INTRODUCTION TO MORPHOLOGY

© 2002 Mark C. Baker and Jonathan David Bobaljik

Rutgers and McGill

Draft – for class use

1. BASIC CONCEPTS

Recall from the introduction that two fundamental questions for a theory of morphology are:

- 1. What is the relationship between phonology and semantics in words?
- 2. Does this differ from the relationship between phonology and semantics **among words** / in sentences (= syntax) ?

In this chapter we will lay out some basic ideas and give preliminary definitions of some of the terminology that we will use. Some of what we present here will be revised in the course of the textbook, but we hope that this will provide a reasonable starting point. Keeping in mind that the idea of this book is to lead you, the reader, by example through the process of theory construction, this chapter is primarily concerned with the kinds of considerations that might make a linguist think that there is something to have a theory of, i.e., questions to be asked and answered.

1.1 Morphemes

1.1.1 Arbitrariness and Lexical Entries

If one were forced to sum up the fundamental questions of linguistic inquiry in a single glib question, it might be something like that in (1).

(1) What characterizes the relationship between sound/signal and meaning/function?

In certain cases, as famously discussed by the French linguist Ferdinand de Saussure, the relationship is arbitrary. The fact that the meaning 'a domesticated carnivorous mammal (*canis famliaris*) related to the foxes and wolves' is tied to the phonological string /dog/ is a peculiar property of English, a mere accident of the language's history. There is no logical reason for this particular pairing of sound and meaning, and clearly it is not universal: the same meaning is tied just as intimately to quite different phonological strings in different languages, such as French / $jj\ell$ /, Japanese /inu/, and Itelmen /qas χ /. Likewise, the same basic phonological string is tied to different meanings in other languages: the string /tak/, which in English means something like 'to converse by means of spoken language,' means 'day' in German, 'mountain' in the Turkic language Uighur, and 'so' or 'thus' in Russian. Thus, a part of learning English (or French, or

Itelmen, or Uighur) is learning the arbitrary pairings of sound and meaning, and storing these in a sort of mental dictionary. This mental dictionary is called the **lexicon** and at a minimum, we know from the arbitrariness just discussed that entries in the lexicon (henceforth **lexical entries**) have two parts: a phonological **representation** and a **semantic** one. Following convention, we will use **IPA** to indicate the phonological representation (discussed more in chapter P), while for the semantic representation in place of a dictionary-like definition we will often just write the corresponding English word in ALL CAPS. Thus "DOG" should be read as "whatever the word 'dog' means". Some lexical entries for some simple English words are given in (2).

(2) English lexical entries (preliminary):

Label	"dog"	"fish"	"teach"
phonology:	/dag/	/fɪʃ/	/tijt∫/
semantics:	DOG	FISH	TEACH

Note that our lexical entries have three parts. In addition to the phonological and semantic representations we have discussed above, we have given each pairing a label. The label has no significance in our theory and we do not mean to assert that such a label is part of the mental representation. For now, the label is simply a convenient device for referring back to a lexical entry without privileging either the phonological or semantic representation. Note that in casual discussions, we often talk about 'the word *dog* which has the meaning such-and-such'. This way of speaking suggests that the phonological representation is basic and that the meaning is attached to the phonology. This makes potentially misleading assumptions about the nature of mental representations, assumptions for which there is no evidence a priori. Using a label is intended to keep this clear in our discussions. Note that we have used English orthographic words as labels, but this is just for convenience; we could just as easily have used numbers, or any other device that would serve to formally identify each lexical entry and keep it distinct from the next.

1.1.2 Compositionality

Not all pairings of sound and meaning are completely arbitrary. In fact, by far the majority of expressions that we encounter in language involve non-arbitrary connections between phonology and semantics. For example, the association of the phonological string in (3) with the particular meaning it has is to a large degree not arbitrary, and it would be incorrect to suppose that this string is represented with a single lexical entry as an arbitrary pairing of sound and meaning. Indeed, you should know the meaning of this expression even if you have never heard this exact expression before and therefore cannot possibly have memorized the connection. The meaning associated with (3) is compositional, a predictable function of the meaning of its parts.

(3) /pæts tijtʃər wakt kwiklij/ Pat's teacher walked quickly.

The compositionality of the meaning in (3) is manifest at two levels. At one level, the meaning of the entire utterance is a function of the individual meanings of each of the four words. At another level, the meaning of each of the four words is in turn compositional, a function of the individual parts of each word. The word *Pat's* contains the word *Pat* (which happens to be a proper name) and the meaning of the word *Pat's* is a function of the name

Pat and the marker of possession (represented orthographically as the sequence apostrophe + s). Likewise, the word *teacher* contains the word *teach* and the meaning of the word *teacher* is compositional, the product of its pieces.

Standard practice divides these two levels of compositionality into two domains. The combination of independent words into clauses and sentences, and the resulting meanings, are the domain of Syntax, while combinations within words are said to be the domain of Morphology. A question that we will we raise later in the textbook, but which we will put aside for now, is whether this division reflects some actual modularity in linguistic competence. This has been a point of no small debate within many different linguistic traditions, but since the question is really about whether grammatical principles are the same within words as among them, we may fruitfully—indeed we must—restrict our attention here to word-internal processes in order to have something to compare to the syntax.

