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·i·za·tion`

  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* [əs]

  (2) a. *matk·a* ‘mother\textsubscript{nom}’ – *matek* `mothers\textsubscript{gen}` – *matc·e* `mother\textsubscript{dat}` – *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 ‘do2pl’), -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), af, the, or
  - hrad ‘castle’, žen ‘woman’, root = 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 či le fan.
   He eat past meal.
   ‘He ate the meal.’

b. Ta či 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: Būlas 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_masc.sg.nom’, kup-o-vat ‘buy Imperf’
  - 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-am-asa ‘readpast’ – sulat ‘write’ – s-am-ualat ‘wrote’
    very rare in English: abso-bloody-lately,
  - 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, *vltavi);
    Pobaltí, pohorší, 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 (3rd sg)
  - *jsen* ‘past aux1sg’, -a ‘gen.sg’ (*měst-a ‘towngen’*)

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*
  - *učím – učíš – učí – učím*

  - **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
  - *slow – slow-ly – slow-ness*

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 + (believe + able),
not *(un + believe) + able.*
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: *orang* ‘man’ – *orang-orang* ‘all sorts of men’ (Cf. *orangutan*)
  - Samoan:
    - *alofo* ‘love\(_Sg\)*
    - *a-lo-lofa* ‘love\(_Pl\)*
    - *galue* ‘work\(_Sg\)*
    - *ga-la-lue* ‘work\(_Pl\)*
    - *la-po\(_a\)* ‘to be large\(_Sg\)*
    - *la-po-po\(_a\)* ‘to be large\(_Pl\)*
    - *tamo\(_e\)* ‘run\(_Sg\)*
    - *ta-mo-mo\(_e\)* ‘run\(_Pl\)*
  - 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:
    - *lo*\(_med\)* ‘learn\(_masc\)*
    - *shata\(_k\)* ‘be-quiet\(_pres.masc\)*
    - *la*\(_m\)*\(_ad\)* ‘learnt\(_masc.sg.3rd\)*
    - *shata\(_k\)* ‘was-quiet\(_masc.sg.3rd\)*
    - *lima\(_d\)*\(_med\)* ‘taught\(_masc.sg.3rd\)*
    - *shite\(_k\)* ‘made-sb-to-be-quiet\(_masc.sg.3rd\)*
    - *luma\(_d\)*\(_med\)* ‘was-taught\(_masc.sg.3rd\)*
    - *shutak\(_e\)* ‘was-made-to-be-quiet\(_masc.sg.3rd\)*

- **Morpheme internal changes** (apophony, ablaut) – the word changes internally
  - English: *sing* – *sang* – *sung*, *man* – *men*, *goose* – *geese* (not productive anymore)
– German: Mann ‘man’ – Männchen ‘small man’, Hund ‘dog’ – Hündchen ‘small dog’
– Czech: kráva – krav, něs-t – nes-ť – nosím

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

 – Papago (a native American language in Arizona) imperfective → perfective
   him ‘walking’ \( \rightarrow \) hi ‘walking’
   him ‘walkingpl’ \( \rightarrow \) hihi ‘walkingpl’

 – French, feminine adjective → masculine adj. (much less clear)
   grande [grăd] ‘big’ \( \rightarrow \) grand [grăn] ‘bigm’
   fausse [foz] ‘false’ \( \rightarrow \) faux [fo] ‘falselm’

• **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 \( \rightarrow \) Adj: predict-able
  – V + -er \( \rightarrow \) N: sing-er
  – un- + A \( \rightarrow \) A: un-productive
  – V + -en \( \rightarrow \) V: deep-en, thick-en

Czech:
  – N + -ov-ý \( \rightarrow \) Adj: motor-ový
  – V + -tel \( \rightarrow \) N: spisova-tel, stavi-tel
  – N + -ova-t \( \rightarrow \) V: pan-ova-t, parazit-ova-t,

• **Compounding** – words are formed by combining two or more words.
  English:
  – Adj + Adj \( \rightarrow \) Adj: bitter-sweet
  – N + N \( \rightarrow \) N: rain-bow, Internet Security Association Key Management Protocol
  – V + N \( \rightarrow \) V: pick-pocket
  – P + V \( \rightarrow \) V: over-do

Czech:
  – N + N \( \rightarrow \) N: maso-žravec ‘carnivore’, vzducho-řod ‘airship’
  – A + N \( \rightarrow \) N: černo-zem ‘black soil’, pno-voz ‘beard’
  – A(dv) + A \( \rightarrow \) A: star-o-česk-ý ‘Old Czech’, tmav-o-modr-ý ‘dark blue’
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:

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom.</td>
<td>ev</td>
<td>ev-lar</td>
</tr>
<tr>
<td>gen.</td>
<td>ev-in</td>
<td>ev-lar-in</td>
</tr>
<tr>
<td>dat.</td>
<td>ev-e</td>
<td>ev-lar-e</td>
</tr>
<tr>
<td>acc.</td>
<td>ev-i</td>
<td>ev-lar-i</td>
</tr>
<tr>
<td>loc.</td>
<td>ev-de</td>
<td>ev-lar-de</td>
</tr>
<tr>
<td>ins.</td>
<td>ev-den</td>
<td>ev-lar-den</td>
</tr>
</tbody>
</table>

• **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:

<table>
<thead>
<tr>
<th>Form</th>
<th>Lemma</th>
<th>Gender</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>měst-a</td>
<td>město</td>
<td>NS2</td>
<td>noun neut sg gen</td>
</tr>
<tr>
<td>NP1 (5)</td>
<td>noun neut pl nom (voc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP4</td>
<td>noun neut pl acc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tém-a</td>
<td>téma</td>
<td>NS1 (5)</td>
<td>noun neut sg nom (voc)</td>
</tr>
<tr>
<td>NP4</td>
<td>noun neut sg acc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>žen-a</td>
<td>žena</td>
<td>FS1</td>
<td>noun fem sg nom</td>
</tr>
<tr>
<td>pán-a</td>
<td>pán</td>
<td>MS2</td>
<td>noun masc anim sg gen</td>
</tr>
<tr>
<td>MS4</td>
<td>noun masc anim sg acc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ostrov-a</td>
<td>ostrov</td>
<td>IS2</td>
<td>noun masc inanim sg gen</td>
</tr>
<tr>
<td>předseda</td>
<td>předseda</td>
<td>MS1</td>
<td>noun masc anim sg nom</td>
</tr>
<tr>
<td>víde-l-a</td>
<td>vídeť</td>
<td>noun</td>
<td>verb past fem sg</td>
</tr>
<tr>
<td>MS1</td>
<td>verb passive fem sg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vid-a</td>
<td>verb transgressive masc sg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dv-a</td>
<td>dv-a</td>
<td>numeral</td>
<td>noun</td>
</tr>
<tr>
<td>numeral masc sg nom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>numeral masc sg acc</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Case</th>
<th>Form</th>
<th>Gender</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom.</td>
<td>kur-e</td>
<td>neuter</td>
<td>chicken</td>
</tr>
<tr>
<td>gen.</td>
<td>muž-e</td>
<td>masc.anim.</td>
<td>man</td>
</tr>
<tr>
<td>dat.</td>
<td>mouš-e</td>
<td>feminine</td>
<td>fly</td>
</tr>
<tr>
<td>acc.</td>
<td>muž-e</td>
<td>masc.anim.</td>
<td>man</td>
</tr>
<tr>
<td>voc.</td>
<td>pan-e</td>
<td>masc.anim.</td>
<td>mister</td>
</tr>
<tr>
<td>loc.</td>
<td>mouš-e</td>
<td>feminine</td>
<td>fly</td>
</tr>
<tr>
<td>inst.</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

• **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)
*palyamunurringkutjamanurtu* – ‘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\_fem\_sg’ -n- – passive, -a – fem+sg

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

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

<table>
<thead>
<tr>
<th>Language</th>
<th>Ration of morphemes per word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenlandic Eskimo</td>
<td>3.72</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>2.59</td>
</tr>
<tr>
<td>Swahili</td>
<td>2.55</td>
</tr>
<tr>
<td>Old English</td>
<td>2.12</td>
</tr>
<tr>
<td>Lezgian</td>
<td>1.93</td>
</tr>
<tr>
<td>German</td>
<td>1.92</td>
</tr>
<tr>
<td>Modern English</td>
<td>1.68</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>1.06</td>
</tr>
</tbody>
</table>