

1 *Options in constructing a morphological framework*

1.0 Locating generalizations

Morphology is pervaded by varying degrees of exceptionality, and any morphological framework must be able to accommodate morphology’s highly non-systematic nature while defining in precise terms the systematic facts. Success in meeting this aim depends partly on finding the right generalizations. And that in turn depends on where we decide to ‘pitch’ generalizations. For instance, suppose we wanted to get a better understanding of an under-documented language’s verbal system. As more data become available, we can begin to generalize over patterns of form and meaning. For example, in Shughni, a threatened Eastern Iranian language spoken mostly in the mountainous Badakhshan region of Tajikistan, the following sentences and their analyses would lead to several initial generalizations.¹

- | | | |
|-----|---|--|
| (1) | a. Present intransitive
wuz wirāfc-um.
I stand-1SG
‘I stand.’ | b. Present transitive
wuz kud win-um.
I dog see-1SG
‘I see a dog.’ |
| (2) | a. Present intransitive
tu wirāfc-i.
you (sg) stand-2SG
‘You stand.’ | b. Present transitive
tu kud win-i.
you (sg) dog see-2SG
‘You see a dog.’ |
| (3) | a. Present intransitive
yu/ya wirofs-t.
he/she stand-3SG
‘She stands.’ | b. Present transitive
yu/ya kud wīn-t.
he/she dog see-3SG
‘She sees a dog.’ |

As the examples above indicate, the language has agreement, specifically head marking where formatives expressing person and number agreement with the subject argument are aligned to the right of a verbal root. One way of generalizing over these data is to view both verb stem and (agreement) formative as lexical entries. So, for example, *-um* is a lexical entry with the meaning

2 Options in constructing a morphological framework

‘first-person singular agreement’, *-i* is a separate entry with ‘second-person singular agreement’, and *-t* yet another lexical entry. The formative lexical pieces would be distinguished from free morpheme lexical entries such as *kud* ‘dog’ and the first-person personal pronoun *wuz*, and have to carry instructions about combinability to capture the fact they are bound, and more specifically bound to the right edge of verb pieces. Something akin to a subcategorization frame could be employed to account for agreement formative morphotactics (e.g. Lieber 1992).

When an attempt at analysis assumes that the basic object of enquiry in morphology is the morpheme, then the morphological generalizations being sought are akin to those that hold for phrase structure. By placing phrase structure and word structure in the same (or nearly the same) problem space, the natural expectation is that they also occupy the same (or nearly the same) solution space. In other words, if word structure is viewed as a kind of phrase structure, then principles that apply to the latter may also apply to the former. In such an approach, generalization is located at a relatively high level and holds over different types of structure, including word structure. For example, subcategorization frames used to constrain possible *syntactic* configurations could also be used to express configurations of lexical pieces below the word level. Generalization at this level carries a number of expectations. Returning to Shughni, we expect that if plural subjects trigger plural agreement, this will be manifested by a formally differentiated set of affixes that can be treated as counterparts to singular lexical entries. In other words, there is a bias in our expectations towards affixation, i.e. that the plural exponents will also be affixes; and there is symmetry in function and form, i.e. one affix type is for singular and another for plural. Another expectation will be about linearization: if singular affixes are right-aligned with respect to the verbal head, the plural affixes will be too. That is, the expectation is not just affixation but suffixation. If linearization plays a role in phrase structure where syntactic constituents of the same type are distributed in the same way, lexical entries of the same type (root type, functional type, functional involving agreement subtype, etc.) should share similar positions in a complex word.

Another approach to the Shughni data is to try to pitch generalization at a lower level. Namely, a verb’s *paradigm* of word forms is the starting point for generalization. The present-tense subparadigms of ‘stand’ and ‘see’ are given in (4).