1.1.3 Complex and simple words

Above, we have identified an important division among two classes of words. Some, like *Pat's*, *teacher*, and *quickly* are *morphologically complex words*. We assert this because their meanings are not arbitrary, but rather in large part a function of their parts. At a minimum, what it means to be a *teacher* depends on the meaning of the verb *teach*. Other English words that are likewise morphologically complex are given in (4) - (6); for each of these the meaning of the whole word is not entirely arbitrary, but depends on the meanings of the smaller words, contained in them.

- (4) teacher, presidential, ungrammatical, antidisestablishmentarianism
- (5) hospitalize, hospitalization; refutable, irrefutable, irrefutability
- (6) chalkboard, snowboard, greenhouse, blackberry, workshop

These morphologically complex words differ in this regard from the words in (7), which we may call *morphologically simple words* (or just simple words, for short).

(7) dog board black berry spider spit teach fish chef widow

Now, as a first characterization of speakers' intuitions, this division into complex and simple words may seem straightforward, but reflection on the examples given leads to some important questions. For example, we were happy to call *teacher* complex because it contains the word *teach* (along with some other stuff), but doesn't *spit* contain *pit* which is also an English word? Similarly, we said that *teacher*—which means roughly 'one who teaches'—contains the meaning of *teach*, but in that case, might we not be tempted to say that *chef* contains the meaning of *cook*, and maybe that *widow* contains the meaning *woman*? Yet we classified *spit*, *chef* and *widow* as simple words. Was something wrong with our classification above, or for the basis of it?

Let's look more closely at the different examples and try to get a sense of what our initial intuitions were telling us.

1.1.4 The parts of words

In what way can we assert that *teacher* contains the word *teach* but that at the same time *spit* does not contain the word *pit*. In both cases, the phonology (sound) of the shorter word is contained in the phonology of the longer word. But as we have seen above, a word is not just a string of sounds. Both /tijt \int / and /kijt \int / are strings of sounds in English, and both are possible words—they violate no rules of sound combination in English (i.e., **phonotactics**)—but only one is an actual word of English, namely /tijt \int /. What is it that makes this an actual word of English? It is that this string of sounds is paired with a meaning, whereas the string /kijt \int / is not (except, perhaps as a proper name). So, a word is not just a string of sounds, it is a **pair of sound and meaning**, of phonology and semantics. When we want to assert that *teacher* contains the word *teach*, what we really mean is not just that string of sounds /tijt \int / (or the corresponding orthographic string) is contained in the word /tijt \int -r/, but more importantly that the pair of sound and meaning that constitutes the word *teach*—that is, the lexical entry of "teach" from (2)—is contained inside the pair of sound and meaning that constitutes the word *teacher*.

Now we can see why we can distinguish the words in (4) - (6) in a principled manner from the words in (7). Even though the word *spit* contains the string of sounds /pɪt/ and even thought the same string of sounds does correspond to an English word, the word *spit* does not contain the (pairing of sound and meaning that constitutes the) word *pit*. In this manner, the word *spit* is simple, it does not contain another word inside it.

We can now return to the example of *chef*. The same logic as we used with *spit* applies here. The meaning of *chef* may involve or include the meaning of *cooking*, but the word *chef* does not contain the word *cook*, because one half of the sound-meaning pair is absent, in this case the sound.

Finally, we can observe that a word may be simple from a morphological perspective (i.e., containing no other words), but it may nevertheless be made up of linguistically relevant pieces. The word *berry* /beri:/ for example contains four phonemes grouped into two syllables, which are in turn collected into a metrical foot. Studying these pieces and how they can and cannot interact is an important part of linguistics (**phonology** and **phonetics**), but from the perspective of word-formation that we are interested in, these pieces (phonemes, syllables, feet) do not directly enter into the pairings of phonology and semantics. From the perspective of phonology-semantics pairings, *berry* is indivisible; an arbitrary pairing of sound and meaning.

It follows that each of the simple words in (7) corresponds to a lexical entry and should be represented in the lexicon in the manner illustrated in (2).

1.1.5 Morphemes - the pieces of complex words

Let us now return to the complex words in (4) - (6). We have said that they are complex in the specific sense that they have other words inside them, where we take words to be pairings of phonology and semantics. If we look at examples like *chalkboard*, *blackberry*, *snowboard*, etc., we see that these contain two words each, and that although the specifics vary from example to example, the meaning is in general a function of the pieces. Thus a *chalkboard* and a *snowboard*

are both kinds of boards, one associated with chalk, the other with snow, although the precise nature of the relation "associated with" cannot be determined just from the fact that the words are combined, but must be supplemented with our knowledge of the real world. Complex words that contain more than one word (more accurately, more than one **root**, see below) in this manner are called **compounds** and will form the subject of Chapter C and part of Chapter Arg.

Not all the complex words in (4) - (6) are compounds. The words *teacher*, *hospitalize*, and *ungrammatical* are complex, but when we extract the smaller words contained in them, the remaining pieces are not themselves words (or even roots). For example, while *chalkboard* is made up of the two words *chalk* and *board* and nothing else (that we can see), the word *teacher* is made up of the word *teach* and the piece -er, and *ungrammatical* of the word *grammatical* and the piece *un*. What are these pieces? Clearly, they have phonological properties—do they have anything else that may be relevant? In particular, is there anything we can say about them in terms of meaning? To answer this, consider the examples in (8).

(8) teacher, driver, lecturer, reader, softener, whitener, blender, modernizer ...

The first word we've seen before, and we've said that the meaning of the whole *teacher* contains the meaning of the embedded word *teach*. The same is true of the other words in the list. But we can be more precise. The meaning of the whole is in each case related to the meaning of the embedded word in a straightforward and rather narrow manner; specifically, the complex words in (8) all mean "one who does X" (or "that which does X") where X stands for the verb that is contained inside the complex word. This is a pattern; it is systematic. We see a recurring phonological pattern (in this case the string of sounds represented orthographically as *er*) with a recurring meaning or function. Moreover, we can demonstrate that this pattern is not an accident of the particular words we happen to have chosen in (8). This pattern is **productive** in the sense that we can **generate** and **understand novel forms**. As new words enter English as verbs, English speakers automatically know how to produce and understand the words that correspond to "one who does X" or "that which does X". Some verbs which have entered English relatively recently are given in (9), and the corresponding nouns are given in (10).

- (9) a. I need to **debug** this computer program.
 - b. Do you **snowboard**?
 - c. I saw Krum **blatch** Harry.
- (10) a. I need to run a **debugger**.
 - b. My sister is an excellent **snowboarder**.
 - c. Krum is a regular **blatcher**.

That this is automatic can be shown by using **nonce** words (words that are made up for a single occasion, for example to illustrate a point); given any sentence of the form in (11a) an English speaker will be able to correctly fill in the blank in (11b), or vice-versa, as in (12), even without knowing the meaning of the nonce words involved.

- (11) a. My friend earns her living by *kreeling*.
 - b. She is a professional _____.

- (12) a. Fido is a good *blicker*.
 - b. I see him _____ all the time.

What these examples illustrate is that the pattern relating *teach* and *teacher* is systematic and that the meaning of complex words in (8) is a function of the suffix *-er*. Similar patterns emerge when we consider the pieces of the other complex words in (4)-(6). Thus *un*- occurs to the immediate left of adjectives in complex words with the meaning not X *ungrammatical, untrue, unhappy*, etc., *-able* occurs in complex words meaning roughly 'able to be X-ed', where X is the verb with which *able* is combined, as in *refutable* 'able to be refuted', *believable* 'able to be believed' etc.

Note importantly that the pairings of sound and meaning in the string $/ \Im r /$ and the meaning 'one who does X' or in the string $/ \Lambda n /$ and the meaning 'not' are just as arbitrary as the pairings in (2) and (7). And indeed, it is a part of a speaker's knowledge of English that they know these indivisible sound : meaning pairings as well as those in (2) and (7). So, we must include this information in our lexicon.

(13) More English lexical entries:

Label	"-er"	"un-"	"-able"
phonology:	/ər/	/ʌn/	/ʌbəl/
semantics:	'one who does	'not X'	'able to be
	X'	X = Adjective	X-ed'
	X = verb		X = verb

We have used lexical entries to represent the knowledge that a speaker of a language (in this case English) has about the arbitrary pairings of sound and meaning in their language. Thus far, all of the lexical entries that we have introduced correspond to the arbitrary pairings, i.e., the minimal units of phonology : semantics correspondence, or, put differently, the minimal meaningful units of grammar. Since not all lexical entries correspond to words, we need to give these minimal units a name, and we will call them *morphemes*, by analogy to phonemes in phonology.

(14) morpheme $=_{def}$ The minimal unit of phonology : meaning correspondence.

Armed with this term, we are now able to give somewhat more precise definitions of simple and complex words.

- (15) a. simple word $=_{def}$ A word which consists of a single morpheme
 - b. complex word $=_{def}$ A word which contains more than one morpheme

1.1.6 Complex words and lexical entries

Let us return briefly to complex words, such as *teacher* (or more precisely, the pairing of the phonology /tijtʃər/ with the meaning TEACHER). There is an intuitive sense in which we know the word *teacher*. Indeed, we could demonstrate this experimentally, with simple word-recognition and decision tasks. But does "knowing the word *teacher*" entail that our lexicon

contains a lexical entry for *teacher* that is separate for the lexical entries for its constituent pieces, namely *teach* (in (2)) and *er* (in (13))?

Despite the ease with which we can pose the question, this has proven to be a very thorny issue in morphological theory, and marks one of the most important current debates. To understand the question, it is important to think about what is at stake, and what the theory we are trying to construct is a theory of.

Let us think for a moment about syntax. Chances are as a speaker of English you have encountered some or all of the following sentences before.

- (16) a. John left.
 - b. I will never give up.
 - c. The cow jumped over the moon.

All of these sentences are fully compositional in terms of their meaning, so there is nothing particularly special about them from a linguistic viewpoint. Nevertheless, a native speaker of English may well recognize that they've heard these sentences before. Indeed, in the case of (16c), a speaker of English may well know about this sentence that it occurs in a nursery rhyme, and may have various memories and feelings associated with it. But what does this knowledge imply for the theory of linguistics? Is there anything special about the properties of the sentence (16c)? The answer is no—the nursery rhyme attributes a rather improbable feat to a particular bovine, but the meaning is entirely compositional (as a cursory glance at any illustrated collection of rhymes will show). There is a cow, and it is asserted that it leapt over the Earth's lunar satellite. It is true that we know the sentence (16c), and that there is a representation of this sentence as a whole that we know its meaning.

