# Morphology

## 1    Why morphology matters?

*Innate bias for words*

- If it is true that words are not marked in the physical stream of speech, two questions:

    - Why did our brains in childhood (infancy) go to the trouble of working out how to parse the speech stream into words?

    - Why do our brains now go to the trouble of segmenting the speech stream into words?

    - Why is the nature of the analysis so nearly identical across languages?

- To question 1:  One can give a functional explanation for the existence of words.

    - It is advantageous to construct a flexible communication system out of a smallish set of units, combined and recombined.

        Keeps memory use under control.
        Allows novel utterances.

    - But does this tell us how or why the baby acquires the words?  Does she do a little communication theory and decide that speech must contain words?  Probably not.

    - Answer is an **innate bias** to perform some sort of analysis of speech (explaining 1 and 2), and probably **innate knowledge** of what the product of this analysis will look like (explaining 3).

*Arbitrariness*

- But remember:
        The words themselves are not innately specified, or else they would not differ from speech community to speech community (***arbitrariness)***.

- Arbitratriness pervades language...
        English *dog*, French *chien*, Russian *sobaka*...

- ...even in "onomatopoeia" — how we represent animal sounds linguistically is fixed:[1]

    *cow*:  Dutch *boeh*, Hungarian *bú*,  Thai *maw maw* (with mid tone), Bengali *hamba*
    *dog*:  Russian *gav-gav*, Catalan *bup-bup*, Indonesian *gong-gong*

---

[1] http://en.wikipedia.org/wiki/Cross-linguistic_onomatopoeias#Animal_sounds

**What is Innate (in this domain):**
        Drive to analyze speech into words.

**What is Learned (in this domain):**
        What the words actually are.

---

*Morpheme:*   the minimal subpart of a word that expresses a meaning (or grammatical property)

**Example:**  *dogs* contains two morphemes:

    *dogs*    = *dog* 'furry beast' + *s* 'plura
    *dogs*    ≠ *d* 'furry' + *og* 'beast' + *s* 'plural'

---

## 2    How we learn words and morphemes

**Rapidly!**

- First words appear around age 1.  (**One-word stage.**)

Kids already know more than they say:
- In 11-month old, selective looking reveals preference for pauses that coincide with word boundaries over pauses inserted between syllables of words.
- In 9-month old, no preference.

So what happened between 9 and 11?
        How do you segment speech into words if you don't know the words?
        No magic bullet, it seems!  Part of the story must be a statistical analysis of input data carried out by the child — and there are explicit proposals about this.

        *But a deeper question remains:  Why* does the baby perform this statistical analysis performed, and *why* are the results saved in memory (ultimately, in the lexicon)?

 **An approach to an answer:**
        An *instinct* for language acquistion.
        **Innate knowledge coupled with environment-dependent learning.**

---

## 3    The knowledge underlying use of a complex word

- *Productivity* of morphology:
(1) industry, industrial, industrialize, industrialization., industrializational, industrializationalize, industrializationalization, ...

- There is no termination to the process, except that the words become too long and their meanings too complex to keep track of all the morphemes.

- An interesting difference between words and sentences: We have no problem parsing sentences with as many words as *industrializationalization* has morphemes (7 of them!) -- compare even *the process of turning a question about something else into a question about industrialization*, which has the same meaning but 14 words! The multi-word phrase is much easier to understand than the multi-morpheme word. We are not as good at parsing words as we are at parsing sentences.

---

- **Distinctions: bound morpheme vs. free morpheme / open-class vs. closed-class**

*Industry* is a word on its own.        **free morpheme**

*-ize, -al*, and *-ation* are not.        **bound morpheme**

- The free morpheme *industry* belongs to an **open class** of **lexical items**, an extendible class of items with a wide range of meanings. You can make up new ones, and are not surprised to learn new ones even in adulthood: *screeve, Obama, cromulent*.

- The opposite of "open class" is "**closed class** (also known as "function words").

---

Since open class items can be coined at will (more or less), it is really the closed-class items that identify the language — that make English English, make French French etc. Lewis Carroll's poem Jabberwocky contains lots of unfamiliar ("nonsense") open-class items, but all its closed-class items — all its function words — are English. Crazy though the poem may sound, it is clearly a poem written in English!

*'Twas brillig, and the slithy toves*
*Did gyre and gimble in the wabe;*
*All mimsy were the borogoves,*
*And the mome raths outgrabe.*

Compare a version in which the open-class words are replaced with normal English, but the function words are replaced with "non-sense". It no longer looks like English:

*Glorm cloudy, yurk lim noisy frogs*
*Rur croak yurk fidget uf lim lake;*
*Caj nervous trid lim hunting dogs,*
*Yurk lim hungry cats lay awake.*

Or worse, replace both function words and open-class items with made-up words:

*Glorm brillig, yurk lim slithnit tovem*
*Rur gyre yurk gimble uf lim wabe;*
*Caj mimsnit trid lim borogovem,*
*Yurk lim mome rathem nusgrabe.*

Because the real poem truly is written in English (because its closed-class items are English), the notion "translating Jabberwocky" into a foreign language is a meaningful one, and lots of translations have been made.

---

- **free vs. bound ≠ open vs. closed class**

- **For example:**
    There are **free** closed-class morphemes (prepositions, pronouns, articles...)
    There are **bound** open-class morphemes (*un+kempt, sanit+ize*)

---

## 4   How to build a morphologically complex word:  the rule Merge

You have to start somewhere!

*Step 0:*

- Take a morpheme from the lexicon whose lexical entry does not say "I am a suffix" or "I am a prefix"... This starting point, which (as the textbook says) carries "the main meaning of the sentence", is the called **root**.

- Putting aside a special kind of complex word called a compound (to which we return next week) — if there's a free morpheme in the word, it's the root.

*Step 1:*

- A morphological rule that we can call "**Merge**" takes an *appropriate* morpheme from the lexicon and glues it to the root. (We return soon to what we mean by "appropriate".)
- A non-root morpheme in a word is called an **affix.** So morphologically complex words are made up of a root and some number of affixes.

**The *lexical entry* for each morpheme specifies**

> 1. its sound
> 2. its meaning
> 3. whether it is a root or affix
> 4. its part-of-speech (category): N, V, A...
> 5. if it is an affix, the part of speech it merges with:
> 6. if it is an affix, its place of attachment:
> > a. left edge (prefix)
> > b. right edge (suffix)
> > [c. after the first syllable, before the last syllable, before the main stress...
> > (infix)][2]

*Step 2:*
- The output of step 2 can undergo Merge again.  Another affix can be attached to it.
- The element to which an affix attaches (whether a root or a morphologically complex word) is called its **base**.
- When an affix attaches to the root, the root is also its base.  But when an affix attaches to a an already complex form, its base is not the root.

## 5   Trees

- A tree diagram can illustrate the internal structure of a word, and reflects how it was formed by one or more operations of Merge.

  [see the textbook for lots of actual trees]

- **Words really are trees (structures) -- not just strings of morphemes**

For example ...
  the suffix *-ment* attaches to a verb and produces a noun:

| | |
|---|---|
| govern | government |
| adjourn | adjournment |
| treat | treatment |
| amuse | amusement |

body  —X→ *bodiment
trance  —X→ *trancement
power  —X→ *powerment

- You might say: "the morpheme next to *-ment* must be a verb" as an alternative to "the base to which  *-ment* attaches must be a verb".

---

[2] Discussed below.

- An argument against this: find a *prefix* that makes a noun (or adjective) into a verb!

  The prefix *en-* (sometimes *em-*) forms verbs out of nouns (and adjectives):

body --> embody
trance --> entrance
power --> empower

Now notice  that these verbs can take the suffix *-ment*.  Even though the morpheme next to *-ment* is not a verb morpheme, structurally *-ment* is being added to a verb.

---

## 6   "Sound" in more detail

- Whether a morpheme even *has* a sound is arbitrary.

  In English, the fact that a noun is singular is not represented by an overt morpheme (one can say there is a **zero morpheme**) — though plural is a morpheme with a sound of its own (*dog* vs. *dog-s*).

- Some morphemes that do have a sound do not have a sound of their own.  Their sound is a copy of sounds already present in the base.  This is called **reduplication**:

**Example 1: Lakhota (Siouan, S. Dakota)**

| sg | inanimate pl | |
|---|---|---|
| gí | gigí | 'to be rusty brown' |
| ská | skaská | 'to be white' |
| ʃǎ | ʃaʃǎ | 'to be red' |
| thó | thothó | 'to be blue or green' |
| ʒí | ʒiʒí | 'to be yellow' |

Sound:   Copy first CV of the base.

**Example 2: Ilocano (Austronesian, Phillippines)**

| sg | plural | |
|---|---|---|
| píŋgan | piŋpíŋgan | 'dish' |
| tálon | taltálon | 'field' |
| dálon | daldálon | 'road' |
| bíag | bibíag | 'lives' |
| nuáŋ | nunuáŋ | 'carabao' |
| úlo | ulúlo | 'head' |
| múla | mumúla | 'plant' |
| táwa | tawtáwa | 'window' |

- In the lexical entry for plural, under the heading "sound", it says not -*ness*, or *un-*, but:

  **"Copy first (C)V(C) of the base."**
  [C = consonant, V = vowel, parentheses = "if present...")

- Reduplication is extremely common in languages of the Austronesian group (Phillipines, Malaysia, Indonesia)

---

## 7   Allomorphs

- Sometimes the sound of a morpheme depends on the base to which it is attached, but is not a copy of that base. Such a morpheme will have different **arbitrary** (i.e. not copied) sounds depending on the base to which it attaches. These variants are called **allomorphs**.

**Example 1:** English plural depends on the identity of the base to which it attaches.

  dog-**s**, cat-**s**, horse-**s**

  but...

  ox-**en**
  child-**ren**
  sheep-**ø,** deer-**ø**

  Note that the -s plural is **productive**. It is the **default** plural marker which we apply in the absence of special knowledge about an irregular plural marker.

  What is the plural of *wug*?

  The situation is more complicated still: the default affix is **-z**, which regular phonology changes to /s/ for reasons we will discuss later. But after -s, -z, -ch, -sh, -j we get a vowel as well. This is more allomorphy...

  What is the plural of *gruss*?

  Sometimes the plural involves a change of vowel: ***mouse/mice, woman/women***.

**Example 2: English past tense:**

  Default is **-d**, but here too we have vowel changes, and sometimes both:

  walk/walked, industrialize/industrialized
  run/ran, sing/sang, lead/led, strive/strove
  sleep/slept

**Example 3: Latin forms of the verb meaning 'love'**

|           | *present* | *past*      | *future*    |
|-----------|-----------|-------------|-------------|
| *you (sg)...* | ama-s     | ama-ba-s    | ama-bi-s    |
| *she/he/it...* | ama-t    | ama-ba-t    | ama-bi-t    |
| *we...*   | ama-mus   | ama-ba-mus  | ama-bi-mus  |
| *you (pl.)* | ama-tis  | ama-ba-tis  | ama-bi-tis  |
| *but:*    |           |             |             |
| *I ...*   | am-(a)-o  | ama-ba-**m** | ama-b(i)-o  |
| *they ...* | ama-nt   | amaba-nt    | ama-**bu**-nt |

The underlined bits are unexpected, and show allomorphs of the suffixes visible elsewhere in the chart.

---

## 8   "Place" in more detail

<u>Point 1:</u> The place of attachment must be listed in the lexicon, because it's arbitrary.

- For example, whether a morpheme with a particular affix is a prefix or a suffix varies across languages.

*Example 1:* Russian vs. Bulgarian comparative and superlative affixes.

[I think I didn't do this one in class, but here it is anyway]

| <u>Bulgarian:</u> | smešna / *po*-smešna | umna / *po*-umna |
|-----------|-----------|-----------|
| <u>Russian:</u> | smešna / smešn-*ee* | umna / umn-*ee*[3] |
|           | 'funny' / 'funnier' | 'smart' / 'smarter' |

*Example 2:* The affix meaning (roughly) -*ness* is a prefix in Kanuri, but a suffix (as in English) in Kurdish. It's also worth noting that English -*ness* attaches only (mostly) to adjectives, while its Kurdish cousin also attaches to nouns

**Kanuri (Nilo-Saharan, Nigeria)**

| gana   | 'small'     | nəmgana   | 'smallness'  |
|--------|-------------|-----------|--------------|
| kura   | 'big'       | nəmkura   | 'bigness'    |
| kurugu | 'long'      | nəmkurugu | 'length'     |
| karite | 'excellent' | nəmkarite | 'excellence' |
| dibi   | 'bad'       | nəmdibi   | 'badness'    |

---

[3] Ignore for present purposes the -*a* of *umna*, which is a morpheme marking it as a feminine adjective of a particular sort.

