

1 Introduction¹

- Morphology deals with
 - word form (sound shapes)
 - word structure (abstracting from the shape)
 - word generation (building complex structures)
- It aims to define notions “morpheme”, “word”, “inflection”
- But why a cognitive scientist should care?
- Because human beings seem to have intuitions (read as implicit knowledge) about these concepts.
- How do we know that they do?
- It is revealed in behavior and learning.

2 Morphemes, morphs, allomorphy

- What comes to your mind when you hear “word structure”, or just “structure”?
- Let’s start with something basic and intuitive: part-whole structure of a word.
- **Morpheme**: smallest linguistic unit with a grammatical function \simeq smallest meaningful part. For the notion of morpheme, we take meaning to be equivalent to function. By “smallest”, we mean a unit that is not decomposable further.
- Another definition would be “sound-meaning pairing”. In some cases this proves problematic, we will come to them below.

¹The material up to section 7 is largely adapted from Payne (2006).

Exercise 2.1 (Aztec)

List the morphemes:

a.	ikalwewe	‘his big house’	i.	petatc-in	‘little mat’
b.	ikalsosol	‘his old house’	j.	ikalmeh	‘his houses’
c.	ikalc-in	‘his little house’	k.	komitmeh	‘cooking-pots’
d.	komitwewe	‘big cooking-pot’	l.	petatmeh	‘mats’
e.	komitsosol	‘old cooking-pot’	m.	ko-yamec-in	‘little pig’
f.	komitc-in	‘little cooking-pot’	n.	ko-yamewewe	‘big male pig’
g.	petahwewe	‘big mat’	o.	ko-yameilama	‘big female pig’
h.	petatsosol	‘old mat’	p.	ko-yamemeh	‘pigs’

- Morpheme classification:

Free morphemes can stand alone, while **bound** morphemes always need some other linguistic material to realize a complete(?) expression.

Another typing:

- Root (bound or free)
- Affix (bound)
- Clitic (bound)

- Root is the semantically main part of a word, which cannot be further divided into morphemes. Examples?
- Sometimes linguists talk of **stems** as well. The stem root distinction is subtle and sometimes confusing.
- Affixes are added to a root or stem to form new words.
- Clitics are bound morphemes that differ from affixes by being related to a structure larger than those at word-level. E.g. Turkish *de/da* are clitics.

2.1 Some complications with the notion “morpheme”

- Non-segmental morphemes: the morpheme does not correspond to a designated sequence of sounds (segment), but is more abstract.

English past tense:

run	ran
speak	spoke
eat	ate
read	read

Noun to verb morphology:

Noun	Verb
breath	breathe
cloth	clothe
house	house

where the voiceless fricative endings ([θ,s]) turn to voiced fricatives ([ð,z])

- Zero morphemes:

Turkish:

gidiyor-um	gidiyor-∅
going-1sg	going-3sg

English:

a fish
ten fish-∅

- Two lessons:

Some morphological operations need to be thought as processes.

At an abstract level a morpheme defines a contrast.

Exercise 2.2 (German plural)

What indicates plurality in German?

Väter	'fathers'	Auge	'eye'
Kinder	'children'	Adler	'eagle'
Pferd	'horse'	Kind	'child'
Männer	'men'	Augen	'eyes'
Vater	'father'	Kuh	'cow'
Mann	'man'	Frauen	'women'
Adler	'eagles'	Auto	'car'
Kühe	'cows'	Autos	'cars'
Pferde	'horses'	Frau	'woman'

- **Morph:** individual **tokens** of a morpheme.
- **Allomorphs:** The set of morphs of a morpheme. Two morphs from this set stand in an allomorphy relation.
- Example: English plural allomorphs?
- Note that in deciding on morphemehood, function (meaning) prevails over form. Compare *quickly* vs. *lovely*.

Exercise 2.3

Think of cases from Turkish where we need the abstraction of allomorphy.

3 Inflection vs. derivation

- Inflections express grammatical or functional categories – this will make more sense when we come to syntax.

therefore they are general. E.g. In a given language if a verb has tense every verb has tense; if a noun has case, every noun has case.

Inflections are relevant for syntax.

- Derivations form new words.

they are more restricted: *build-ing*, **see-ing*, *view-ing*...