Put differently, among the properties an English speaker may know about the string *teacher* or the string in (16c) is the structural analysis of that string. In addition to whatever special associations there may be with the whole, the English speaker knows that the strings are composed of morphemes, each with its own lexical entry, and combined according to the rules of English (and UG) morphology. It is this aspect of linguistic knowledge which most interests us here, and which must therefore be reflected in the theoretical representations we will choose to use.

Note that this structural compositionality holds also of combinations such as the phrasal verb (or verb-particle combination) give up in (16b). Although the meaning of this combination is not a trivial result of combining the meanings that give and up have in isolation, it is nevertheless the case that the morpheme give is contained within give up, and this morpheme retains many of its characteristic morphological properties, for example, in taking the irregular past tense form gave(up). Our model of linguistic competence must accommodate this aspect of compositionality as well.

In order to make the properties of the theory as clear as possible, we will give lexical entries only to morphemes. That is, the lexicon, as a complete list of lexical entries, will be taken to constitute a list of all and only the morphemes in a given speaker's grammar. We can call this the

narrow lexicon, and the theory of morphology will be about the properties of morphemes and the rules and restrictions on their combination.

By calling this the narrow lexicon, we can delineate the object of study from another use of *lexicon* in the literature. This other usage is meant to include the list of all properties of words and phrases that must be learned, i.e., that are not a part of the theory of concatenation (not a part of morphology). This broad lexicon will include phrases that have non-transparent meanings (such as idioms), but also phrases such as those in (16a,c) where the compositional meaning is entirely transparent, and yet we might still want to attribute some knowledge of an expression to a speaker. The property of being included in the broad lexicon is formally referred to as **listedness**. The broad lexicon conceived of in this manner would also include favourite phrases, memorized pieces of songs and poetry, and indeed all linguistic objects about which one has knowledge. Expressions listed in the broad lexicon will nevertheless have a structural analysis in terms of the appropriate theories of syntax and morphology, and it is of course the latter that interest us here.

1.2 Words and morphemes in other languages

1.2.1 Wordhood

In the discussion above we have identified and defined *morphemes*, and have shown that morphemes may form the basis of words. A simple word is a word consisting of a single morpheme, while a complex word consists of more than one morpheme. Phrasing the discussion in this way appeals to our pre-theoretic intuition as speakers of English that we know what a "word" is. To be sure, there is fairly broad consensus in most cases. We all agree that *happy* is a word of English, but that *un*- is not. What properties of these strings do we appeal to when we consult these intuitions? For example, when we agree that *snowboard* is a complex word (a compound in fact), why do we say this, as opposed to saying that it is simply two words occurring very close together?

The same question can be asked of other languages as well. For example, speakers of Mohawk, an Iroquoian language, share the intuition that (17a) is a single word, whereas (17b) is a collection of words . What kind of linguistic knowledge underlies these shared intuitions?

(17) a. /wahuwaja?dawitsherahetk Λ :?d Λ ?/

'she made the thing one puts on one's body ugly for him'

'Mary likes the taste of that fish.'

The morphemes that occur in (17) are given in (18).

(18)	wa 'past'	j 'self's'	huwa 'she did s.th. to him'
	a?d 'body'	awi 'put on'	tsher 'thing'
	ahetka 'ugly'	?d 'make'	Λ 'for someone'
	? 'one time'		
	gígл 'this'	s 'regularly'	gńdzu 'fish'
	ne '?'	yakaw 'she'	uwá:ri 'Mary'
	éka 'likes the ta	ste of'	

The intuition that (17a) is a word, whereas (17b) is a string of words is based on phonological and syntactic properties of these strings. Taking phonology first, certain sound properties are best defined over a unit which generally (but not always) coincides with what we intuitively want to call words. This unit is called the **phonological word**.

In English, a phonological word must have at least one syllable. Thus /kæt/ is a word, and /bltk/ is a possible word, even if it is not an actual one. However, /k/ is not a possible word of English, neither is /z/ Note importantly that /z/ is a morpheme of English (for example, the plural in dog-s), and thus there is no minimal size for morphemes / lexical entries. Rather, the minimality requirement applies to phonological words.

In addition to minimality requirements, in some languages the phonological word is a maximal domain for some process. One example is stress, in many languages, words have a maximum of one stressed syllable. This is true of French, for example. Stress is also consistently placed in French at the last syllable of each phonological word. Thus, the fact that there are three stressed syllables in (19) (meaning: 'I buy a lot of fruit') tells us that this string contains at least three phonological words , and gives us clues as to where the word boundaries must lie (at a minimum, the "o" is part of the following, nit the preceding phonological word).

(19) [ʒa∫étbokúdfr^wí]

Other phonological requirements that signal the edges of phonological words include special **phonotactic** properties. Certain sounds occur only in particular positions within words, and certain combinations of sounds are excluded within words, but may be possible across word boundaries. For example, the sequence [ksstr] is impossible internal to a word in English, but occurs in the string in (20). The phonology thus dictates a word-boundary in the string.

