ESSLLI 2013: Computational Morphology

Instructors: Jirka Hana & Anna Feldman
August 5-9, 2013
Overview of the course

1. Basics of Phonetics, Phonology & Morphology
2. Classical approaches to morphological analysis
3. Classical tagging techniques
4. Tagset Design and Morphosyntactically Annotated Corpora
5. Unsupervised and Resource-light Approaches to Computational Morphology:
   - Monolingual approaches
   - Cross-lingual approaches
Basics of Phonetics, Phonology & Morphology

ESSLLI 2013: Computational Morphology

Jirka Hana & Anna Feldman
Phonology/Phonetics in 5 slides

What is morphology
Words, Morphemes, etc.
Morphological processes
Typology
Complications

Jirka Hana & Anna Feldman
Basics of Phonetics, Phonology & Morphology
Consonants vs. vowels:

- consonants – involve some constriction
- vowels – no constriction; can always be held indefinitely.
Consonants vs. vowels:

- consonants – involve some constriction
- vowels – no constriction; can always be held indefinitely.

Sounds are encoded by some phonetic alphabet, e.g. IPA (International P.A.)
Describing Consonants

Voicing – do vocal folds vibrate?
- voiced – yes: [b], [d], [g], [m], [n], [ŋ], [z], [ʒ], ...
- voiceless – no: [p], [t], [k], [s], [ʃ], ...

Place of Articulation – place of obstruction (lips, teeth, ...)
- Bilabial – [p], [b], [m], [w]
- Labiodental – [f], [v]
- Interdental – [θ], [ð]
- Alveolar: – [t], [d], [s], [z], [n], ...
- Palatal: – [ʃ] (ship), [ʒ] (visual), [j] (yes)
- Velar: – [k], [g], [ŋ] (walking)
- etc.

Manner of Articulation – degree of obstruction
- Stops: flow of air is stopped – [p, b, t, d, k, g]
- Fricatives: narrow constriction – [f, v, θ, ð, s, ʃ, z, ʒ, x, h]
- Nasals (nasal stops): air passes also through the nose – [m, n, ŋ]
Phonological Rules

Phonological rules translate some mental/underlying representation of words to the real sounds

(1) /p/ → [p]h
(2) /p/ → [p]h at the beginning of the word
Similarly for /t/ and /k/ ([p,t,k] are voiceless stops):
(3) voiceless stop → aspirated at the beginning of the word
(4) [–voiced, +stop] → [+aspirated] / #
Phonological Rules

Phonological rules translate some mental/underlying representation of words to the real sounds

(1) a. /pɪt/ → [pʰɪt]
   b. /spɪt/ → [spɪt]
Phonological Rules

Phonological rules translate some mental/underlying representation of words to the real sounds

(1) a. /pɪt/ → [pʰɪt]
   b. /spɪt/ → [spɪt]

(2) /p/ → [pʰ] at the beginning of the word
Phonological Rules

Phonological rules translate some mental/underlying representation of words to the real sounds

(1) a. /pɪt/ → [pʰɪt]
   b. /spɪt/ → [spɪt]

(2) /p/ → [pʰ]   at the beginning of the word

Similarly for /t/ and /k/ ( [p,t,k] are voiceless stops):
Phonological Rules

Phonological rules translate some mental/underlying representation of words to the real sounds

(1) a. /pɪt/ → [pʰɪt]
   b. /spɪt/ → [spɪt]

(2) /p/ → [pʰ] at the beginning of the word

Similarly for /t/ and /k/ ( [p,t,k] are voiceless stops):

(3) voiceless stop → aspirated at the beginning of the word
Phonological Rules

Phonological rules translate some mental/underlying representation of words to the real sounds

(1) a. /pɪt/ → [pʰɪt]
   b. /spɪt/ → [spɪt]

(2) /p/ → [pʰ] at the beginning of the word

Similarly for /t/ and /k/ ( [p,t,k] are voiceless stops):

