as phonology or semantics, showing there is a substantial grey zone in between what we might call motivated and arbitrary morphology. In Chapter 5 we focus on conditions on paradigms; these are generalizations which cross-cut the lexical generalizations which are inflection classes. These conditions, ranging from the very general to the quite specific, have an interesting and surprisingly complete typology. This typology is based on the antecedent of the condition (semantic, syntactic, morphological, or phonological) and the type of paradigm in the consequent of the condition. Chapter 6 steps back and looks at the abstract properties of paradigm structure that are revealed through inflection classes, in terms of predictability both across cells of the paradigm, and across different components of the inflectional system. We identify three basic types of systems. In Chapter 7 we show how these basic types can be associated with three different ways of viewing morphological complexity. The first of these is organization, where the system is easy to specify in the grammar. The second of these is the opposite of organization – emergent complexity, where the major burden is associated with lexical stipulation and there is virtually no role for the morphological grammar. The third is central-system complexity, the trade-off between the two other types, where morphological systems that are high in central system complexity represent a balance between stipulation in the lexicon and the grammatical-rule system. These three types are then associated with three off-the-shelf measures available from Stump & Finkel (2013). In real world systems, the situation is much more intricate, of course, and we provide a case study of Tlatepuzco Chinantec tone classes, illustrating the interplay of lexical and grammatical knowledge, as represented in the system of default-inheritance classes. Chapter 8 sums our

overview of morphological complexity in general, and inflectional classes in particular.

A thread running throughout this volume is the cross-linguistic pervasiveness of morphological complexity as manifested through inflection class distinctions. Although the most familiar examples in the literature come from Indo-European languages, few language families that have inflectional morphology are entirely free from such quirks (and those that seem to be may well just be victims of our own ignorance). We have drawn our examples here from a wide range of languages spanning all the inhabited continents, representing typologically very different inflectional systems. That said, it must also be admitted that some language families are especially prone to morphological complexity, the Oto-Manguean family being a particular stand out, and we have tempered genetic balance with the sort of typological diversity that can be gained only from dipping repeatedly into the same well.

The types of linguistic structures that we are concerned with here are often overlooked in more general discussions of language for partly understandable reasons. They are not necessary for transmitting information, they are not necessary for the smooth running of the sound system. Indeed, they are simply not necessary (many languages have limited morphology). And yet they are there, complex, difficult, and persistent, as we shall see.