1 Introduction[1]

- Morphology basically 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.
- Another definition would be "sound-meaning pairing". In some cases this proves problematic, we will come to them below.

  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.	koyamec·in	'little pig'
f.	kommitc·in	'little cooking-pot'	n.	koyamewewe	'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 linnguistic 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. English *the*, *a*, Turkish *de/da* are clitics.

2.1 Some complications with the notion "morpheme"

- Non-segmental morphemes:

  English past tense:

[1] The material up to section 5.1 is largely adapted from Payne (2006); the rest is developed by me and the Fall 2015 class.

run	ran
speak	spoke
eat	ate
read	read

Noun to verb morphology:

Noun	Verb
breath	breathe
cloth	clothe
house	house

where the voicless 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 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"?

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'.

Yet more interestingly... see exercise (5.1) below.

Non-concatenative morphology:

In Semitic languages like Arabic and Hebrew.

ktb	Root	no meaning
kətob	imperative	'write!'
kɑtob	infinitive	'to write'
kotɛb	present participle	'writing'
kɑtub	past participle	'written'
kɑtab	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

4 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*...

5 Rules and representations in morphology

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

  (1) N_{sg} + -*s* = N_{pl}

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

  (2) *Name of the process*: *input* → *output*

  (3) Plural: N → 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 → muslim
b. Person nominalization: ktb → muktib

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

(6) 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 information that might be needed by another rule that would act on the result of the above rules.

- 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 nominal (since the rule is nominalization), so that the following syntactic processes can make a proper use of it.

5.1 Morphophonemics (morphology-phonology interaction)

- As we saw while discussing the notion of allomorphy the particular form (the morph) of a morpheme may depend on the particular **environment** of the morph. Rules governing such processes fall within **morphophonemics** (or **morphophonology**.

- For a process to fall under morphophonemics it must be necessary to refer to morphological units like morpheme, stem, affix, and so on, when specifying the process. Otherwise processes involving only sound (or form) change without reference to morphology falls in the realm of **phonology**. Examples?

- Let's start with Turkish perfective. The allomorphs are:
$\{$*-dı,-di,-du,-dü,-tı,-ti,-tu,-tü*$\}$.

  or
  $\{\text{-}\alpha\beta \mid \alpha \in \{d,t\}, \beta \in \{ı,i,u,ü\}\}$

- In what follows, we will concentrate on consonants and leave aside vowels and vowel harmony. So let H stand for a vowel which is picked from $\{ı,i,u,ü\}$ according to vowel harmony. Also C stands for all the consonants, and V, for vowels. By convention, these set names can stand both for the sets and an element picked from the sets, determined by the context.

  Given these convention, the allomorphs of Turkish perfective are {*-dH, -tH*}. Now what is the rule that governs whether we have *t* or *d*?

- We are faced with the tasks of specifying:

  A. which consonant appears in which environments.

  B. what will be the input to the morphophonemic rule(s).

- Now, let's make an assumption which has the potential of simplifying things. Let's assume that the following rule attaches an abstract morpheme of perfective to a verbal stem. If we can maintain this, our sound changing rules do not need to "know" about the details of the morphemes. Abstraction for the consonants can be accomplished by defining $D = \{t,d\}$, like we did for vowels – but this time we will specify what D will become ourselves.[2]

[2]Here is an idealization. For the sake of simplicity we will treat expressions like *tuhaftı* comprising of a verb ending with the sound 'f' and the perfective marker. A more thorough analysis is to have the adjective *tuhaf* followed by the copula *i* and the perfective marker. Turkish morphophonemics gives the final form *tuhaftı* to the composition of all these.

(7) Perfective: Verb → Verb + *DH*

- Now we can devise a rule that further specifies *D* as *t* or *d* according to the specific environment. Here is a tentative rule for that – environments are shown by giving a **context** after the '/'.

(8) a. $DH \rightarrow tH \;/\; \{p,ç,t,k,s,ş,f,h\}\_\_$
b. $DH \rightarrow dH \;/\; V \text{ or } C- \{p,ç,t,k,s,ş,f,h\}\_\_$

- We can take the *H* repeated in the input specification to the environment specification:

(9) a. $D \rightarrow t \;/\; \{p,ç,t,k,s,ş,f,h\}\_\_H$
b. $D \rightarrow d \;/\; V \text{ or } C- \{p,ç,t,k,s,ş,f,h\}\_\_H$

- We can simplify this rule by assuming that one of *d* and *t* is the default case, and it is converted to the other in certain environments. What should be the basis of which form to take as the basic? One way to tackle such issues, which is quite common in science, is to make an assumption and look at the consequences of this assumption. Let's take *t* as default and see what happens. Our rule becomes:

(10) $t \rightarrow d \;/\; V \text{ or } C- \{p,ç,t,k,s,ş,f,h\}\_\_H$

- This rule changes all the *t*'s to *d*'s if they are found to be preceded by a vowel or a consonant from the set specified in the environment part and followed by *H*, and leaves untouched other *t*'s. Observe that the effect is the same with that of rule (9), but now we have a simpler rule.