(3) voiceless stop → aspirated at the beginning of the word

(4) [−voiced, +stop] → [+aspirated] / #
Natural classes

In these rules we can refer to classes of sounds like:

- voiced consonants ([b, d, g, ð, z, ʒ, n, m, ...]),
- rounded vowels ([u, u, o, ɔ]),
- nasals ([m, n, ɳ]),
- sibilants (hissy sounds [s, z, ʃ, ʒ, tʃ, dʒ]),
- etc.
Common types of phonological rules

- Assimilation – a sound becomes more like a nearby sound.
  - place assimilation: *comfort* [mʃ], *input* [mp]
  - voicing assimilation *talks* [ks] vs. *dogs* [gz]
  - *bit* [i] vs. *bin* [ɪ] – /ɪ/ assimilates to the following /n/ (nasal)
    easier to pronounce

- Insertion
  - *hamster* /hæmstr/ → [hæm(p)str]: [p] is sometimes inserted

- Deletion
  - *okay* [okey] → [ʔkey] (optional)
    easier and faster to say

- etc
What is morphology
What is morphology?

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

<table>
<thead>
<tr>
<th>Sumerian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>badu</em></td>
<td>‘he goes away’</td>
</tr>
<tr>
<td><em>baddun</em></td>
<td>‘I go away’</td>
</tr>
<tr>
<td><em>bašidu</em></td>
<td>‘he goes away to him’</td>
</tr>
<tr>
<td><em>bašiduun</em></td>
<td>‘I go away to him’</td>
</tr>
<tr>
<td><em>ingēn</em></td>
<td>‘he went’</td>
</tr>
<tr>
<td><em>ingēnen</em></td>
<td>‘I went’</td>
</tr>
<tr>
<td><em>inšiğen</em></td>
<td>‘he went to him’</td>
</tr>
<tr>
<td><em>inšiğenen</em></td>
<td>‘I went to him’</td>
</tr>
</tbody>
</table>
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).
<table>
<thead>
<tr>
<th>Phonology/Phonetics in 5 slides</th>
<th>Morphemes, Morph, Allomorphy</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is morphology</td>
<td>Bound &amp; Free</td>
</tr>
<tr>
<td>Words, Morphemes, etc.</td>
<td>Root &amp; Affix, Affixes</td>
</tr>
<tr>
<td>Morphological processes</td>
<td>Content &amp; Functional</td>
</tr>
<tr>
<td>Typology</td>
<td>Inflection (\times) Derivation</td>
</tr>
</tbody>
</table>

**Words, Morphemes, etc.**
**Word-form**

- **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.
**Morphemes** are the smallest meaningful constituents of words; Words are composed of morphemes (one or more).

*sing-er-s, home-work, un-kind-ly, flipp-ed, de-nation-al-iz-ation*
Morpheme, Morph

- **Morphemes** are the smallest meaningful constituents of words; Words are composed of morphemes (one or more).
  
  *sing-er-s, home-work, un-kind-ly, flipp-ed, de-nation-al-iz-ation*

- **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 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.

(5) a. indefinite article: an orange – a building
    b. plural morpheme: cat-s [s] – dog-s [z] – judg-es [əz]
The order of morphemes/morphs matters:

\[ \text{talk-ed} \neq *\text{ed-talk}, \text{re-write} \neq *\text{write-re}, \text{un-kind-ly} \neq *\text{kind-un-ly} \]
**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)

(6) 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 × 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*
Root \times 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*

- **Affix** – a morpheme that is not a root; it is always bound
Root \times Affix

- **Root** – nucleus of the word that affixes attach to. 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*

- **Affix** – a morpheme that is not a root; it is always bound
  - suffix
  - prefix
  - infix
  - circumfix
Types of affixes

- **suffix** – follows the root
  - *talk-ing, quick-ly;*
  - Russian: *ruk-a* ‘hand’