(4)

	'stand'		'see'		pattern	
	SG	PL	SG	PL	SG	PL
1st	wirāfc-um	wirāfc-ām	win-um	win-ām	-um	-ām
2nd	wirāfc-i	wirāfc-et	win-i	win-et	-i	-et
3rd	wirofs-t	wirāfc-en	wīn-t	win-en	-t	-t

Association of meaning and form occurs through a stem’s place in a paradigm of *cells*, groupings of any number of morphosyntactic feature values, including a single feature value. So *wirāfc-um* in (4) occupies the cell which is the grouping of {NUM:SG} and {PER:1}.² In this approach, the suffix *-um* is not a lexical entry that pairs form and meaning; rather it is an exponent of a particular morphosyntactic grouping. The generalization being made is found in the third column: a pattern of exponence of morphosyntactic groupings shared by two verbs. In paradigm-based approaches, it is this abstracted pattern to which *paradigm* refers (Spencer 2004: 7).³ Just as in the first approach, meaning–form association as well morphotactics are being encoded in the generalization (exponents are identified as suffixes in the third column). But because the formatives are not lexical entries the paradigm-based approach does not entail a syntax of words, hence generalizations are at a ‘lower’ level. Word structure is not similar to phrase structure and thus should not be expected to have syntactic correlates. For example, in the paradigm of ‘see’, the exponence of {PER:3, NUM:SG} is complex, consisting of the affix *-t* together with a change in the feature of the stem’s vowel from [– long] to [+ long]. Multiple exponence of a single function is not the expected situation in syntax,⁴ and non-concatenative operations seem to be the reserve of morphology. At the same time, the special character of morphology is itself subject to generalization. Vowel lengthening in the stem used for the third-person singular present is found elsewhere in this verb’s paradigm, namely in the past tense, as shown in (5).

(5)

wuz=um	kud	wīn-t.
I=1.SG	dog	see-PST
‘I saw a dog.’		

Using the paradigm of ‘see’ as the location of generalization allows us to capture the fact that any process that operates over the stem in the {PER:3, NUM:SG, TENSE:PRS} (3) does the same in a subset of cells elsewhere in the paradigm, namely those containing {TENSE:PST} (5). This is a special kind

4 Options in constructing a morphological framework

of word-structure behaviour found in a particular Shughni verb that could itself be generalizable across all verbs in the language, or at least a subclass of verbs.

1.0.1 Paradigmatic relations

The {PER:3, NUM:SG, TENSE:PRS} = {TENSE:PST} generalization has the paradigm as its starting point. A paradigm-based approach to word structure attempts to discover generalizations about paradigmatic organization. To illustrate, consider the paradigms of four Russian nouns *zakon* ‘law’, *karta* ‘map’, *rukop’is’* ‘manuscript’ and *boloto* ‘marsh’.⁵

(6)

	I	II	III	IV
<i>Singular</i>				
NOMINATIVE	zakon	kart-a	rukop’is’	bolot-o
ACCUSATIVE	zakon	kart-u	rukop’is’	bolot-o
GENITIVE	zakon-a	kart-i	rukop’is’-i	bolot-a
DATIVE	zakon-u	kart-e	rukop’is’-i	bolot-u
INSTRUMENTAL	zakon-om	kart-oj	rukop’is’-ju	bolot-om
PREPOSITIONAL ⁶	zakon-e	kart-e	rukop’is’-i	bolot-e
<i>Plural</i>				
NOMINATIVE	zakon-i	kart-i	rukop’is’-i	bolot-a
ACCUSATIVE	zakon-i	kart-i	rukop’is’-i	bolot-a
GENITIVE	zakon-ov	kart	rukop’is’-ej	bolot
DATIVE	zakon-am	kart-am	rukop’is’am	bolot-am
INSTRUMENTAL	zakon-am’i	kart-am’i	rukop’is’am’i	bolot-am’i
PREPOSITIONAL	zakon-ax	kart-ax	rukop’is’-ax	bolot-ax

A characteristic of the Russian noun morphology system is that there are different patterns of exponence, expressed as four major declensional classes, which partition the noun lexicon.⁷ In such a system the simplest organizing principle would be that each class has a unique set of forms to set it apart from the other classes: a ‘canonical’ system of inflectional classes (Corbett 2009). So looking horizontally across a particular cell pausing at, for instance, {NUM:SG, CASE:INS} there would be four different values because there are four different patterns of exponence. Rather surprisingly this does not happen here: for Declension I the suffix *-om* is used; for Declension II *-oj*; for Declension III *-ju*; and Declension IV uses the same *-om* as Declension I. Even more surprising is that there is not a single cell where a four-way distinction is made. Another expectation is that within

1.0 Locating generalizations 5

a class each cell would be different from the other, so for example forming a nominative singular is different from a nominative plural. While there is a tendency for *vertical* distinctions across cells, it is only a tendency. So for Declension II, dative singular is in *-e*, but so is prepositional singular. In fact, in the world's languages that have inflectional classes fully horizontal and fully vertical distinctions are rare (Corbett 2009). The paradigm-based approach addresses departures from the canonical inflectional class situation in terms of special *horizontal* or *vertical* relations. For instance, lack of distinction in vertical relations is expressed in terms of syncretism, a topic explored in Chapter 4.