**Kurdish (Indo-Iranian, Iraq)**

| | | | |
|---|---|---|---|
| aaqil | 'wise' | aaqilii | 'forethought' |
| diz | 'robber' | dizii | 'robbery' |
| draiʒ | 'long' | draiʒii | 'length' |
| zaanaa | 'wise' | zaanaaii | 'erudition' |
| garm | 'warm' | garmii | 'warmth' |

We know: **prefix** (attach to left edge of base), **suffix** (attach to right edge of base)

- Now: **Infix**! Attach inside the base, according to a rule that refers to the base's sound structure.

**Example 1**: **Bontoc (Austronesian, Philippines)**
The infix *-um-* is positioned following the initial consonant of the base.

| | *'he is becoming ...'* | |
|---|---|---|
| fikas | fumikas | 'strong' |
| kilad | kumilad | 'red' |
| bato | bumato | 'stone' |
| fusul | fumusul | 'enemy' |

**Example 2: Samoan (Austronesian, Samoa)**
This example is interesting because it is also *reduplication*:
"copy penultimate syllable and position it next to the original"

| *'he ...'* | *'they ...'* | |
|---|---|---|
| manao | mananao | 'wish' |
| matua | matutua | 'be old' |
| malosi | malolosi | 'be strong' |
| punou | punonou | 'bend' |
| savali | savavali | 'travel' |
| pese | pepese | 'sing' |
| laga | lalaga | 'weave' |
| atamaʔi | atamamaʔi | 'be wise' |
| galue | galulue | 'work' |
| aloga | alologa | 'love' |

In these cases, the morpheme is attached in the middle of the sounds belonging to another morpheme. This type of affixation is called **infixation**.

**Example 3:** *fuckin'* infixation (fan-*fuckin'*-tastic, Massa-*fuckin'*-chusetts, Missi-*fuckin'*-ssippi)[4]

- The infixation of *fuckin'*, which denotes a strong positive or negative attitude, appears to be determined by *stress* (accent). A first approximation at the correct rule is:

---

(2)  *fuckin'*-**infixation**
Infix *fuckin'* to the left of the syllable that bears main stress.

Thus:
(3)  a. Massa-fuckin'-chusetts
but not :  b. *Massachu-fuckin's setts.

— because the syllable *-chu-* bears the main stress. Likewise:

(4)  a. Conn-fuckin'-ecticut
[it helps to pause after the first syllable, or to draw it out]
but not:  b. *Connect-fuckin'-icut.
nor:  c. *Connecti-fuckin'-cut.

- This example shows that the rule does involve stress, and not counting syllables from the end. since the stress in *Connecticut* falls on the third syllable from the end (the antepenult), while the stress in *Massachusetts* falls on the next-to-last syllable (the penult), as we've seen.

(5)  un-fuckin'-believable

- What is especially worth noting about the whole phenomenon is that it really is **rule-governed**! We have clear intuitions about good vs. bad places to infix *fuckin'* within bases to which we have surely never heard this infixation rule applied.

- People sometimes think that parental input and correction is a key to language acquisition. Pretty unlikely here: ("No my little sweetheart, you know we don't say *Massachu-fuckin'-setts*. Say it properly, or you can't have dessert: *Massa-fuckin'-chusetts*. Very good!").

## 9   Multiple affixes

- In many languages, verbs or nouns bear a whole group of affixes. Often these affixes mark **agreement**: they represent properties like "plural" or "3rd person" that have no meaning on the verb, but agree with a plural or 3rd person word or phrase somewhere else in the sentence.

- These agreement affixes often make it possible to leave out the plural or 3rd person words, and just use a verb to express thoughts like "we saw him".

- Languages with complex morphology of this sort present nice puzzles that will plague your problem sets and exams in this class!

---

[4] In British, Irish, Australian and New Zealand English, *bloody* can be used, and appears to follow exactly the same rule: *Massa-bloody-chusetts*, *Massachu-bloody-setts*, etc.

24.900 Introduction to Linguistics
Fall 2012