4 Main morphological processes

Affixation:

- Prefix: *anti-dis-establishment*
- Suffix: *antidisestablish-ment-ari-an-ism*
- Infix:

Tagalog *-um-* makes an agent out of a verb:

sulat s-um-ulat ‘one who wrote’
gradwet gr-um-adwet ‘one who graduated’

- Circumfix:

Indonesian *ke-...-an* makes nouns from adjectives:

besar ‘big’ ke-besar-an ‘bigness, greatness’

Chukchee *a-...-ke* expresses negation:

jatjol ‘fox’ a-jatjol-ka ‘without a fox’
cakett ‘sister’ a-cakettə-ke ‘without a sister’

Stem modification:

Shape change without an affix. Ex.: *ring, rang*.

Autosegmental variation:

What is “autosegmental”? Processes that refer to structures beyond single sounds are autosegmental processes or properties. For instance where the main stress (Tr. *vurgu*) refers to syllables (and their organization) rather than individual sounds; there is no rule like “stress the vowels *a* and *e*, but not *i*, *o*, and so on”.

English stress:

convért → *cónvert*

permít → *pérmit*

Reduplication:

Indonesian plural is made by duplicating the root: *anak* ‘child’, *anakanak* ‘children’.

More interestingly, Ilakona, an Austronesian language, duplicates the first syllable only for plural: *ulo* ‘head’, *ululo* ‘heads’.

Non-concatenative morphology:

In Semitic languages like Arabic and Hebrew.

kṭb	Root	no meaning
kəṭob	imperative	‘write!’
katob	infinitive	‘to write’
koṭēb	present participle	‘writing’
katub	past participle	‘written’
katab	perfective	‘wrote’

Subtractive morphology:

Murle (East Africa) plurals are made by removing the final consonant: *nyoon* ‘lamb’, *nyoo* ‘lambs’, *onyiit* ‘rib’, *onyii* ‘ribs’.

Don’t confuse with languages that mark the singular rather than the plural:

Arbore (Ethiopia) is such a language: *tiisin* ‘a maize cob’, *tiise* ‘maize cobs’, *nebelin* ‘an ostrich’, *nebel* ‘ostriches’.

Compounding:

black + board = blackboard

5 Morphotactics (morphology-syntax/semantics² interaction)

- Morphotactics is the part of the grammar that regulates the order of morphemes. For instance in Turkish, the order causative > passive > aspect > person holds, e.g. *kandırılmışız*.
- Morphotactics involves notions from both syntax and semantics. For instance, the reason why passive cannot “apply” before causative has a syntactico-semantic explanation.

²It will become clear in the coming weeks why we write “syntax/semantics”.

6 Rules and representations in morphology

- Let's take the plural marking in English (N = noun, pl = plural, sg = singular):

$$(1) \quad N_{sg} + -s = N_{pl}$$

- We can represent the process in an input/output format:

$$(2) \quad \text{Name of the process: } input \rightarrow output$$

$$(3) \quad \text{Plural: } N \rightarrow N + -s$$

- Now some more abstraction. Take the following data from Arabic.

(4)	Root:	slm	Root:	ktb
	a.	muslim 'person of peace'	c.	muktib 'literate person'
	b.	salima 'he was safe'	d.	katiba 'he was reading'

On the surface the rule appears as:

- (5) a. Person nominalization: $slm \rightarrow muslim$
- b. Person nominalization: $ktb \rightarrow muktib$

It is not hard to see that there is room for being more general:

$$(6) \quad \text{Person nominalization: } C_1C_2C_3 \rightarrow muC_1C_2iC_3$$

- Observe that we have a type of rule which is able to take apart the consonants in an Arabic root and store them in variables. Once this is done, with similar rules you can perform any operation of shuffling, insertion and/or deletion on the input. Also observe that a rule capable of breaking into parts a string of an arbitrary length (not just 3) would need to be more “intelligent” than what we have here. We will see such rules in coming weeks.
- By the way, can you see an important shortcoming of these rules?

- Some crucial information that might be needed by another rule that would act on the result of the above rules is missing.
- Usually morphology is not a business done only for its own sake. Many properties that are decided by morphological processes are used by the rules of syntax; just like morphology sometimes uses outputs of phonological processes. Therefore in a more thoroughly conceived version of (6), we would have the information that the output is a noun (since the rule is nominalization), so that the upcoming syntactic processes can make a proper use of it.