(20) [hızsıksstren θ s] 'his six strengths'

Phonological conditions influence Mohawk speakers' intuitions regarding (17) as well. Some properties of phonological words in Mohawk, which apply to all content words (that is, to nouns and verbs, but not necessarily to function words such as pronouns) are:

- (21) a. must have at least two syllables
 - b. exactly one stressed syllable, generally the **penultimate** (=second to last)
 - c. The sequences /sn/ and /eu/ cannot occur inside a word (these are parts of more general rules of Mohawk phonology).

What do these phonological properties force us to conclude about the Mohawk examples? The string in (17a) contains exactly one stressed syllable, and that syllable is the penultimate one. Thus, (17a) is a single phonological word. The string in (17b) by contrast has four stressed syllables. Given their distribution, we must conclude that there are four content words, as indicated in (22). There is also a function word *ne*, which we identify as such on the basis of property (21c).

(22) gí:gA gÁ:dzu yakawé:kas ne uwá:ri

There are, however, some murky cases where our intuitions break down somewhat.

Consider, for example, the underlined string in (23). Phonologically, this is pronounced as [hæts], with voicing assimilation affecting the underlying voiced z of the auxiliary /IZ/. This suggests that from a phonological perspective, the string is indeed a word, but we're in real trouble if we try to build our theory of syntax on words of this sort—the two pieces *hat* and *s* belong to independent syntactic constituents (the former is a part of a prepositional phrase, modifying the subject, while the latter is the finite auxiliary).

(23) The man with the <u>hat's</u> my brother.

Luckily, we need not start our investigation in the grey areas. There is more than enough material of interest in the clear cases on which speakers agree that we can begin to develop our theory based on the clear cases. If it turns out that we do discover properties that distinguish word-internal composition from word-external composition, then we can use these properties to determine which way the questionable sequences fall.

Syntactic generalizations are also stated in terms of words (and bigger units called phrases). For example, words (and phrases) can be moved around for stylistic effect, while morphemes alone cannot:

- (24) a. I really don't like snowboarders.
 - b. Snowboarders, I really don't like.
 - c. *Snow, I really don't like boarders.
 - d. *Snowboard, I really don't like ers.
 - e. *Ers, I really don't like snowboard.

For the Mohawk examples, the phonological and syntactic criteria for word-hood coincide. Thus, the following re-orderings of the string in (17b) are possible, but no reordering of the string in (17a) will yield a grammatical word of Mohawk.

- (25) a. uwá:ri yakaw-é:ka-s ne gí:gA gÁ:dzu
 - b. yakaw-é:ka-s ne uwá:ri gí:gA gÁ:dzu
 - c. uwá:ri ne gí:gA yakaw-é:ka-s gÁ:dzu

In addition to movement and rearrangement, another syntactic criterion for word-hood is distributional. For example, in English certain verbs can take only a single Noun Phrase (NP) as a direct object (see (26a-b)). An attempt to put two nouns or noun phrases in the direct object position yields an unacceptable string (see (26c-d)).

- (26) a. I saw a ski.
 - b. I saw a toboggan.
 - c. *I saw [a ski toboggan]
 - d. *I saw [a ski] [a toboggan].

Contrast the examples in (26c-d) with compounds, where the occurrence of two noun roots is fully acceptable, precisely because the two together count as a single word for the puposes of syntax.

(27) I saw [$_{NP}$ a snowboard] / [$_{NP}$ a snow-shovel] / [$_{NP}$ a snow fence] etc...

Thus, we see that there are phonological and syntactic criteria for word-hood. When these criteria coincide, as in the Mohawk examples above, speakers have robust intuitions about words in their language. It is not always the case that the intuitions coincide so neatly. There are cases in which the phonological and syntactic criteria conflict, and in these cases, intuitions sometimes break down. Likewise, especially in the case of compounding, English orthography can often be very misleading. In some languages, like Mandarin Chinese, word boundaries are simply not indicated. All of these considerations define grey areas which should ultimately be brought within the scope of the investigation, but there is ample material to work with even if we restrict ourselves to the relatively clear cases where the phonological or syntactic criteria can give us an independent classification.

1.2.2 Segmentation

The starting point for morphological analysis is to segment a string of sounds into morphemes. The means for doing this is distributional, and there are two tools that we use in doing a distributional analysis, one of which we used implicitly in segmenting the English strings above, the other of which becomes more important in looking at data from an unfamiliar language. Consider the following words from different languages which may or may not be familiar (some phonological processes have been suppressed in (28b)).

(28) a.	Inuktitut:	tuktu	'caribou'
		tuktusiuqtuŋa	'I am seeking caribou' (e.g., hunting)
b.		qukiuti qukiutisiuqtuna	'gun' ι 'I am seeking guns'

Are these words simple or complex? Without knowing anything more about the languages in question, we may hypothesize that the second word in each pair is complex. The grounds for this hypothesis is that we see (or at least we think we do) the first word in each pair contained inside the second word.

It is important at this point to recall that when we say that we see the "word" *tuktu* inside the word *tuktusiuqtuŋa*, what we observe is not just the same string of sounds, but a correspondence between phonology and semantics, between sound and meaning. Thus, we can be fairly confident of this hypothesis, assuming that the data given are representative of Inuktitut generally.

The nature of what we have just done is the same as what we did in English, namely, we have identified a recurring correspondence between phonology and semantics, and we may therefore posit hypothetical morphemes, as in (29).

(29) Some tentative Inuktitut morphemes:



It is important to stress that the positing of morphemes in this manner constitutes a **hypothesis**; the best we can do given the data to hand. As theoretical linguists, we make hypotheses of this sort based on the data available, and then use these hypotheses to develop further questions, i.e., to identify further data that will be relevant to our analysis. And we must always be open to the possibility that our initial hypothesis may be in need of revision.

In particular, hypotheses made on the basis of a very limited array of data are far less reliable than those based on a massive study of a language. For example, it turns out that we are correct in isolating *tuktu* as a simple word and thus a morpheme in (29a), but the following Inuktitut words should lead us to revise our hypothesis concerning (29b).

(30) qukiuti 'gun' qukiqtara 'I shot it'

titirauti 'pencil' titiraqtara 'I wrote it'

Applying our segmentation here will yield the conclusion that *qukiuti* is itself morphologically complex and thus does not constitute a single morpheme. On the basis of this, we would delete the lexical entry in (29b) in favour of the two in (31).

(31) Revised Inuktitut morphemes:

Label	a."quki"	b. "uti"
phonology:	/quki/	/uti/
semantics:	SHOOT	INSTRUMENT
		FOR X-ING

Note that the data point quite clearly to the bi-morphemic nature of the words for *gun* and *pencil* in Inuktitut, even though these are mono-morphemic in English. It is important not to fall into the trap of assuming that there should be a 1:1 correspondence between morphemes in one language and morphemes in another. Quite often, this is not the case and what might be expressed as a single morpheme in one language may take a string of morphemes or words to translate in another.

It will turn out that (29c) is also internally complex (possibly as many as 4 morphemes!), and so when we have the tools to identify and isolate these, we will revise our analysis accordingly.

1.2.3 More segmenting

The following table gives you some more practice in segmenting, and in making hypotheses about morphemes. Though we will discuss the data immediately below, we suggest working through the data on your own first, then checking your hypotheses against the discussion that follows.

(32) Some Inuktitut words FONT: $[g = \gamma, r = \mu, j = d_3 / C_{;} = j / V_V]$

		· · · · · · · · · · · · · · · · · · ·			
iyluya	my house	iylumi	in a house	iylurd3uaq	big house
iyluŋa	her house	iylutut	like a house	umiaĸdʒua	qbig boat
uiya	my husband	uitut	like a husband	umialik	someone with
uiŋa	her husband	tupiqtut	like a tent		a boat
nunaya	my land	paniktut	like a daughter	uilik	someone with
qukiutiya	my gun	ayyaktut	like a hand		a husband
nunait	your land	nunakkut	across the land	umilik	someone with
		aputikkut	across the snow		a beard

Let's begin by looking at the four top, left cells. Is there a recurring string of sounds and a corresponding recurring meaning? Clearly yes, the string /iɣlu/ and the meaning HOUSE are common to all four forms. Can we make an even stronger claim looking through the remainder of the table? Yes, in fact we can say that the phonological string /iɣlu/ occurs in **all** the forms meaning HOUSE, and in **only** those forms. An 'all and only' pairing of this sort is the strongest basis on which to hypothesize a morpheme, and thus we hypothesize:

(33) An Inuktitut morpheme:



Note that the morpheme "igloo" in each of its occurrences in our table is followed by additional phonological material. When we collect together all the words that contain the morpheme

"igloo", and factor out that morpheme (i.e., both its sound and meaning) we are left with the mini-table in (34).

(34)

уа	my	mi	in a	 big
ŋа	her	tut	like a	

By factoring out the sound : meaning correspondence /i γ lu/ : HOUSE, we are left with a set of further sound : meaning correspondences, which are therefore good candidates for additional morphemes. The next step is then to examine the remainder of the data to see if these recur. Four of the five pairings do show up in other forms, and again, we are able to identify 'all and only' correspondences. Thus, all and only those words ending in the sequence / γ a/ contain the meaning 'my', all and only those words ending in the sequence / κ dʒuaq/ contain the meaning 'big', etc. We may therefore confidently hypothesize that these sequences are morphemes and posit lexical entries accordingly. Note that doing so allows us to factor out these pieces from other examples, and we are able to isolate, for example the pairing /nuna/ : 'land' from the word nuna γ a 'my land' minus the morpheme / γ a/ : 'my'. And so on throughout the table.

Now, one of the pairings in (34) occurs only once in the data set, specifically the pairing /mi/ : 'in a'. Are we justified on the basis of one example in positing a morpheme? Keeping in mind that our analysis is always open to revision on the basis of further data, within the data set thus presented we are justified in positing a morpheme here by the process of elimination. We know that the word iylumi means 'in a house' and we are fairly confident that the string /iylu/ contributes the meaning 'house' to this word. The residue of sound and the residue of meaning that coincide are thus /mi/ : 'in a', moreover, we have no indication that there is any problem with assuming this pairing to be a morpheme (even though there is only one data point, it does satisfy the 'all and only' criterion: there are no other word-final *mi* sequences, nor any other words containing the meaning 'in a'). Not only are we therefore justified in proposing this morpheme, it is the most reasonable hypothesis on the basis of the data available.

