

2

Dimensions of morphological complexity

STEPHEN R. ANDERSON

The question of what aspects of a language’s word structure patterns contribute to its overall complexity has a long tradition. With roots in the notions of linguistic typology associated with traditional grammar, Sapir (1921; see also Anderson 1992: §12.2) organizes these matters along three dimensions. One of these concerns the range of concepts represented by morphological markers, and refers to the extent of elaboration of the inflectional and derivational category structure of a language. A second refers to the range of marker types, and thereby differentiates transparent affixation of the sort associated with pure ‘agglutinating’ languages from a variety of other formal processes by which morphological information can be conveyed. The third dimension is that of the overall internal complexity of words, the sheer number of distinct pieces of information that are combined in a single word, ranging from the simplest case of ‘isolating’ languages that involve (little or) no morphological combination up to the ‘polysynthetic’ type¹ in which most or all of the components of a full sentence are expressed within a single word.

My goal in this chapter is to develop and elaborate a characterization of the morphological characterization of languages along lines similar to Sapir’s, so as not only to serve similar typological goals but also to provide a framework for understanding the questions of linguistic typology that motivate other authors in this volume. In that spirit, I will feel free to propose an agenda of questions to be asked about languages without being obliged to offer a comprehensive set of answers. Before proceeding to that enterprise, however, I want to step back from the details and ask what it is about morphology that constitutes ‘complexity’ in the broader picture of human natural language.

¹ Not to be confused with the distinct technical sense of this term in Baker (1995) and related work.

2.1 What is 'complex' about morphology?

Different observers will see different things about a language as making it 'complex'. In traditional terms, for example, the outer limit of morphological complexity in natural languages was often seen as represented by languages of the polysynthetic type, highlighting the sheer quantity of morphological elaboration as the main contributor to complexity.

While teaching a course on the diversity of the world's languages to a class of Chinese students in Beijing recently, I was exposed to what was, for me, a somewhat unusual perspective on this matter. The students had been reading some of Mark Baker's work on the parametric description of grammars, in which he talks about 'polysynthesis' as a parameter of grammatical structure. They did not really know anything about the specific languages he discussed in this connection, but it was clear to them that this notion was associated with being very complex. In trying to figure out what it might mean for a language to be 'polysynthetic', one of my students offered a clarification in a written exercise:

A "Polysynthetic" language is one in which words are very complex. That is, they have more than one meaning element combined into a single word: for instance, English *cat-s*.

From the perspective of a speaker of a Chinese language, apparently, *any* morphological structure seems to be complex. While initially merely amusing to speakers of languages of the 'Standard Average European' type, I will suggest that this is a rather more coherent and principled view than it may seem at first sight.

What aspects of a system contribute to its complexity? A standard rhetorical move would be to consult the dictionary for a starting point. The equivalent source of wisdom in the present age is Wikipedia, which provides the following:

A complex system is a system composed of interconnected parts that as a whole exhibit one or more properties (behavior among the possible properties) not obvious from the properties of the individual parts.

'Complexity', then, can be seen as the consequence of a system's displaying characteristics that do not follow as theorems from its nature, as based on its irreducible components. How, then, does this apply to language? What aspects of language are essential, and what properties that languages display represent complications that are not logically necessary?

Languages are systems that provide mappings between meaning, or conceptual structure on the one hand, and expression in sound (or signs) on the other. In order to fulfil this function, there are some kinds of organization that they have to display by virtue of their essential character.

Of necessity, a language has to have a syntax, because it is the syntactic organization of meaningful elements as we find it in human languages that gives them their

expressive power, providing their open-ended ability to express and accommodate a full range of novel meanings.

It is also plausible to suggest that languages need to have phonologies. That is because individual meaningful elements—linguistic signs—must have characteristic expressive properties to serve their essential purpose, as stressed by linguists since de Saussure. When these are combined by means of the syntactic system, however, the result may be at odds with the properties of the system through which they are to be implemented: the properties of the vocal tract, or of the signing articulators. There is thus a conflict between the need to preserve the distinguishing characteristics of meaningful elements and the need to express them through a system with its own independent requirements. Optimality Theory articulates this explicitly as the conflict between considerations of Faithfulness and Markedness, but similar considerations are at the foundation of every theory of phonology. Some account of how this tension is to be resolved is inherently necessary, and thus the presence of phonology, like that of syntax, follows from the nature of language.

Both syntax and phonology are thus inherent in any system that is to fulfil the basic requirements of a human language, and their presence (as opposed to their specific stipulated properties) cannot be seen as constituting complexity in itself. The same cannot be said for morphological structure, however, as pointed out forcefully by Carstairs-McCarthy (2010), and that makes it hard to understand why humans should have evolved in such a way that their languages display this kind of organization at all. To clarify this, let us note that the content we ascribe to morphology can be divided into two parts, 'morphotactics' and 'allomorphy', and in both cases it is difficult to see the existence of such structure as following inherently from the nature of language.

Morphotactics provides a system by which morphological material (grossly, but inaccurately identified with the members of a set of 'morphemes') can be organized into larger wholes, the surface words of the language. But in fact language already involves another system for organizing meaningful units into larger structures, the syntax, and so to the extent the morphotactics of a language can be distinguished from its syntax, this would seem to be a superfluous complication.

Allomorphy is the description of the ways in which the 'same' element of content can be realized in a variety of distinct expressions. When that variation follows from the language's particular resolution of the conflict between the requirements of Faithfulness and Markedness, as described previously, this is just phonology, and can be seen as necessary. But when we find allomorphic variation that does not have its roots in the properties of the expression system, it does not have this character of necessity, and so constitutes added complexity.

2.1.1 *Morphotactics ≠ syntax*

In fact, the two *a priori* unmotivated kinds of complication are essentially constitutive of morphology. Carstairs-McCarthy (2010) cites examples of morphotactic organi-

zation that are unrelated to (and in fact contradict) the syntax of the languages in which they occur; another such example was discussed in Anderson (1992: 22–37) in Kwakw'ala, a Wakashan language of coastal British Columbia.

To summarize, Kwakw'ala is a language with a rather rigid surface syntax. Sentences conform to a fairly strict template, with the verb (or possibly a sequence of verbs) coming in absolute initial position, followed by the subject DP, an object DP marked with a preceding particle beginning with *x-* (if the verb calls for this), possibly another object marked with *s-* (again, if the verb subcategorizes for an object of this type), optionally followed by a series of PPs. Adjectives strictly precede nouns they modify, and other word order relations are similarly quite narrowly constrained by the grammar.

However, Kwakw'ala also has a rich system of 'lexical suffixes' constituting its morphology, and these correspond functionally to independent words in other languages. The point to notice is that when meaningful elements are combined in the morphology, as in the abundant variety of complex words consisting of a stem and one or more lexical suffixes, the regularities of order found in the syntax are quite regularly and systematically violated.

For instance, as just noted, verbs are initial in syntactic constructions, with any objects coming later. When elements corresponding to a verb and its object are combined within a single word, however, they typically appear in the order O-V rather than V-O. Thus, *λ'ena-gila* 'fish.oil-make, make fish oil' has the object of 'make' initially, and the element with verbal semantics following.

Similarly, although the subject invariably precedes any objects in a syntactic construction, if (and only if) the element expressing the object of the verb is a lexical suffix attached to the verb in the morphology, it can precede the subject DP. Thus, in *na'w-əm'y-ida bəgʷanəm* 'cover-cheek-the man, the man covered (his) cheek' the object of 'cover' is the suffix *-əm'y* 'cheek', and as a result it precedes the subject *-ida bəgʷanəm* 'the man'—an ordering that would be quite impossible if the object were separately expressed in the syntax.

As another example, exactly when they are combined in a single word by the morphology rather than composed in the syntax, an adjectival modifier can follow the expression of the noun it modified. The lexical suffix *-dzi* 'large' not only can but must follow an associated nominal stem, as in *λ'aqʷa-dzi* 'copper-large, large copper (ceremonial object)'. Once again, the ordering imposed by the morphotactics of the language is directly contrary to that which would be given in the syntax.

There is actually something of an argument-generating algorithm here: find any systematic regularity of order in the syntax of Kwakw'ala, and it is quite likely that the principles of morphotactic organization will systematically violate it. Overall, the morphotactics of the language bear little or no resemblance to its syntax, which naturally raises the question of why a language should have two quite distinct systems

that both serve the purpose of combining meaningful units in potentially novel ways to express potentially novel meanings.

Given the existence of two separate and distinct combinatory systems in Kwakw'ala, there is potential duplication of function: a new complex meaning might be constructed either on the basis of the morphology or on that of the syntax. And this is indeed the case, as illustrated in (1). Kwakw'ala has a suffix *-exsd* that adds the notion 'want' to the semantics of a stem, and the meaning 'want to (X)' can be conveyed through the addition of this element to the stem representing a verb. But there is also a semantically empty stem *ax-*, and the suffix can be added to this to yield an independent verb *axexsd* 'to want', which can in turn take another verb as its syntactic complement to yield essentially the same sense.

- (1) a. *k^wak^w'ala-exsd-ən*
 speak.Kwakw'ala-want-1SG
 I want to speak Kwakw'ala
- b. *ax-exsd-ən q-ən k^wak^w'ala*
 Ø-want-1SG that-1SG speak.Kwakw'ala
 I want to speak Kwakw'ala

Interestingly, there appears to have been a subtle but significant shift in the language: where traditional speakers relied heavily on the morphology for the composition of novel expressions, modern speakers are much more likely to combine meanings in the syntax. Importantly, neither the morphology nor the syntax has actually changed in any relevant way: what has happened is just that the expressive burden has shifted substantially from one combinatory system to the other. My understanding is that similar developments have occurred in other languages with complex morphology of this sort, as a function of declining active command of that system (though without loss of the ability to interpret morphologically complex words).

2.1.2 Allomorphy ≠ phonology

A similar case can be made concerning the relation between allomorphic alternation and patterns of variation dictated by the phonology of a language. Again, we can illustrate this from Kwakw'ala. As in the other Wakashan languages, Kwakw'ala lexical suffixes each belong to one of three categories, depending on their effect on the stems to which they are attached. These are illustrated in (2).

- (2) **Hardening** (roughly, glottalizing) suffixes, e.g. /*qap + alud*/ → [*qap'alud*] 'to upset on rock'
- Softening** (roughly, voicing), e.g. /*qap + is*/ → [*qabis*] 'to upset on beach'
- Neutral** (no change), e.g. /*qap + a*/ → [*qapa*] '(hollow thing is) upside down'

Importantly, the category of a given suffix is not predictable from its phonological shape: it is an arbitrary property of each suffix that it is either ‘hardening’, ‘softening’, or neutral. As a result, the different shapes taken by stem final consonants with different suffixes cannot be regarded as accommodations to the requirements of Markedness conditions.

This example is of course quite similar to that of the initial mutations in Celtic languages and other such systems. There is little doubt that, like these, there was a point in the history of the Wakashan languages at which the ancestors of these suffixes did differ in phonological form, and the changes we see today are the reflexes of what were originally purely phonological alternations. But the important thing is that in the modern languages, by which I mean everything for which we have documentary evidence since the late nineteenth century, this motivation is no longer present, but the allomorphic variation persists.

Languages are perfectly content, that is, to employ principles of variation whose properties do not follow from the necessary resolution of the competing demands of Faithfulness and Markedness. Such variation is the content of the component of morphology we call allomorphy, and its presence in natural language must be regarded as not following from their nature, and thus as adding complexity.

So we must conclude that from the point of view of what language does and what it needs to fulfil that task, morphological structure is superfluous: neither morphotactic regularities nor non-phonologically conditioned allomorphic variation follow from the basic requirements of linguistic structure. Nonetheless, virtually all languages—even Chinese languages—have at least some morphology that is not reducible to syntax and/or phonology. Since it appears to be the case that in a very basic sense any morphology at all constitutes ‘complexity’, that fact stands in need of an explanation.

A genuinely explanatory account of the basis for morphological organization would have to lie in the evolutionary history by which the human language faculty has emerged in the history of our species. Like many aspects of the structure of language, the search for such an account runs up against a general lack of firm evidence, since language in general leaves no direct trace in the physical record. Of course, just how much morphological complexity there is and where it is located can vary enormously from language to language. It is the structure of this variability to which I turn in the remainder of this chapter.

2.2 The structure of ‘complexity space’ in morphology

While some languages display very little organization that is autonomously morphological, others provide us with rather more to explore. Three North American languages that are notably robust in their morphology are exemplified in (3).²

² My thanks to Marianne Mithun for the Central Alaskan Yupik and Mohawk examples here.

- (3) **Kwakw'ala:** huṣ^w-sanola-gil-eł
 vomit-some-continuous-in.house
 some of them vomit in the house
- Central Alaskan Yupik:** Piyugngayaaqellrianga-wa
 pi-yugnga-yaaqe-lria-nga=wa
 do-able-probably-INTR.PARTICIPIAL-1SG=suppose
 I suppose I could probably do that
- Mohawk:** wa'koniatahron'kha'tshero'ktáhkwen
 wa'-koni-at-ahronkha-'tsher-o'kt-ahkw-en
 FACTUAL-1SG/2SG-MIDDLE-speech-NMZR-run.out.of-CAUS-STATIVE
 I stumped you (left you speechless)

While obviously displaying more complex morphology than familiar European languages, these differ somewhat from one another. Kwakw'ala presents us with many word forms that incorporate a more diverse collection of information than we are used to in languages like English, but the individual components are relatively transparent and the degree of elaboration in words that occur in actual texts is moderate. 'Eskimo'-Aleut languages like Central Alaskan Yupik are commonly cited as falling at the extreme end of morphological complexity, because they make use of essentially open-ended combinations of meaningful elements to construct expressions of arbitrary complexity, but the individual components of a word are still relatively easy to tease apart, and the actual degree of complexity in common use is only somewhat greater than in Kwakw'ala. Iroquoian languages like Mohawk or Oneida are somewhat more intricately organized, and the complexities of combination are much harder to disentangle (see Koenig and Michelson, this volume). While the results are often very elaborate, there is a sort of upper bound imposed by the fact that, unlike the other two but similar to the Athabaskan languages, their morphology is based on a word structure template with a limited (if large) number of slots, such that the degree of complexity of the material filling any particular slot is (with some exceptions) bounded.

Other chapters of this volume present a variety of examples of complex morphological systems. My goal here is to characterize the logical space within which such complexity falls, the major dimensions along which languages elaborate the structure of words in ways that do not derive transparently from the essential nature of these linguistic elements. These fall into two broad categories: properties of the overall system, and characteristics of the relation between individual morphological elements and their exponents, ways in which the realization of content in overt form does not follow from the nature of either.

2.2.1 Overall system complexity

Morphological systems taken as wholes can differ in several ways. Some languages simply have more robust inventories of morphological material, more non-root meaningful elements (typically, but not always, affixes) than others. This difference is at least

logically distinct from the extent to which languages combine multiple morphological elements within a single word. And where such combined expression is found, the extent to which relations between elements (such as linear order within the word) can be predicted from their nature can also vary.

2.2.1.1 Number of elements in the system A basic sort of complexity derives from the simple matter of how much morphological elaboration a language makes available. In this regard, the languages of the Eskimo-Aleut family may be more or less at the extreme end: these languages commonly have more than five hundred derivational suffixes, and an inflectional system that involves at least as many more. The Salish and Wakashan languages of the Pacific Northwest are also rich in derivational morphology, though perhaps not quite to the same extent: Kwakw'ala, for instance, has about two hundred and fifty suffixes of this type (along with a few reduplicative processes), as documented by Boas (1947). English plays in a slightly lower league, though it is still not trivial: Marchand (1969) identifies about one hundred and fifty prefixes and suffixes in the language. Even Chinese languages, which are sometimes claimed to have 'no morphology', do in fact display some. Packard (2000) describes seven prefixes and eight suffixes in standard Mandarin, and provides arguments for thinking of these as morphological elements. Perhaps there are languages that are absolutely uncomplicated in this respect—Vietnamese is sometimes suggested, although this language exhibits extensive compounding, which is surely morphological structure in the relevant sense—but this is certainly an extremely rare state of affairs, if indeed it exists at all.

Even a system with a comparatively small number of markers may exhibit complexity of a different sort when the factors determining the choice of a particular marker and the conditions of exponence (e.g. as a prefix, suffix, or infix) are themselves complex, as in the case of gender marking in Archi (Chumakina and Corbett, this volume).

2.2.1.2 Number of affixes in a word The extent to which a language makes full use of its morphological capabilities can vary independently of the structure of the system itself. For example, the Eskimo languages all have more or less the same inventories of morphological possibilities, but some of them seem to put more weight on this aspect of their grammar than others. de Reuse (1994) observes that Central Siberian Yupik

'postbases are most often productive and semantically transparent, and can be added one after another in sequences of usually two or three, the maximum encountered being seven. These sequences are relatively short in comparison to other Eskimo languages, such as C[entral] A[laskan] Y[upik], where one can find more than six postbases in a word, and where it is possible to have more than a dozen.'

What is at stake here is a difference related to the change mentioned previously in the extent to which modern Kwakw'ala speakers rely more on syntactic than on morphological elaboration to express complex meanings, although the poten-

tial expressive capacity of the morphology remains unreduced. For comparison, Kwakw'ala is roughly similar to Central Siberian Yupik in the degree of observed complexity of individual words.

2.2.1.3 Principles of morphological combination Apart from sheer numbers of possible morphological elaborations of a basic stem, either in the size of the language's system or in what can be observed in individual words, another dimension of a language's morphological complexity is the principles that govern combinations of morphological markers. In many cases, the content (or 'meaning') of various parts of a word's morphology corresponds to a structure in which some elements take semantic scope over others. The most straightforward way in which the formal correspondents of these elements can be related is for their combination to reflect such scope relations directly. Where all of the markers in question are identifiable affixes, this is achieved by having these added one after another (working out from the root), with each one taking all of the material inside it (i.e. preceding if a suffix or following if a prefix) as its scope.

We can see this in Kwakw'ala, where the same affixes can combine in different orders depending on the meaning to be expressed as shown in (4).

- (4) a. 'cause to want':
 ne'nak^w-exsda-mas-ux^w John gax-ən
 go.home-want-cause-3SG John to-1SG
 John made me want to go home
- b. 'want to cause':
 q'aq'ol'a-madz-exsd-ux^w John gax-ən q-ən guk^wile
 learn-cause-want-3SG John to-1SG that-1SG build.house
 John wants to teach me to build a house

Here the order follows from the content properties of the elements involved, and so does not contribute complexity.

Contrast that situation with one in which the order of elements within a word is specified as an autonomously morphological property, rather than following from their semantics (or something else). This situation is often referred to in terms of morphological templates, such as what we find in the Athabaskan languages. An example is the templatic order of markers within the Babine-Witsuwit'en verb, given in (5) and derived from Hargus (1997, *apud* Rice 2000).

- (5) Preverb + iterative + multiple + negative + incorporate + inceptive + distributive # pronominal + qualifier + conjugation/negative + tense + subject + classifier + stem

This is actually one of the simpler and more straightforward template types found within the Athabaskan family (and is chosen here from among the many examples provided in Appendix I to Rice (2000) in part because all of the marker categories are

comparatively self-explanatory). For each of these languages, we can give somewhat similar templates, specifying the order in which morphological elements appear within a rather complex whole. The principles governing such templates are not wholly arbitrary, but various factors are involved including at least some stipulation: the ordering of element classes is partly based on semantics, partly on phonology (with prosodically weaker elements located closer to the stem), and partly arbitrary.

While the partial arbitrariness of such templatic morphology obviously adds complexity to the language in the sense being developed here, it is notable that such templates appear to be highly stable, at least grossly, over very long periods. Similarity in template structure, for example, is a significant factor in the evidence that supports a kinship between Tlingit-Athabaskan-Eyak on the one hand, and the Yeniseian languages of Siberia on the other (Vajda 2010).

We may ask what factors are ‘natural’ predictors of element order within words. Like other ordering relations we find in grammar, the relative order of morphological operations—commonly, but not exclusively represented by ‘morpheme order’—is governed by more than one principle, and these do not always agree.

Basic, of course, is the notion of semantic scope: a morphological operator is expected to take all of the content of the form to which it applies as its base. Another factor, though, is the typical relation between derivational and inflectional material, with the latter coming ‘outside’ the former in the general case. This relation has been asserted (in Anderson 1992 and elsewhere) to be a theorem of the architecture of grammar; a large and contentious literature testifies to the fact that this may need to be qualified in various ways, but in the present context it is only the general effect that matters.

There may also be rather finer-grained ordering tendencies of a similar sort (e.g. mood inside of tense inside of agreement, etc.), as suggested in various work of Joan Bybee (e.g. Bybee et al. 1994), although those also tend to have rather a lot of apparent exceptions. Linguistic theory needs to clarify the issue of which of these effects, if any, follow from the architecture of grammar, and which are simply strong tendencies, grounded in some other aspect of language.