- One immediate problem with this rule is its behavior in cases like *şartı* (as in *üyelik şartı*). Our rule would erroneously convert *şartı* to **şardı*.⁴ This can be avoided by a simple modification: we just make our rule aware of morpheme boundaries so that it does not operate on the *t* in *şartı*, which is now *şart-ı*, and thereby does not match the input specification of our new rule in (11).

$$(11) \quad -t \rightarrow -d \quad / \quad V \text{ or } C - \{p, \zeta, t, k, s, \zeta, f, h\} \text{---} H$$

- However, another problem arises in cases like *karar-tı*, *ürper-ti*, and the like. Our rule would erroneously turn these to *karar-dı* and *ürper-di*, which are, though well-formed, entirely different than what we intend. To remedy this, we can make our rule aware of the distinction between derivational vs. inflectional morpheme boundaries. If we agree that ‘+’ signifies derivational, and ‘-’ signifies inflectional boundaries, then our rule would not make the changes above, as the actual representations of these are *karar+tı* and *ürper+ti*.
- It is crucial to observe the chain of reasoning which lead us from the decision to take *t* as the default, to the conclusion that our rules need to be aware of the derivation vs. inflection distinction. Now let us **backtrack** to an earlier decision point, retract our “*t* is default” assumption and instead take *d* as the default. Now the rule becomes:

$$(12) \quad -d \rightarrow -t \quad / \quad \{p, \zeta, t, k, s, \zeta, f, h\} \text{---} H$$

- To the best of our knowledge, (12) does not run into the problem we had with default *t* in the face of expressions like *karartı*. In other words, our conjecture is that there will be no cases where a *d* right after a morpheme boundary would erroneously be turned to *t*. Of course the rule would need to be revised upon encountering a possible counterexample.
- The above observation, if correct, also frees our hands of the derivational vs.

⁴To be precise, here we assume that *H* has not been turned to *ı* yet at the time our rule is applied. Therefore what happens is: first our rule turns *şartH* to *şardH*, and the vowel harmony rule, which we leave unspecified here, turns this into *şardı*. This is to say that vowel harmony rules apply after the rule we are trying to specify.

inflectional morpheme distinction. The ‘-’ can now mean “any morpheme boundary”. Upon closer inspection perhaps we may even drop the morpheme boundary sign, but we will not pursue this investigation here.

- There is still room for further simplification, this time of a different character. The simplification involves finding a **natural class** for the set of consonants in the environment specification of the rule, something that serves a common denominator for the consonants in the list, by the use of which we can represent all of them at once. For that, we need to digress a little.

Very basic phonetics/phonology:

- Think of a consonant, say *t* or *b*, as being specified by three **features**: (i) where (**place**) in the vocal tract it is articulated (alveolar, dental, glottal, and so on), (ii) how (**manner**) it is articulated (fricative, plosive, nasal, and so on); and (iii) whether the vocal chords are involved or not (**voice**). A particular consonant can be represented as a set of features.
- For example, *t* and *d* are both “alveolar plosives”, articulated by restricting (hence plosive) the air flow by tongue and alveolar ridge (hence alveolar). The former is voiceless, while the latter is voiced. More formally, *t* is {[place alveolar],[manner plosive],[voice -]}, while *d* is {[place alveolar],[manner plosive],[voice +]}. We adopt an abbreviating convention, where we omit the words “place” and “manner” and write ‘[+voice]’ instead of ‘[voice +]’, etc.
- Returning to our track, the set of consonants that call for a *d* to *t* conversion in (12) is the class of voiceless consonants (of Turkish)([-voice]). If we designate this set as $C_{[-voice]}$, then the rule becomes:

$$(13) \quad -d \rightarrow -t \quad / \quad C_{[-voice]} \text{---} H$$

- To further generalize our rule, observe the use of suffixes for locative (-*de*, -*da*, -*te*, -*ta*), adjective forming (-*ken*, -*kan*, -*gen*, -*gan*), occupational (-*cH*, -*çH*). Again the consonants are alternating between voiced and voiceless variants of the same type. Therefore, why not have rules like the following, entirely leaving the consideration of ensuing vowels out:

