

# Intro to Linguistics – Morphology

Jirka Hana – October 31, 2011

## Overview of topics

1. Basic terminology
2. Classification of morphemes
3. Structure of words
4. Morphological processes
5. Word formation
6. Language Typology
7. Processing morphology

## 1 Basic terminology

- **Morphology** – study of internal structure of words
- **Morpheme** – the smallest linguistic unit which has a meaning or grammatical function. Words are composed of morphemes (one or more). There are some complications with this simple definition.

*sing·er·s, home·work, moon·light, un·kind·ly, talk·s, ten·th, flipp·ed, de·nation·al·iz·ation*

The order of morphemes matters:

*talk·ed ≠ \*ed·talk, re·write ≠ \*write·re*

- **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.
- **Allomorphs** – morphemes having the same function but different form. Unlike the synonyms they usually cannot be replaced one by the other.

- (1) a. indefinite article: *an orange – a building*  
b. plural morpheme: *cat·s [s] – dog·s [z] – judg·es [əz]*

- (2) a. *matk·a* ‘mother<sub>nom</sub>’ – *matek* ‘mothers<sub>gen</sub>’ – *matc·e* ‘mother<sub>dat</sub>’ – *matč·in* ‘mother’s’

## 2 Classification Of Morphemes

### 2.1 Bound × Free

- **Bound** – cannot appear as a word by itself.  
-s (*dog·s*), -ly (*quick·ly*), -ed (*walk·ed*);  
-te (*dělá·te* ‘do<sub>2pl</sub>’), -y (*žen·y* ‘women’), vy- (*vy·jít* ‘walk out’)
- **Free** – can appear as a word by itself; often can combine with other morphemes too.  
*house* (*house·s*), *walk* (*walk·ed*), *of*, *the*, *or*  
*hrad* ‘castle’, *žen* ‘woman<sub>root</sub> = gen.pl.’, *přes* ‘over’, *nebo* ‘or’

Past tense morpheme is a bound morpheme in English (-ed) but a free morpheme in Mandarin 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.’

### 2.2 Root × Affix

- **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*).  
Compounds contain more than one root: *home·work*; *železo·beton* ‘reinforced concrete’
- **affix** – a morpheme that is not a root; it is always bound
  - **suffix**: *talk·ing*, *quick·ly*; *mal·ý* ‘small<sub>masc.sg.nom</sub>’, *kup·ova·t* ‘buy<sub>imperf</sub>’
  - **prefix**: *un·happy*, *pre·existing*; *do·psat* ‘finish writing’, *nej·méně* ‘least’
  - **infix**: common in Austronesian and Austroasiatic lgs (Tagalog, Khmer)  
Tagalog: *basa* ‘read’ *b·um·asa* ‘read<sub>past</sub>’ – *sulat* ‘write’ – *s·um·ulat* ‘wrote’  
very rare in English: *abso·bloody·lutely*,
  - **circumfix**: morpheme having two parts that are placed around a stem  
Dutch collectives:  
*berg* ‘mountain’ *ge·berg·te* ‘mountains’ \**geberg*, \**bergte*  
*vogel* ‘bird’ *ge·vogel·te* ‘poultry’ \**gevogel*, \**vogelte*  
Czech *po+...+í*:  
*Vltava* → *Po·vltav·í* ‘Vltava river area’ (\**povltava*, \**vltaví*);  
*Pobaltí*, *pohoří*, *pohraničí*, *potrubí*, *pobřeží*, *polesí*

Suffixes more common than prefixes which are more common than infixes/circumfixes

## 2.3 Content × Functional

- **Content** morphemes – carry some semantic content  
*car, -able, un-*
- **Functional** morphemes – provide grammatical information  
*the, and, -s (plural), -s (3<sup>rd</sup> sg)*  
*jsem ‘past aux<sub>1sg</sub>’, -a ‘gen.sg’ (měst·a ‘town<sub>gen</sub>’)*

## 2.4 Derivation vs. Inflection

- **inflection** – creating various forms of the same word  
**lexeme** – an abstract entity; the set of all forms related by inflection (but not derivation).  
*table – table·s*  
*uč·í·m – uč·í·š – uč·í – uč·í·me*  
**lemma:** A form from a lexeme chosen by convention (e.g., nom.sg. for nouns, infinitive for verbs) to represent that set.  
Also called the canonical/base/dictionary/citation form.  
E.g., *break, breaks, broke, broken, breaking* have the same lemma *break*  
**ending** – inflectional suffix
- **derivation** – creating new words  
*slow – slow·ly – slow·ness*  
*uč·í·t – uč·í·tel – uč·í·tel·ka – uč·í·tel·ský – uč·í·tel·ova·t – vy·uč·ova·t*

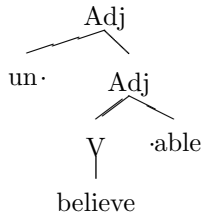
Inflection vs. Derivation:

- Derivation tends to affect 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.