There are some other points which can be established on the basis of the data in this table. For example, the segmentation process outlined above should have led you to posit morphemes for which the meaning components are 'boat' and 'beard'. Let's focus on the second one. The meaning 'beard' occurs only in one form, namely *umilik* 'someone with a beard'. The string /lik/ though is identifiable as corresponding to the meaning 'someone with a', on the basis of its occurrence in *umialik* 'someone with a boat' and *uilik* 'someone with a husband' (for the latter, we can also independently confirm that /ui/ corresponds to the meaning 'husband'. Therefore, by deduction, we note the correlation of sound and meaning /umi/ : 'beard'. Now, does this phonological string occur in all and only the forms meaning 'beard'?

At this point the answer is no. While the phonological string /umi/ occurs in all the forms containing the meaning 'beard' (of which there is only one), the phonological string also occurs in all the forms containing the meaning 'boat', namely: *umialik* 'someone with a boat' and

umiasdzuaq 'someone with a boat'. This brings us back to the English examples *pit* and *spit* though. A recurring phonological string is not by itself of interest to morphologists. What is of interest to us is regular correspondences of sound and meaning. There is no apparent overlap in meaning between beards and boats, and thus, it is unlikely that this is any more significant than the fact that both words begin with the sound /b/ in English. Put differently, if the string /umia/ meaning 'boat' were supposed to contain the morpheme /umi/ = 'beard', what meaning could we assign to the piece /a/ such that when applied to beard it yields 'boat'? This is not to say that the correspondence is impossible, but it is highly implausible, and in the absence of compelling evidence to the contrary, this must be treated as a case of (partial) **homophony**, i.e., coincidental overlap in phonological representations among two morphemes.

Quick quizz:

Segment the following into morphemes. How many of them contain a morpheme with the phonological representation $/\Lambda n/?$

(35)	unfaithful	unlawful	untreated	untouched
	unthinkable	unhappy	unable	uncle
	untied	until	unsuspecting	unpretentious
	underwear	undershirt	unprepared	underprepared

Summing up, the list of morphemes that we should have posited on the basis of the Inuktitut data in (32) is given in (36).

(36) a.

Label	igloo	husband	land	gun
phon:	/iɣlu/	/ui/	/nuna/	/qukiuti/
sem:	HOUSE	HUSBAND	LAND	GUN
Label	boat	tent	daughter	hand
phon:	/umia/	/tupiq/	/panik/	/ayyak/
sem:	BOAT	TENT	DAUGHTER	HAND
Label	snow	beard		
phon:	/aputi/	/umi/		
sem:	SNOW	BEARD		
b.				
Label	my	her	your	in
phon:	/ya/	/ŋa/	/it/	/mi/
sem:	MY	HER	YOUR	IN A

Label	like	across	 big	s.o. with
phon:	/tut/	/kkut/	/rd3nad/	/lik/
sem:	LIKE A	ACROSS	BIG	SOMEONE
				WITH A

We have divided the list in (36) into two groups, for reasons that we will discuss in the next section. Some important things to notice about the morphemes we have posited are:

- (a) morpheme boundaries do not necessarily correspond to other linguistic divisions such as syllable boundaries (for example, the Inuktitut word for 'big house' which is divided syllabically as: /iɣluʁ.dʒuaq/ but morphologically as /iɣlu-ʁdʒuaq/).
- (b) a single morpheme in one language may correspond to (be translated by) more than one morpheme—or even more than one word—in another language (for example, the meaning associated with the Inuktitut phonological string /lik/ is expressed in English as the entire expression 'someone with a', consisting of four distinct morphemes; an example in the other direction is English *gun*, Inuktitut *quikuiti* shoot-tool, which we saw above.
- (c) we have been careful to refer to the lexical entries in (36) as morphemes and not as words; thus, you will see that we have talked above about the Inuktitut word *iyluya* 'my house' but that we have not referred to an Inuktitut "word" meaning 'house' but only to an Inuktitut morpheme meaning 'house'. Why is that? It is to this question that we turn in the next subsection.

1.2.4 Words and Roots

While we are confident on the basis of the data that we have examined so far that the Inuktitut phonological string /*iylu*/ means 'house', and while *house* is a word in English, we have not yet seen any evidence that the corresponding string can be a word in Inuktitut. In many languages, morphemes that correspond to words in English are unable to stand on their own. Consider the following examples from Ojibwa:

(37)	Ojibwa	(Algonquian;	Canada a	nd US)
------	--------	--------------	----------	--------

nidoon	'my mouth'	nigwis	'my son'	niʒi∫ibim	'my duck'
gidoon	'your mouth'	gigwis	'your son'		
*doon	(mouth)	*gwis	(son)	3i∫ip	'duck'

Based on the first two lines, we can easily identify morphemes corresponding to the English words *my*, *your*, *mouth and son*. Nevertheless, the morphemes meaning 'mouth' and 'son' cannot stand on their own as words. In Ojibwa (as in many other languages) these morphemes can only occur in complex words containing other morphemes, for example, possessors.