- 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*.[3] 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) $\text{-}t \rightarrow \text{-}d \;/\; V \text{ or } C- \{p,ç,t,k,s,ş,f,h\}\_\_H$

- 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, than 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) $\text{-}d \rightarrow \text{-}t \;/\; \{p,ç,t,k,s,ş,f,h\}\_\_H$

- 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. 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

[3] To 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 tyring to specify.

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 a 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 $C_{[\text{-voice}]}$, then the rule become:

(13) $\textit{-d} \rightarrow \textit{-t} \;/\; C_{[\textit{-voice}]}\_\_H$

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

(14) $\textit{-d} \rightarrow \textit{-t} \;/\; C_{[\textit{-voice}]}\_\_$

(15) $\textit{-g} \rightarrow \textit{-k} \;/\; C_{[\textit{-voice}]}\_\_$

(16) $\textit{-c} \rightarrow \textit{-ç} \;/\; C_{[\textit{-voice}]}\_\_$

- Would it be possible to collapse these rules into one? Actually same thing is going on in all of them. What is that?

- First, let's agree that all the vowels are [+voice] by definition.

  Second, let's agree to use variables that can stand for the values of features, the minus or plus signs for voice, glottal, dental, etc. for place, and so on.

  Third, let's use '?' as the value of a feature, whose value we do not care – whatever comes in its place is fine for us.

  Forth, let's define a *meta* feature '[cons. +]' or '[cons. -]', which is possessed by all consonants.[4]

  Now, all this in place, we can state:

(17) $\text{-}\{[\text{+cons.}],[\text{?voice}]\} \cup A \rightarrow \text{-}\{[\text{+cons.}],[\alpha\text{voice}]\} \cup A \;/\; Z \cup \{[\alpha\text{voice}]\}\_\_$

- Let's look at what's going on in (17) in detail. First our underlying form, which is the input to our rule, leaves the voice feature of the consonant **underspecified**. This is to say that whether the consonant will be voiced or not is left to (17) to decide. Our rule says that the voice feature of the initial consonant of a morpheme has to be same with the voice feature of the consonant preceding the morpheme boundary.

- This move brings a huge simplification to our morphophonemics, but there is a complication, namely, the problem with *karartı*, *ürperti* came back. What is more, this time we cannot solve the problem by counting on the derivation vs. inflection distinction, since our generalization about voice was meant to apply to both cases. At this point one need to decide whether to keep the generalization in (17) or not. If one decides to keep, then a way of covering the "exceptional" case of *karartı* needs to be found. Maybe there is "something" in that case that makes the morphophonemics behave differently. The task would be to find that "something" and revise the rules so that they become

[4] One way to define is to introduce a conditional rule: If a feature set F has at least one feature which can only be possessed by consonants, add '[cons. +]' to F, otherwise add '[cons. -]'.

"aware" of the necessary distinction. I leave the further development of the account as an exercise.

- So far we have seen rules that modify a sound or an underlying form. One can also devise rules that insert and/or delete – even shuffle – sounds. It is also possible to write rules that are sensitive to syllable boundaries or any other information. Just don't forget that any extra information you introduce needs to be justified on the grounds that it is required for a correct and/or simple account of the phenomenon.
- When you try other examples, you will realize that in most cases rules should be ordered. For a very simple instance, rules that delete the '+'s and '-'s in the input should come last.
- Finally, there is still room for improvement in the rule schema we have used. We will come back to it after we have a more formal look at the subject in the next weeks.

Exercise 5.1.

i. In our first lecture we observed that when the accusative case marker is suffixed to a possesive compound (e.g. *trafik cezası*), the sound *n* needs to be inserted in between. Give an account of this process. One crucial question is whether the *n* is part of the accusative marker or the possessive suffix.

ii. Give an account of emphatic reduplication in Turkish, e.g. *sarı→sapsarı*, *yeşil→yemyeşil*, and so on.

5.2 Morphotactics (morphology-syntax/semantics[5] 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.
- More on morphotactics after some syntax...

[5] It will become clear in the coming weeks why we write "syntax/semantics".