Somewhat surprisingly, phonological effects also show up, as argued by Rice for various Athabaskan systems. This is an effect known from clitic systems: for instance, Stanley Insler argues (in unpublished work) that second position clitics in Vedic Sanskrit show regularities such as high vowels before low, vowel-initial clitics before consonant-initial, etc. Perhaps the appearance of similar effects in morphology is another example of how at least special clitics are to be seen as the morphology of phrases (Anderson 2005).

2.2.2 *Complexity of exponence*

The other fundamental ways in which morphological structure can contribute to linguistic complexity derive from the non-trivial ways in which individual

morphological elements and their surface realizations can be related. As argued by Stump and Finkel (this volume), the nature of this complexity depends in part on the perspective adopted, whether that of the speaker producing a complex form or that of the hearer attempting to recover the information it contains. The 'ideal' morphological element, with what might be called canonical realization, corresponds to the classical Structuralist morpheme, with a single discrete, indivisible unit of form linked to exactly one discrete unit of content. But as we know, real morphology in real languages is only occasionally like that, and commonly deviates from this ideal in a variety of ways.

2.2.2.1 Complexity in the realization of individual elements The simplest cases involve discontinuous aspects of a form that correspond to a single aspect of its content, including circumfixes and infixes. These are really two sides of the same coin, since an infixed form can be regarded as coming to instantiate a 'circumfixed' root. In both cases, a single morphological element has a discontinuous realization. Both, in turn, are simply the limiting, simplest variety of multiple exponence. Some examples are rather more exuberant than this, with my personal favourite being the way negation is multiply marked in Muskogean languages such as Choctaw. All of these are exemplified in (6)

- (6) **Circumfixes:** Slavey *ya-t̥* 'preach, bark, say'; cf. *yaht̥* 's/he preaches, barks, says', *xayadat̥* 's/he prayed', *náya'ewít̥* 'we will discuss' (Rice 2012)

Infixes: Mběngokre [Jé] -g- 'plural', cf. *fāgnān* 'to spend almost all (PL)', Sg *fānān* (Salanova 2012)

Multiple Exponence: Choctaw *ak̥iyokiit̥took* 'I didn't go'; cf. *iyalit̥took* 'I went' (Broadwell 2006)

In the Choctaw form, negation is marked in five independent ways: (a) substitution of a- for -li as 1SG subject marker; (b) prefixed k-; (c) suffixed -o(k); (d) an accentual feature of length on stem; and (e) suffixed -kii.

Just as real languages involve cases in which a single element of content corresponds to multiple components of the form of words, the opposite is also true: a single element of form can correspond to several distinct parts of a word's content, each signalled separately in other circumstances. This is the case of *cumulative* morphs, typified by the -ō ending of Latin *amō* 'I love'. Indeed, particular elements of form may correspond to *no* part of a word's content, in the case of 'empty' or 'superfluous' morphs. Conversely, a significant part of a word's content may correspond to no part of its form. The usual 'solution' to this difficulty for the classical morphemes is to posit morphological zeroes as the exponents of the content involved, but it is important to realize that this is simply a name for the problem, not a real resolution of it.

A variety of forms of morphological complexity introduced by these and other non-canonical types of exponence are abundantly documented in the literature, beginning

explicitly with Hockett (1947) and surveyed in Anderson (1992, to appear). These also include a variety of cases in which content is indicated not by some augmentation of the form, but rather by a systematic change of one of several sorts: subtractive morphology; Umlaut, Ablaut, and other kinds of apophony; consonant mutation; metathesis; exchange relations, and others. The bottom line is that languages abound in relations between form and content that are complex in the basic sense of violating the most natural way of expressing the one by the other.

2.2.2.2 Complexity of inter-word relations Complexities of exponence are not, of course, limited to those presented by the relation between form and content in individual words referred to in section 2.2.2.1. Another class of complications to the canonical involves the range of forms built on the same base—the paradigm of a given lexeme. Several chapters in this volume are devoted to the complexity of paradigms, and the issue is thus well illustrated elsewhere, so it will suffice here simply to indicate this as a contribution to morphological complexity overall.

The paradigm of a lexeme can be regarded as a structured space of surface word forms. The independent dimensions of that space are provided by the set of morphosyntactic properties that are potentially relevant to a lexeme of its type (defined by its syntactic properties); each dimension has a number of distinct values corresponding to the range of variation in its defining morphosyntactic property. Since each combination of possible values for the morphosyntactic features relevant to a given lexeme represents a different inflectional ‘meaning’, it follows that we should expect a one-to-one correspondence between distinct morphosyntactic representations and distinct word forms. To the extent we do not find that, the system exhibits additional morphological complexity.

Several types of complexity of this sort can be distinguished, and each is the subject of a literature of its own. *Syncretism* (Baerman et al. 2005) describes the situation in which multiple morphosyntactic representations map to the same word form for a given lexeme (e.g. [hit] represents both the present and the past of the English lexeme {HIT}). The opposite situation, *variation*, where multiple word forms correspond to the same morphosyntactic representation, is less discussed but still exists: e.g. for many speakers of American English both the forms [dowv] and [dajvd] can represent the past tense form of {DIVE}. In some cases, a paradigm may be *defective* (Baerman et al. 2010), in that one or more possible morphosyntactic representations correspond to no word form at all for the given lexeme. A fourth kind of anomaly arises in the case of *deponency* (Baerman et al. 2007), where the word forms corresponding to certain morphosyntactic representations appear to bear formal markers appropriate to some other, distinct morphosyntactic content (as when the Latin active verb *sequitur* ‘follows’ appears to bear a marker which is elsewhere distinctive of passive verbs). All of these types of deviation from the expected mapping between relations of content and relations of form contribute to morphological complexity.

2.2.2.3 *Complexity of allomorphy* To conclude this typology of the complexity introduced into language by morphological structure, it is necessary to mention the factors that determine how a given morphological element is to be realized. The simplest type here, of course, would be for each morphological unit to have a single, distinct realization in the forms of words in which it appears, but of course morphology has always attended to the fact that a single morphological element can take multiple shapes, the very definition of ‘allomorphy’. Allomorphy can contribute to the complexity of the system to varying degrees, though, depending on the bases of the principles underlying its conditioning.

Where the variation results from the independently motivated phonology of the language, this does not contribute additional complexity to the language in the sense under discussion here, as already noted in section 2.1.2. In other cases, though, although the conditions for allomorphic variation can be stated in purely phonological terms, the actual variants that appear are not predictable from the phonology itself. Thus, in Warlpiri the marker of ergative case is *-ŋku* if the stem is exactly bisyllabic, but *-lu* when added to stems that are trisyllabic or longer. A rather more extensive instance of such phonologically conditioned allomorphy is found in Surmiran (a Rumantsch language of Switzerland), where essentially every stem in the language takes two unpredictably related forms, depending on whether the predictable stress conditioned by its attendant morphology falls on the stem itself or on an ending (Anderson 2011).

Since phonological conditioning factors are, at least in principle, transparent, they contribute less complexity (again, in principle) than cases in which unpredictable allomorphy is based on specific morphological categories or on semantically or grammatically coherent sets of categories. These, in turn, appear less complex than ones in which the allomorphy is conditioned by (synchronically) arbitrary subsets of the lexicon, such as the Celtic mutations alluded to in section 2.1.2. Perhaps the summit of complexity with regard to the conditioning of allomorphy is the case where specific, unpredictable variants appear in a set of semantically and grammatically unrelated categories whose only unity is the role it plays in determining allomorphy. Such collections of categories, called ‘morphemes’ by Aronoff (1994), may recur in a number of rules within a language without having any particular coherence beyond this fact. See also Round (this volume) for a further elaboration of this notion and illustrations of several distinct types of ‘morphomic’ structure.

Another type of allomorphic complexity is presented by formally parallel elements that behave in different ways, something that has to be specified idiosyncratically for the individual elements. This is the issue of distinct arbitrary inflectional classes into which phonologically and grammatically similar lexemes may be divided. The members of a single word class, while all projecting onto the same paradigm space, may nonetheless differ in the ways in which those paradigm cells are filled. Conversely, affixes that are formally similar may induce different sorts of modifications in the

stems to which they attach, as we saw already in section 2.1.2 with the three affixal types in Kwakw'ala.

A related kind of complexity is found in languages where morphological elements can display one of several distinct types of phonological behaviour. In earlier theories such as that of Chomsky and Halle (1968), this was typically represented as a difference in the type of boundary associated with the element. Lexical Phonology incorporated this into the architecture of the grammar as the difference between the morphology of 'Level I', 'Level II', etc., where specific elements are characterized by level, distinguished from clitics. Included in this category perhaps should be the case of clitics attached at various prosodic levels, as in Anderson (2005).