Morphemes with this behaviour are called **bound morphemes**. They may form the basis for a well-formed word, but they cannot stand on their own as words. Bound morphemes are quite common in the verbal systems of various languages. Indeed, using the bare stem as an infinitive is somewhat of a typological oddity of Modern English, especially within Indo-European. For example segmentation of the data in (38) allows us to easily isolate the Spanish morphemes corresponding to the English verbs *speak* and *live (habl and viv, respectively)*, but as (38c) indicates, these morphemes cannot stand on their own as verbs. They must occur in combination with other morphemes, and hence, we refer to these morphemes as **bound**.

(38) Bound roots are common in some languages. Spanish (rough paraphrases):

a.	habl-ar	'to speak' (infinitive)	viv-ir	'to live' (infinitive)
b.	habl-o habl-as habl-ábamos habl-aríamos habl-a	'I speak' 'you speak' 'we were speaking' 'If we spoke' 'Speak!'	viv-o viv-es viv-íamos viv-iríamos viv-e	'I live' 'you live' 'we were living' 'If we lived' 'Live!'
c.	*habl		*viv	

Note that adjectives and even nouns in languages like Spanish often consist of bound roots, and must be combined minimally with a vowel that indicates the gender (see Chapter D):

(39) a. Nouns

	perr-o	'dog' (Masc.)	perr-a	'dog' (Fem)
	perr-o-s	'dogs' (Masc)	perr-a-s	'dogs' (Fem)
	*perr	Not a word.		
b.	Adjectives	octives		
	roj-o	'red' (Masc, Sg)	roj-a	'red' (Fem, Sg)
	roj-o-s	'red' (Masc, Pl)	roj-a-s	'red' (Fem, Pl)
	*roj	Not a word.		

Our principles of segmentation allow us to identify a noun root *perr* meaning 'dog', and an adjective root *roj* meaning 'red'. These are morphemes of Spanish, and thus have lexical entries. But they are not words of Spanish, and thus constitute bound roots.

Extra! (40) Some Bound Roots in English: *scissor scirssors, scissor-kick, scissor-hands ... *capt capture, captive

Why are these not "traps" like *under* or Inuktitut *umi*, *umia*?

In these words, the prefixes and suffixes seem to have their normal meaning, e.g., un = "not" in *un-kempt* as contrasted with *well-kempt*.

nature. native

unenviable, undecided, unfair, unsaid

1.3 Sections still to come...

*nat

ROOTS AND AFFIXES

The first addition to lexical entires: "position"

ALLOMORPHY

Some allomorphy is phonology

But some allomorphy (Hungarian, English plurals in -en, < Nida) is predictable, but not derivable. This kind of allomorphy illustrates why it is important not to conceive of the morpheme as being a particular sound unit. The morpheme is a pairing of sound:meaning (including function), but the 'sound' part can be complex.

Analogy is to phoneme : allophone (partly). The abstract unit behaves in a coherent manner, as if it is a single entity for phonological purposes, although it may have distinct surface manifestations or **realizations**.

Note that we have already seen the same thing on the meaning side: the *-er* suffix can mean 'one who' or 'that which'. The difference in meaning is predictable from the context (whether it refers to a person or an object). In this case, there is a common component of meaning, with different, but predictable surface manifestations. (This isn't quite the same thing: this is underspecification, the appearance of allosemy arises due to the paraphrases, the fact that pronouns must specify animacy. The meaning is just 'x | verb (x)'. There are better cases later)

1.4 Asides and additional material

Aside A: One of the words in (4) is *antidisestablishmentarianism*. This is a curious word, and is often given as "the longest word in the English language" in trivia contests (medical terminology excluded). It may be the longest word in some particular dictionary, or perhaps in all printed dictionaries thus far, but we can demonstrate that it is not the longest possible word in English, and, perhaps more importantly, that any speaker of English who knows the word *antidisestablishmentarianism* already knows a longer word (even if they haven't consciously thought about it). Can you see it?

The point of departure is to consider the meaning of the word. We won't get into the details of xx-century English politics, but it is sufficient to know that antidisestablishmentarianism denotes a certain political platform / position / ideology held by a group of people (distinguishing themselves from the disestablishmentarianism position). Given that knowledge, would you, as an English speaker, know how you might refer to a person subscribing to this belief? Clearly, they would have to be an antidisestablishmentarianist., just as a capitalist subscribes to *capitalism* and a *socialist* to socialism. Now, surely more than one person held this position, and in order to talk about them together we would use the plural, which is: antidisestablishmentarianist s. word that is already а longer than antidisestablishmentarianism ! Note that this demonstration appeals only to your knowledge as a speaker of English about the pattern relating *-ism* words (for belief systems) to *-ist* words (for believers), and about the pattern of plural formation; you don't even have to know what it is the antidisestablishmentarianists believe.

Extra thought: can you make an even longer word from this one? Could someone be an *antiantidisestablishmentarianist*? If you're tempted to say "no" on the basis of the two *anti-*'s canceling each other out, think about the meaning of anti-anti-Semitic, or an anti-antiaircraft fire jacket (the official name for a flak jacket).

Later in book: add connection to neuro- representations, cf. Marantz, Pylkkänen on *teacher*, anticonnectionist position. Do not present this as "experimental verification of theory", present instead as "corroborating evidence from an independent domain".

1.5 Sources

Definitions of *dog*, *talk* excerpted from <u>www.dictionary.com</u>

blatch in (9c): [to] fly [so as] to intentionally collide with another player, from the Harry Potter Lexicon www.i2k.com/~svderark/lexicon