Basics of Morphology

Jirka Hana

(Based on slides from an ESSLLI 2010 course
by Anna Feldman & Jirka Hana)

What is morphology?

Morphology is the study of the *internal structure of words*.

- The first linguists were primarily morphologists.
- Well-structured lists of morphological forms of Sumerian words were attested on clay tablets from Ancient Mesopotamia and date from around 1600 BC; e.g. (Jacobsen 1974: 53-4),

badu	'he goes away'	*inǧen*	'he went'
baddun	'I go away'	*inǧenen*	'I went'
bašidu	'he goes away to him'	*inšiǧen*	'he went to him'
bašiduun	'I go away to him'	*inšiǧenen*	'I went to him'

Morphology (cont.)

- Morphology was also prominent in the writings of Pāṇini (5th century BC), and in the Greek and Roman grammatical tradition.
- Until the 19th century, Western linguists often thought of grammar as consisting primarily of rules determining word structure (because Greek and Latin, the classical languages had fairly rich morphological patterns).

- **Word-form**, **form**: A concrete word as it occurs in real speech or text.

Some terminology

- **Word-form**, **form**: A concrete word as it occurs in real speech or text. For our purposes, word is a string of characters separated by spaces in writing.

Some terminology

- **Word-form**, **form**: A concrete word as it occurs in real speech or text. For our purposes, word is a string of characters separated by spaces in writing.
- **Lemma**: A distinguished form from a set of morphologically related forms, chosen by convention (e.g., nominative singular for nouns, infinitive for verbs) to represent that set. Also called the canonical/base/dictionary/citation form. For every form, there is a corresponding lemma.

Some terminology (cont.)

- **Lexeme**: An abstract entity, a dictionary word; it can be thought of as a set of word-forms. Every form belongs to one lexeme, referred to by its lemma.
  For example, in English, *steal*, *stole*, *steals*, *stealing* are forms of the same lexeme STEAL; *steal* is traditionally used as the lemma denoting this lexeme.
- **Paradigm**: The set of word-forms that belong to a single lexeme.

(1) The paradigm of the Latin INSULA 'island'

	singular	plural
nominative	*insula*	*insulae*
accusative	*insulam*	*insulas*
genitive	*insulae*	*insularum*
dative	*insulae*	*insulis*
ablative	*insula*	*insulis*

Complications with terminology

The terminology is not universally accepted, for example:

- lemma and lexeme are often used interchangeably
- sometimes lemma is used to denote all forms related by derivation (see below).
- Paradigm can stand for the following:
  1. Set of forms of one lexeme
  2. A particular way of inflecting a class of lexemes (e.g. plural is formed by adding -s).
  3. Mixture of the previous two: Set of forms of an arbitrarily chosen lexeme, showing the way a certain set of lexemes is inflected.

Note: In our further discussion, we use lemma and lexeme interchangeably; and we use them both as an arbitrary chosen representative form standing for forms related by the same paradigm.

- **Morphemes** are the smallest meaningful constituents of words; e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme. Words are composed of morphemes (one or more).
  sing-er-s, home-work, un-kind-ly, flipp-ed, de-nation-al-iz-ation
  nej-ne-ob-hospod.ař-ova-tel-nějš-ího, auto-servis-u

Morpheme, Morph, Allomorph

- **Morphemes** are the smallest meaningful constituents of words; e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme. Words are composed of morphemes (one or more).
  sing-er-s, home-work, un-kind-ly, flipp-ed, de-nation-al-iz-ation
  nej-ne-ob-hospod.ař-ova-tel-nějš-ího, auto-servis-u

- **Morph**. The term morpheme is used both to refer to an abstract entity and its concrete realization(s) in speech or writing. When it is needed to maintain the signified and signifier distinction, the term **morph** is used to refer to the concrete entity, while the term morpheme is reserved for the abstract entity only.

Allomorphy

- **Allomorphs** are variants of the same morpheme, i.e., morphs corresponding to the same morpheme; they have the same function but different forms. Unlike the synonyms they usually cannot be replaced one by the other.

  (2) a. indefinite article: *an orange – a building*

  b. plural morpheme: *cat-s* [s] – *dog-s* [z] – *judg-es* [əz]

  c. opposite: *un-happy – in-comprehensive – im-possible – ir-rational*

Morphemes (cont.)

The order of morphemes/morphs matters:
talk-ed ≠ **ed-talk*, *re-write* ≠ **write-re*, *un-kind-ly* ≠ **kind-un-ly*

The order of morphemes/morphs matters:
talk-ed ≠ **ed-talk*, *re-write* ≠ **write-re*, *un-kind-ly* ≠ **kind-un-ly*

It is not always obvious how to separate a word into morphemes. For example, consider the *cranberry*-type morphemes. These are a type of bound morphemes that cannot be assigned a meaning or a grammatical function. The *cran* is unrelated to the etymology of the word *cranberry* (*crane* (the bird) + *berry*). Similarly, *mul* exists only in *mulberry*. There are other complications, e.g., zero-morphemes and empty morphemes.

Bound × Free Morphemes

- **Bound** – cannot appear as a word by itself.
  -*s* (*dog-s*), -*ly* (*quick-ly*), -*ed* (*walk-ed*)
- **Free** – can appear as a word by itself; often can combine with other morphemes too.
  house (*house-s*), *walk* (*walk-ed*), *of*, *the*, *or*

Past tense morpheme is a bound morpheme in English (*-ed*) but a free morpheme in Mandarine Chinese (*le*)

(3) a. *Ta chi le   fan*.
He eat past meal.
'He ate the meal.'

b. *Ta chi fan   le*.
He eat meal past.
'He ate the meal.'

- **Root** – nucleus of the word that affixes attach too.
  In English, most of the roots are free. In some languages that is less common (Lithuanian: *Billas Clintonas*).
  Some words (compounds) contain more than one root:
  home-work

Root × Affix (cont.)

- **Affix** – a morpheme that is not a root; it is always bound
    - **suffix**: follows the root
      English: *-ful* in *event-ful*, *talk-ing*, *quick-ly*, *neighbor-hood*
      Russian: *-a* in *ruk-a* 'hand'
    - **prefix**: precedes the root
      English: *un-* in *unhappy*, *pre-existing*, *re-view*
      Classical Nahuatl: *no-cal* 'my house'
    - **infix**: occurs inside the root
      English: very rare: *abso-bloody-lutely*
      Khmer: *-b-* in *lbeun* 'speed' from *leun* 'fast'; Tagalog: *-um-* in
      s-um-ulat 'write'
    - **circumfix**: occurs on both sides of the root
      Tuwali Ifugao *baddang* 'help', *ka-baddang-an* 'helpfulness',
      *ka-baddang, *baddang-an;
      Dutch: *berg* 'mountain' *ge-berg-te*, 'mountains', *geberg,
      *bergte; *vogel* 'bird', *ge-vogel-te* 'poultry', *gevogel, *vogelte

Affixing

- Suffixing is more frequent than prefixing and far more frequent than infixing/circumfixing (**sapir:1921**; **greenberg:1957**; **hawkins:gilligan:1988**).

  - Postpositional and head-final languages use suffixes and no prefixes;
  - But prepositional and head-initial languages use not only prefixes, as expected, but also suffixes.
  - Many languages use exclusively suffixes and no prefixes (e.g., Basque, Finnish),
  - Very few languages use only prefixes and no suffixes (e.g., Thai, but in derivation, not in inflection).

Affixing

- Suffixing is more frequent than prefixing and far more frequent than infixing/circumfixing (**sapir:1921**; **greenberg:1957**; **hawkins:gilligan:1988**).
    - Postpositional and head-final languages use suffixes and no prefixes;
    - But prepositional and head-initial languages use not only prefixes, as expected, but also suffixes.
    - Many languages use exclusively suffixes and no prefixes (e.g., Basque, Finnish),
    - Very few languages use only prefixes and no suffixes (e.g., Thai, but in derivation, not in inflection).

- Several attempts to explain this asymmetry (**hana:culicover:2008**):
    - processing arguments (**cutler:etal:1985**; **hawkins:gilligan:1988**),
    - historical arguments (**givon:1979**), and
    - combinations of both (**hall:1988**).

- **Content** morphemes – carry some semantic content
  car, *-able*, *un-*
- **Functional** morphemes – provide grammatical information
  the, *and*, *-s* (plural), *-s* (3rd sg)

There are two rather different kinds of morphological relationship among words, for which two technical terms are commonly used:

- **Inflection**: creates new forms of the same lexeme.
  E.g., *bring*, *brought*, *brings*, *bringing* are inflected forms of the lexeme BRING.

- **Derivation**: creates new lexemes
  E.g., *logic*, *logical*, *illogical*, *illogicality*, *logician*, etc. are derived from *logic*, but they all are different lexemes.

Inflection × Derivation

There are two rather different kinds of morphological relationship among words, for which two technical terms are commonly used:

- **Inflection**: creates new forms of the same lexeme.
  E.g., *bring*, *brought*, *brings*, *bringing* are inflected forms of the lexeme BRING.

- **Derivation**: creates new lexemes
  E.g., *logic*, *logical*, *illogical*, *illogicality*, *logician*, etc. are derived from *logic*, but they all are different lexemes.

- **Ending** – inflectional suffix

- **Stem** – word without its inflectional affixes = root + all derivational affixes.

- Derivation tends to affects the meaning of the word, while inflection tends to affect only its syntactic function.
- Derivation tends to be more irregular – there are more gaps, the meaning is more idiosyncratic and less compositional.
- However, the boundary between derivation and inflection is often fuzzy and unclear.

Inflection × Derivation (cont.)

	Derivational	Inflectional
category-changing	often	generally not
paradigmatic	no	yes
productivity	limited & variable	highly productive
type of meaning	often lexical	often purely grammatical
semantic regularity	often unpredictable	regular
restricted to specific syntactic env.	no	yes
position	central	peripheral
portmanteau forms	rarely	often
repeatable?	sometimes	never

(Based on R. Sproat's course notes and (Kroeger 2005:253)

- **Concatenation** (adding continuous affixes, without splitting the stem) – the most common process:
  - hope+less, un+happy, anti+capital+ist+s

Often, there are phonological changes on morpheme boundaries:
  - book+s [s], shoe+s [z]
  - happy+er → happi+er

Morphological processes (cont.)

- **Reduplication** – part of the word or the entire word is doubled:
  - Tagalog: *basa* 'read' – *ba-basa* 'will read'; *sulat* 'write' – *su-sulat* 'will write'
  - Afrikaans: *amper* 'nearly' – *amper-amper* 'very nearly'; *dik* 'thick' – *dik-dik* 'very thick'
  - Indonesian: *oraŋ* 'man' – *oraŋ-oraŋ* 'all sorts of men' (Cf. *orangutan*)
  - Samoan:

alofa	'love$_{Sg}$'	*a-lo-lofa*	'love$_{Pl}$'
galue	'work$_{Sg}$'	*ga-lu-lue*	'work$_{Pl}$'
la:poʔa	'to be large$_{Sg}$'	*la:-po-poʔa*	'to be large$_{Pl}$'
tamoʔe	'run$_{Sg}$'	*ta-mo-moʔe*	'run$_{Pl}$'

  - English: *humpty-dumpty*
  - American English (borrowed from Yiddish): *baby-schmaby*, *pizza-schmizza*

Morphological processes (cont.)

- **Templates** – both the roots and affixes are discontinuous.
  Only Semitic lgs (Arabic, Hebrew).
  Root (3 or 4 consonants, e.g., *l-m-d* – 'learn') is interleaved
  with a (mostly) vocalic pattern

  Hebrew:

lomed	'learn$_{masc}$'	shotek	'be-quiet$_{pres.masc}$'
lamad	'learned$_{masc.sg.3rd}$'	shatak	'was-quiet$_{masc.sg.3rd}$'
limed	'taught$_{masc.sg.3rd}$'	shitek	'made-sb-to-be-quiet$_{masc.sg.3rd}$'
lumad	'was-taught$_{masc.sg.3rd}$'	shutak	'was-made-to-be-quiet$_{masc.sg.3rd}$'

Morphological processes (cont.)

- **Suppletion** – 'irregular' relation between the words. Hopefully quite rare.
  - English:
    be – am – is – was,
    go – went,
    good – better
  - Czech:
    být 'to be' – *jsem* 'am',
    jít 'to go' – *šla* 'went$_{fem.sg}$,
    dobrý 'good' – *lepší* 'better'

Morphological processes (cont.)

- **Morpheme internal changes** (apophony, ablaut) – the word changes internally
    - English: *sing – sang – sung*, *man – men*, *goose – geese* (not productive anymore)
    - German: *Mann* 'man' – *Männ-chen* 'small man', *Hund* 'dog' – *Hünd-chen* 'small dog'
    - Czech: *kráva* 'cow$_{nom}$' – *krav* 'cows$_{gen}$',
      nés-t 'to carry' – *nes-u* 'I am carrying' – *nos-ím* 'I carry'

Morphological processes (cont.)

- **Subtraction (Deletion)**: some material is deleted to create another form
    - Papago (a native American language in Arizona)
      imperfective → perfective
      | *him* | 'walking$_{imperf}$' | → | *hi* | 'walking$_{perf}$' |
      | *hihim* | 'walking$_{pl.imperf}$' | → | *hihi* | 'walking$_{pl.perf}$' |
    - French
      feminine adjective → masculine adj. (much less clear)
      | *grande* [grɑ̃d] | 'big$_f$' | → | *grand* [grɑ̃] | 'big$_m$' |
      | *fausse* [fos] | 'false$_f$' | → | *faux* [fo] | 'false$_m$' |

Word formation: some examples

- **Affixation** – words are formed by adding affixes.
  - V + *-able* → Adj: *predict-able*
  - V + *-er* → N: *sing-er*
  - *un* + A → A: *un-productive*
  - A + *-en* → V: *deep-en*, *thick-en*

Word Formation (cont.)

- **Compounding** – words are formed by combining two or more words.
  - Adj + Adj → Adj: *bitter-sweet*
  - N + N → N: *rain-bow*
  - V + N → V: *pick-pocket*
  - P + V → V: *over-do*

- **Acronyms** – like abbreviations, but acts as a normal word
  laser – l̲ight a̲mplification by s̲imulated e̲mission of r̲adiation
  radar – r̲a̲dio d̲etecting a̲nd r̲anging

Word formation (cont.)

- **Acronyms** – like abbreviations, but acts as a normal word
  laser – <u>l</u>ight <u>a</u>mplification by <u>s</u>imulated <u>e</u>mission of <u>r</u>adiation
  radar – <u>ra</u>dio <u>d</u>etecting <u>a</u>nd <u>r</u>anging
- **Blending** – parts of two different words are combined
  - *breakfast + lunch → brunch*
  - *smoke + fog → smog*
  - *motor + hotel → motel*

Word formation (cont.)

- **Acronyms** – like abbreviations, but acts as a normal word
  laser – <u>l</u>ight <u>a</u>mplification by <u>s</u>imulated <u>e</u>mission of <u>r</u>adiation
  radar – <u>ra</u>dio <u>d</u>etecting <u>a</u>nd <u>r</u>anging
- **Blending** – parts of two different words are combined
  - *breakfast + lunch → brunch*
  - *smoke + fog → smog*
  - *motor + hotel → motel*
- **Clipping** – longer words are shortened
  - *doctor, professional, laboratory, advertisement, dormitory, examination, bicycle* (bike), *refrigerator*

Morphological types of languages

Morphology is not equally prominent in all languages. What one language expresses morphologically may be expressed by different means in another language.

- English: Aspect is expressed by certain syntactic structures:

  (4) a. John wrote (AE)/ has written a letter.   (the action is complete)

  b. John was writing a letter   (process).

- Russian: Aspect is marked mostly by prefixes:

  (5) a. John **na**pisal pis'mo.   (the action is complete)

  b. John pisal pis'mo.   (process).

Morphological types of languages (cont.)

There are two basic morphological types of language structure:

- **Analytic** languages – have only free morphemes, sentences are sequences of single-morpheme words.

  (6) Vietnamese:

khi	tôi	đên	nhà	bạn	tôi,	chúng	tôi	bat	dăù
when	I	come	house	friend	I,	PLURAL	I	begin	do

làm	bài
lesson	

  When I came to my friend's house, we began to do lessons.

- **Synthetic** – both free and bound morphemes. Affixes are added to roots.

Morphological types of languages (cont.)

Synthetic languages have further subtypes:

- **Agglutinating** – each morpheme has a single function, it is easy to separate them.
  E.g., Uralic lgs (Estonian, Finnish, Hungarian), Turkish, Basque, Dravidian lgs (Tamil, Kannada, Telugu), Esperanto

  Turkish:

	singular	plural	
nom.	ev	ev-ler	'house'
gen.	ev-in	ev-ler-in	
dat.	ev-e	ev-ler-e	
acc.	ev-i	ev-ler-i	
loc.	ev-de	ev-ler-de	
ins.	ev-den	ev-ler-den	

Morphological types of languages (cont.)

- **Fusional** – like agglutinating, but affixes tend to "fuse together", one affix has more than one function.
  E.g., Indo-European, Semitic, Sami (Skolt Sami, ...)

  - Czech *matk-a* 'mother' – *-a* means the word is a noun, feminine, singular, nominative.
  - Serbian/Croatian: the number and case of nouns is expressed by one suffix:

	singular	plural	
nominative	ovc-a	ovc-e	'*ovca* 'sheep'
genitive	ovc-e	ovac-a	
dative	ovc-i	ovc-ama	
accusative	ovc-u	ovc-e	
vocative	ovc-o	ovc-e	
instrumental	ovc-om	ovc-ama	

    Clearly, it is not possible to isolate separate singular or plural or nominative or accusative (etc.) morphemes.

- **Polysynthetic**: extremely complex, many roots and affixes combine together, often one word corresponds to a whole sentence in other languages.
  angyaghllangyugtuq – 'he wants to acquire a big boat' (Eskimo)
  palyamunurringkutjamunurtu – 's/he definitely did not become bad' (W Aus.)
  Sora

- English has many analytic properties (future morpheme *will*, perfective morpheme *have*, etc. are separate words) and many synthetic properties (plural (*-s*), etc. are bound morphemes).
- The distinction between analytic and (poly)synthetic languages is not a bipartition or a tripartition, but a continuum, ranging from the most radically isolating to the most highly polysynthetic languages.
  - It is possible to determine the position of a language on this continuum by computing its degree of synthesis, i.e., the ratio of morphemes per word in a random text sample of the language.

Morphological types of languages (cont.)

Language	Ration of morphemes per word
Greenlandic Eskimo	3.72
Sanskrit	2.59
Swahili	2.55
Old English	2.12
Lezgian	1.93
German	1.92
Modern English	1.68
Vietnamese	1.06

Table: The degree of synthesis of some languages (Haspelmath 2002)

Some difficulties in morpheme analysis

- Zero morpheme
  - Coptic:
    - ǰo-i   'my head'
    - ǰo-k   'your (masc.) head'
    - ǰo   'your (fem.) head'
    - ǰo-f   'his head'
    - ǰo-s   'her head'
  - Finnish:
    - oli-n   'I was'
    - oli-t   'you were'
    - oli   'he/she was'
    - oli-mme   'we were'
    - oli-tte   'you (pl.) were'
    - oli-vat   'they were'

Zero morpheme (cont.)

- Should all meanings be assigned to a morpheme?
  - If yes, then one is forced to posit zero morphemes (e.g., *oli*-Ø, where the morpheme Ø stands for the third person singular)
- But the requirement is not necessary, and alternatively one could say, for instance, that Finnish has no marker for the third person singular in verbs.

Empty morphemes

- The opposite of zero morphemes are *empty morphemes*.
  - Four of Lezgian's sixteen cases:

absolutive	sew	fil	Rahim
genitive	sew-**re**-n	fil-**di**-n	Rahim-a-n
dative	sew-**re**-z	fil-**di**-z	Rahim-a-z
subessive	sew-**re**-k	fil-**di**-k	Rahim-a-k
	'bear'	'elephant'	(male name)

- This suffix, called the *oblique stem* suffix in Lezgian grammar, has no meaning, but it must be posited if we want to have an elegant description.

- With the notion of an empty morpheme we can say that different nouns select different suppletive oblique stem suffixes, but that the actual case suffixes that are affixed to the oblique stem are uniform for all nouns.

- What is an alternative analysis?

Some more difficulties

Breton diminutive plurals:

bag	'boat'	bagòu	'boats'
bagig	'little boat'	bagòuigòu	'little boats'

Some more difficulties

Breton diminutive plurals:
bag   'boat'   bagòu   'boats'
bagig   'little boat'   bagòuigòu   'little boats'

English:
pick up – picker upper, *tuck in –tucker-inner*
(notice the regullar consonant doubling)

Momma aka diaper changer, snot wiper, head chef, laundry
specialist, maid, toy gatherer, taxi driver, boo-boo kisser,
tucker-inner*...well you get the point*