2.3 The source(s) of morphological complexity

This survey of ways in which word structure contributes complexity to the grammars of languages naturally raises the question of where these things come from. As argued in section 2.1, they do not follow from the intrinsic nature of the task of mapping between content and form, so where do they come from?

Empirically, it seems clear that most of the ways in which grammars are morphologically complex arise as the outcome of historical change, restructurings of various sorts. Many of these fall under the broad category of 'Grammaticalization'. Canonically, this involves the development of phonologically and semantically reduced forms of originally independent words, leading eventually to grammatical structure. Originally full lexical items may generalize their meanings in such a way as to limit their specific content, leading to their use as markers of very general situation types. When this happens, they may also be accentually reduced, leading to further phonological simplification. This, in turn, may lead to their re-analysis as clitics, with an eventual development into grammatical affixes, and so new morphology is born.

But within linguistic systems, there are other possible paths that can lead to morphology where before there was only phonology and syntax. For example, phonological alternations, when they become opaque in some way, can also be reinterpreted as grounded in the morphology instead. The standard example of such a change is German Umlaut, and the overall pattern of development is quite familiar. A similar point can be made on the basis of the re-analysis of derived syntactic constructions, once they become opaque, being re-analysed as syntactically simple but morphologically complex (Anderson 1988).

Historical developments thus often yield systems that are more complex in morphological terms. But the opposite is true as well: when systems become more complex, that may trigger restructuring which reduces the complexity. Paradoxically, change produces complexity, but complexity can result in change.

It is sometimes assumed that morphology always has its origins in some other part of the grammar, particularly the syntax, as expressed in Givón's (1971) aphorism that today's morphology is yesterday's syntax. But there are examples where that cannot be the case, showing that morphology seems to have some sort of status of its own.

This is demonstrated particularly elegantly in phenomena found in a language that is of demonstrably recent origin, Al Sayyid Bedouin Sign Language as studied by Meir et al. (2010). In this language, more or less the entire history of the emergence of grammatical structure can be observed. The interesting point for our purposes here is that by the third generation of speakers of this language, morphological structure has begun to emerge in the form of regularities of compound formation. One of these is the generalization that in endocentric compounds, modifiers precede their heads: e.g. PRAY[^]HOUSE 'mosque'. This is hardly an exotic structure. But importantly, it is one that cannot have come from the syntax, because we also find that in syntactic formations, heads precede their modifiers. This demonstrates that this bit of morphology really is not entirely parasitic on other areas of grammar, however often the origin of specific morphology can be found elsewhere.

2.4 Conclusion

I started with what seems a logically plausible conception of what makes complexity: basically, some property of a system that cannot be derived from its essential character. When we look at the basic nature of human language, it seems to follow that *any* morphological property is of this sort, since syntax and phonology would seem to suffice unaided to fulfil the needs language serves. But given the pervasiveness of morphology in the world's languages, and the tendency of morphological structure to be created in linguistic change rather than being uniformly eliminated, this would appear to be a sort of *reductio ad absurdum* of this notion of complexity, as applied to language.

In particular, what seems complex to us as scientists of language may or may not pose problems for users of language. Strikingly, we find that little children seem to have no remarkable difficulty in acquiring languages like Georgian, or Mohawk, or Icelandic along more or less the same time course as children learning English or Mandarin. Of course, it might be that little children are just remarkable geniuses at solving problems that seem impenetrable to scientists. But it seems more likely that morphology, despite the fact that *a priori* it seems like nothing but unmotivated and gratuitous complication, is actually deeply embedded in the nature of language. Although morphological structure of any sort would seem to be a serious challenge to the notion that human languages are 'optimal' solutions to the problem of mapping content to form, morphology seems to be a fact of life—and a part of the human language faculty. And that has to give us pause about our ability to say anything serious about what is or is not complex in morphological systems in any deep and basic sense.

Acknowledgements

I am grateful to audiences at a SSILA special session on morphological complexity in the languages of the Americas at the Linguistic Society of America Annual Meeting in Portland, Oregon in January 2012, and at the Morphological Complexity Conference in London, for discussion of this material. I am especially grateful to Mark Aronoff, Andrew Carstairs-McCarthy and Marianne Mithun for ideas, examples, and comments incorporated here. My work on Kwakw'ala was supported by grants from the US National Science Foundation to UCLA.