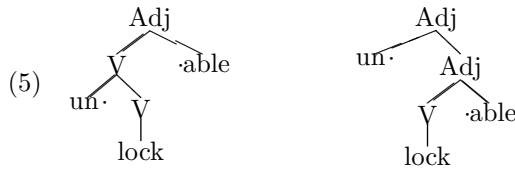
## 3 Structure of words

Structure of words can be captured in a similar way as structure of sentences.

- (4) unbelievable = un + (believ + able),  
not \*(un + believe) + able.



Some words can be ambiguous:



## 4 Morphological processes

- **Concatenation** (adding continuous affixes) – the most common process

Often phonological changes on morpheme boundaries.

- **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: *oraj* ‘man’ – *oraj-oraj* ‘all sorts of men’ (Cf. *orangutan*)
- Samoan:
 

<i>alofa</i>	‘love <sub>sg</sub> ’	<i>a-lo-lofa</i>	‘love <sub>pl</sub> ’
<i>galue</i>	‘work <sub>sg</sub> ’	<i>ga-lu-lue</i>	‘work <sub>pl</sub> ’
<i>la:poʔa</i>	‘to be large <sub>sg</sub> ’	<i>la:po-poʔa</i>	‘to be large <sub>pl</sub> ’
<i>tamoʔe</i>	‘run <sub>sg</sub> ’	<i>ta-mo-moʔe</i>	‘run <sub>pl</sub> ’
- English: *humpty-dumpty*
- American English (borrowed from Yiddish): *baby-schmaby*, *pizza-schmizza*

- **Templates** – both root and affix

Both the roots and affixes are discontinuous. Only Semitic lgs (Arabic, Hebrew). A root (3 or 4 consonants, e.g., *l-m-d* – ‘learn’) is interleaved with a (mostly) vocalic pattern

- Hebrew:
 

lomed	‘learn <sub>masc</sub> ’	shatak	‘be-quiet <sub>pres.masc</sub> ’
lamad	‘learnt <sub>masc.sg.3rd</sub> ’	shatak	‘was-quiet <sub>masc.sg.3rd</sub> ’
limed	‘taught <sub>masc.sg.3rd</sub> ’	shitek	‘made-sb-to-be-quiet <sub>masc.sg.3rd</sub> ’
lumad	‘was-taught <sub>masc.sg.3rd</sub> ’	shutak	‘was-made-to-be-quiet <sub>masc.sg.3rd</sub> ’

- **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* – *krav*, *nés·t* – *nes·u* – *nos·ím*

- **Subtraction (Deletion)**: some material is deleted to create another form

- Papago (a native American language in Arizona) imperfective → perfective
  - hím* ‘walking<sub>imperf</sub>’ → *hí* ‘walking<sub>perf</sub>’
  - híhím* ‘walking<sub>pl.imperf</sub>’ → *híhí* ‘walking<sub>pl.perf</sub>’
- French, feminine adjective → masculine adj. (much less clear)
  - grande* [grãd] ‘big<sub>f</sub>’ → *grand* [grã] ‘big<sub>m</sub>’
  - fausse* [fos] ‘false<sub>f</sub>’ → *faux* [fo] ‘false<sub>m</sub>’

- **Suppletion** – ‘irregular’ relation between the words. Hopefully quite rare.

- Czech: *být* – *jsem*, *jít* – *šla*, *dobrý* – *lepší*
- English: *be* – *am* – *is* – *was*, *go* – *went*, *good* – *better*

## 5 Word formation

- **Affixation** – words are formed by adding affixes.

English:

- V + *-able* → Adj: *predict·able*
- V + *-er* → N: *sing·er*
- *un-* + A → A: *un·productive*
- V + *-en* → V: *deep·en*, *thick·en*

Czech:

- N + *-ov·ý* → Adj: *motor·ový*
- V + *-tel* → N: *spisova·tel*, *stavi·tel*
- N + *-ova·t* → V: *pan·ova·t*, *parazit·ova·t*,

- **Compounding** – words are formed by combining two or more words.