1.0.2 Semi-regularity

We began with the observation that morphological facts may be the products of a system or may fall outside it. Another way of putting this is that a fact may or may not be subject to a given generalization. In the paradigm-based approach, the pattern of inflections is a generalization holding for a given subclass of lexical entries, as illustrated with the facts about Russian noun morphology. But there are items in Russian, as in many languages, whose pattern of exponence deviates in one way or another from what is included in the generalization. We can consider three different situations. The word for 'soldier' *soldat* has nominative singular *soldat* and genitive plural *soldat*. Such a pattern falls outside Declensions I to IV in (6). As a second instance, the word for 'person' is *čelovek* in the nominative singular but *l'ud'i* in the nominative plural. Here the stem that is used in the singular is exchanged for a wholly distinct stem in the plural. And third, the word for 'sledge' *san'i* has no singular morphology; rather it uses plural word-forms in both plural and singular contexts. In one view each instance is non-systematic, since there is a break with the noun inflection generalization given in (6). Alternatively we could think of them as *partially* systematic, since they are not completely disengaged from the inflectional system. So *soldat* behaves like a Declension I noun except for the genitive plural cell, and *san'i* is a perfectly regular Declension I noun with respect to plural forms. In terms of pattern of exponence, *čelovek* does whatever Declension I nouns do. Not shown in (6) is the genitive/accusative syncretism characteristic of animate nouns (we see this later in Chapter 2, Section 2.3.7); *čelovek* as an animate shows this syncretism.⁸ Much of morphological irregularity is like this and is therefore better treated as *semi-regularity*. The paradigmatic system is still relevant, but it is the non-standard way in which it is engaged that makes an item look as though it is excluded from the generalization. If the

6 Options in constructing a morphological framework

generalization is pitched at the level of the paradigm, then accommodation can be made for semi-regularity.

1.0.3 Options and chapter outline

Network Morphology is a paradigm-based framework: morphological generalizations are gathered at the level of the paradigm. In this chapter we create a profile of Network Morphology by outlining options available to a morphological framework and showing which ones are taken. We can think of where to locate generalization – at the morpheme level or at the paradigm – as the first set of options. Other options largely follow from the choice made here. In Section 1.1 we consider options in what is taken to be the fundamental object of enquiry in morphology. A paradigm-based approach entails that this is the lexeme rather than the morpheme. Choosing the lexeme entails other options; for example, that the approach is also *inferential* rather than *lexical*. In Section 1.2 we look at options in how to conceive morphology in relation to the rest of the grammar. We contrast the notion of an autonomous component for morphology, adopted by Network Morphology, with the alternative that the grammar is more like a seamless web where there are no boundaries between syntax and morphology. The alternative naturally follows from the option that generalization is at the morpheme level, such that principles of word structure could in theory be principles of phrase structure. A good representative alternative is Distributed Morphology, and this theory therefore receives most of our attention. The nature of generalization is explored in Section 1.3, where we discuss the concepts of *inheritance hierarchy* and *network* that are fundamental to Network Morphology's way of capturing generalizations, as well as accommodating semi-regular cases. Exactly how inheritance is interpreted provides further options: mandatory or default inheritance, single or multiple inheritance. Network Morphology is a *formal* framework: it is computer interpretable due to the lexical knowledge representation language in which its theories are expressed, the DATR language. Formalization as an option for morphological frameworks is briefly discussed in Section 1.4. A summary of the options taken (and not taken) by Network Morphology is presented in Section 1.5, a character profile of the framework. How this profile is projected into the chapters of the book is briefly outlined, too.

1.1 Object of inquiry: lexeme rather than morpheme

All paradigm-based frameworks adopt the *lexeme* rather than the morpheme as minimal sign, for example A-morphous Morphology (Anderson 1992) and

1.1 Object of inquiry: lexeme rather than morpheme 7

Paradigm Function Morphology (Stump 2001). The lexeme is a unity of syntactic, semantic, phonological and (if relevant) morphological class properties that makes one word distinct from another. We could view the lexical entry for the Declension I noun *zakon* ‘law’ schematically as in (7).

- (7) ZAKON
 syntactic level
 syn cat = Noun
 semantic level
 ‘law’
 phonological level
 stem = /zakon-/
 morphological level
 mor class = Class_I

Levels of lexical representation play different roles (Aronoff 1994). The syntactic level determines what phrase-level configurations the item can appear in; the semantic level indicates its broad meaning as well as any features that can be used to express selectional restrictions; the phonological level defines its basic sound form as a *stem*; and finally the morphological level flags its inflectional class.

The morphological level of representation determines the shape of the lexeme within a given syntactic context because it serves to locate the pattern of correspondences that hold between a cell’s *content*, morphosyntactic feature combination, and *form*, exponent. Generalization at the paradigmatic level assumes that the correspondences are rule-based. A rule of exponence is triggered by the morphosyntactic feature combination of the cell that the stem occupies. So, for example, a stem occupying a cell whose content is {NUM:SG, CASE:DAT} triggers a *-u* suffix exponent if the lexeme carries Declension I at its morphological level (6). Therefore, generalization at the level of the paradigm entails lexemes instead of morphemes as the basic morphological units and an *inferential* rather than a *lexical* approach to word structure (Stump 2001). The approach is inferential since the exponents of the content are introduced by rule, rather than being lexical objects or vocabulary items. Thus, the complex word is *inferred* from the lexeme’s stem.

1.1.1 Radical agglutination

In (6) above there are a number of cells lacking an exponent. Exponentless cells are in fact found in all four classes. For Declension I these are {NUM:SG, CASE:NOM} and {NUM:SG, CASE:ACC}, for Declension II {NUM:PL, CASE:GEN}, for Declension III {NUM:SG, CASE:NOM} and {NUM:SG, CASE:ACC}, and finally

8 Options in constructing a morphological framework

for Declension IV {NUM:PL, CASE:GEN}. A paradigm-based framework has no stake in pervasive exponence since exponence serves a contrastive end rather than constituting the phonological part of a morpheme lexical sign item. But in a lexical entries as morphemes approach, word structure is assumed to be compounding, entailing what Spencer (2004: 76) calls *radical agglutination*, a prerequisite to the compositionality of the complex unit. Just as syntactic compositionality requires consistency in the function:form mapping of words filling terminal nodes in phrase structure trees, so a one-meaning:one-form mapping has to be assumed for the affixes and stems making up a complex word. The reality is that deviations from radical agglutination abound in just about every way possible. Spencer (2004) provides the following deviation taxonomy.

(8)

Meaning	Form	Term
<i>Radical agglutination</i>		
0. one	one	–
<i>Radical agglutination deviations</i>		
1. one	zero	zero morphs
2. zero	one	meaningless morphs
3. > one	one	cumulation
4. one	> one	extended exponence

We have already come across type 1 deviation, zero morphs, in the course of our discussion about Russian nouns where the meaning ‘singular nominative/accusative’ lacks an exponent for Declension I and III nouns, and ‘plural genitive’ is exponentless for Declension II and IV. We can also see the Shughni and Russian data as examples of type 3, cumulation, since more than one feature maps onto a single form. In the Shughni verb examples, these are number and person features, and for the Russian data number and case features. Spencer is careful to point out that systems that are predominantly agglutinative do not escape deviation. Past-tense forms of the Finnish verb ‘to be’ are shown in (9) (Haspelmath 2002: 33).

(9)

Finnish past tense ‘to be’	
ol-i-n	‘I was’
ol-i-t	‘you (sg) were’
ol-i	‘s/he was’
ol-i-mme	‘we were’
ol-i-tte	‘you (pl) were’
ol-i-vat	‘they were’

1.1 Object of inquiry: lexeme rather than morpheme 9

The stem *ol-* lines up with verb lexical semantics, the formative /i/ maps to ‘past’, leaving the remainder of the complex word to be interpreted as person/number feature values. In the third-person singular, we clearly have an example of a zero morph (type 1 deviation). But throughout the paradigm, we also have type 3 deviation (cumulation) going on where more than one feature maps onto a single form. There are no dedicated ‘pieces’ for number and person values. Rather values are fused onto one suffix: number is contrasted when we compare formatives /n/ and /mme/, but person is contrasted when we compare /n/ with /t/.