English:

- Adj + Adj → Adj: *bitter·sweet*
- N + N → N: *rain·bow*, *Internet Security Association Key Management Protocol*
- V + N → V: *pick·pocket*
- P + V → V: *over·do*

Czech:

- N + N → N: *maso·žravec* ‘carnivore’, *vzducho·loď* ‘airship’
- A + N → N: *černo·zem* ‘black soil’, *plno·vous* ‘beard’
- A(dv) + A → A: *star·o·česk·ý* ‘Old Czech’, *tmav·o·modr·ý* ‘dark blue’

German: *Donau·dampf·schiff·fahrts·gesellschafts·kapitän*

- **Acronyms** – like abbreviations, but acts as a normal word  
*laser* – *light amplification by simulated emission of radiation*  
*radar* – *radio detecting and ranging*  
*Čedok* – *Česká dopravní kancelář*
- **Blending** – parts of two different words are combined  
*breakfast + lunch* → *brunch*  
*smoke + fog* → *smog*  
*motor + hotel* → *motel*
- **Clipping** – longer words are shortened  
*doc(tor)*, *prof(essional)*, *lab(oratory)*, *ad(vertisement)*, *dorm(itory)*, *exam(ination)*  
*auto(mobil)*

## 6 Morphological Types Of Languages

Two basic morphological types of languages:

- **Analytic** (isolating) languages – have only free morphemes, sentences are sequences of single-morpheme words.
- **Synthetic** languages – both free and bound morphemes. Affixes are added to roots.

### 6.1 Analytic languages

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

(6) Vietnamese:

khi tôi đến nhà bạn tôi, chúng tôi bắt đầu làm bài (Comrie 1989)  
when I come house friend I, PLURAL I begin do lesson

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

### 6.2 Synthetic languages

Synthetic languages have both free and bound morphemes.

Has 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:

	sg.	pl.	
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	

- **Fusional** – like agglutinating, but affixes tend to “fuse together”, one affix has more than one function. Common homonymy of inflectional affixes.

*matk·a* ‘mother’ – *a* means the word is a noun, feminine, singular, nominative.

E.g., Slavic, Romance languages, Greek

(7) Homonymy of the *a* ending in Czech:

form	lemma	gloss		category
měst-a	město	town	NS2	noun neut sg gen
			NP1 (5)	noun neut pl nom (voc)
			NP4	noun neut pl acc
tém-a	téma	theme	NS1 (5)	noun neut sg nom (voc)
			NS4	noun neut sg acc
žen-a	žena	woman	FS1	noun fem sg nom
pán-a	pán	man	MS2	noun masc anim sg gen
			MS4	noun masc anim sg acc
ostrov-a	ostrov	island	IS2	noun masc inanim sg gen
předsed-a	předseda	president	MS1	noun masc anim sg nom
vidě-l-a	vidět	see		verb past fem sg
				verb past neut pl
vidě-n-a				verb passive fem sg
				verb passive neut pl
vid-a				verb transgressive masc sg
dv-a	dv-a	two		numeral masc sg nom
				numeral masc sg acc

(8) Ending *-e* and noun cases in Czech:

case	form	lemma	gender	gloss
nom	kuř-e	kuře	neuter	chicken
gen	muž-e	muž	masc.anim.	man
dat	mouš-e	moucha	feminine	fly
acc	muž-e	muž	masc.anim.	man
voc	pan-e	pán	masc.anim.	mister
loc	mouš-e	moucha	feminine	fly
inst	–	–		

- **Polysynthetic:** extremely complex, many roots and affixes combine together, often one word corresponds to a whole sentence in other languages.

*angyaghllangyugtug* – ‘he wants to acquire a big boat’ (Eskimo)

*palyamunurriŋkutjamunurtu* – ‘s/he definitely did not become bad’ (W Aus.)

Sora – LF, p. 132

### 6.3 Morphology in real languages

Czech – mostly fusional, but also other properties:

- analytic: future and past tense, conditional, prepositions, ...
- agglutinating: prefixes/suffixes; *vidě·n·a* ‘seen<sub>fem.sg</sub>’ -n- – passive, -a – fem+sg

English – originally fusional, but now both analytic properties (future morpheme *will*, perfective morpheme *have*, etc. are separate words) and synthetic properties (plural (*s*), etc. are bound morphemes)

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

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