A kind of reverse situation is type 4, extended exponence. In (4) we saw the third-person singular of the Shughni verb *win-* ‘to see’ as an example of this where, in addition to the *-t* suffix, the verb stem’s vowel is lengthened. In other words, the meaning has two different phonological reflexes, albeit of different kinds: a suffix and an alternation of the vowel. Stump (1990, 2001) provides an example of extended exponence where the two segments are not only of the same kind – both affixes – but are actually identical. In Breton ‘boat’ is *bag* and its plural *bagou*. However, in the diminutive formed in *-ig*, the *-ou* suffix shows up twice: *bagouigou* meaning ‘little boats’ as opposed to *bagig* ‘little boat’. Examples of this type are counterexamples to what Lieber terms Redundancy Restriction (Lieber 2004: 161–5), a constraint preventing more than one affix lexical entry from introducing the same meaning in a given word. Lieber gives the derivational example *dramatical*, which if compared to analogous *theatric* suggests the *-al* formative is redundant. In fact the form *dramatic* could be viewed as the redundant free co-variant, except even here there is no meaning for the /t/ formative to map to if the base is *drama*. This is an example of type 2 deviation where /t/ and /al/ are meaningless morphs.

Spencer (2004) notes that it is not a breakdown in isomorphism per se that calls into question the affixes-as-words approach. Cases of allomorphy, as found in inflectional classes like those in Russian nouns, are affixal analogues of synonymy; affix syncretism is akin to lexical homonymy or polysemy (systematized homonymy). Even zero morphs have a parallel life in the stem world, albeit marginally. In Basque the copula *ezan* has a ‘zero’ stem in the third-person singular (10). This example has been described as a special instance of stem suppletion (Hippisley, Chumakina, Corbett and Brown 2004: 414).

(10)	Basque copula <i>ezan</i> (see Hualde and Ortiz de Urbina 2003)	
	SG	PL
1st	n- in -tzen	g- in -en
2nd	h- in -tzen (informal)	z- in -eten
3rd	z- Ø -en	z- ir -en ⁹

10 Options in constructing a morphological framework

However, when stems are collocated within a compound (or within a phrase as syntactic words) the meaning:form deviations of types 2, 3 and 4 do not occur.

Because in the paradigm-based approach the meaning–form association is computed by a stem’s place in a particular cell of features and the exponence of that particular situation, Spencer (2004) argues that deviation type 2, meaningless morphs, is entailed by this approach. In other words, exponents do not carry meaning in the same way as words do; they are, as we have said, a mechanism for making a contrast amongst different feature combinations. So provided their contrasting service can be characterized, they can be multiple (deviation 4), zero (deviation 1) or double-duty (deviation 3).

1.1.2 Morphemes and word syntax

When morphemes are the basic object of morphological enquiry, generalizations about phrase structure could be expanded to cover word structure, as has been proposed in word-syntax approaches to morphology, for example Williams (1981), Selkirk (1982) and Lieber (1992). As noted in Lieber (2000: 408), morphotactics in early versions of phrase structure grammar were defined by a special set of rewrite rules. The apparatus used in modern generative grammar could also be pressed into the service of morphological description. For example, in the morpheme-based theory of Lieber (1992) word constituency is subjected to an X-bar treatment with typing of word constituents as heads, complements and specifiers.¹⁰ So in the compound *cat lover* the first element is the complement of *lover* (the head) evidenced by the phrasal equivalent ‘lover of cats’. In the compound *filing cabinet*, on the other hand, the first element is a modifier. Turning to a derivational example, in *happiness* the head is the suffix and the base neither complement nor modifier, so therefore must belong to the functionally ‘heterogeneous’ specifier class (Lieber 1992: 55).¹¹ Just as lexical heads assign theta roles to their complements in phrase structure, so does the head of a word piece to its complement piece. In examples like *debug* and *encase* the category-changing prefix is the head and the noun root its complement. The theta role assigned to the complement is Theme in the first example, i.e. ‘remove the bug [+Theme] from X’, and Location in the second, i.e. ‘put X in a case [+Loc]’ (Lieber 1992: 57). Semanticsyntactic interpretation is simply a matter of computing the semanticsyntax of each morpheme lexical entry. Since all word structure (compounding, derivation, inflection) is endocentric like phrase structure, heads such as *lover* in *cat lover*, *-ness* in *happiness* and *de-* in *debug* determine the category of the word as a whole through feature percolation to the word’s top node, where priority is given