- **prefix** – precedes the root
  - *un-happy, pre-existing;*
  - Czech: *do-psat* ‘finish writing’, *nej-méně* ‘least’
Types of affixes – infix

- **infix**: occurs inside the root
  common in Austronesian and Austroasiatic lgs
Types of affixes – infix

- **infix**: occurs inside the root
  common in Austronesian and Austroasiatic lgs

  Tagalog (Philippines): \( \text{basa} \) ‘read’ \( \text{b-um-asa} \) ‘read_{past}’
  \( \text{sulat} \) ‘write’ \( \text{s-um-ulat} \) ‘wrote’

  Khmer (Cambodia): \( \text{leun} \) ‘fast’ – \( \text{l-b-eun} \) ‘speed’
**Types of affixes – infix**

- **infix**: occurs inside the root
  - common in Austronesian and Austroasiatic lgs

Tagalog (Philippines):

- *basa* ‘read’ → *b-um-asa* ‘read\textsubscript{past}’
- *sulat* ‘write’ → *s-um-ulat* ‘wrote’

Khmer (Cambodia):

- *leun* ‘fast’ → *l-b-eun* ‘speed’

very rare in English: *abso-bloody-lutely,*
Types of affixes – infix

- **infix**: occurs inside the root
  common in Austronesian and Austroasiatic lgs

Tagalog (Philippines):

- *basa* ‘read’  
  *b-um-asa* ‘read<sub>past</sub>’
- *sulat* ‘write’  
  *s-um-ulat* ‘wrote’

Khmer (Cambodia):

- *leun* ‘fast’  
  *l-b-eun* ‘speed’

very rare in English: *abso-bloody-lutely*,

The places in the stem where infixing can occur are quite restricted:  
either in the second or prefinal position, where various units are  
counted – syllables, moras, consonants, vowels, etc. (Hoeksema &  
Types of affixes – circumfix

- **circumfix**: occurs on both sides of the root
Types of affixes – circumfix

- **circumfix**: occurs on both sides of the root

  Dutch collectives:
  
  - *berg* 'mountain'  *ge-berg-te* 'mountains'  *geberg, bergte*
  - *vogel* 'bird'  *ge-vogel-te* 'poultry'  *gevogel, vogelte*
Types of affixes – circumfix

- **circumfix**: occurs on both sides of the root

  **Dutch collectives:**
  
  berg  'mountain'  ge-berg-te  'mountains'  *geberg, *bergte
  vogel  'bird'  ge-vogel-te  'poultry'  *gevogel, *vogelte

  **Tuwali (Philippines):**
  
  baddang  'help',  ka-baddang-an  'helpfulness',  *ka-baddang,  *baddang-an;
Types of affixes – circumfix

- **circumfix**: occurs on both sides of the root

  Dutch collectives:
  - *berg* 'mountain' → *ge-berg-te* 'mountains' → *geberg, *bergte*
  - *vogel* 'bird' → *ge-vogel-te* 'poultry' → *gevogel, *vogelte*

  Tuwali (Philippines):

  Czech *po+...+í:*
  - *Vltava* → *Po-vltav-í* ‘Vltava river area’ (*povltava, *vltaví;*
  - *Pobaltí* ‘Baltics’, *pohranicí* ‘border region’, *pobřeží* ‘sea shore’, *pohoří*, *potrubí*, *polesí*
Suffixed is more frequent than prefixing and far more frequent than infixing/circumfixing (Sapir 1921, Greenberg 1957, Hawkins & Gilligan 1988).
Sufffixing 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 (see Hana & Culicover 2008 for an overview):
  - processing arguments (Cutler et al 1985, Hawkins & Gilligan 1988),
  - historical arguments (Givön 1979), and
  - combinations of both (Hall 1988).
Content × Functional

- **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 word
  - *bring, brought, brings, bringing*

- **Derivation**: creates new words
  - *logic, logical, illogical, illogicality, logician, etc.*
**Lexeme**: An abstract entity; the set of all forms related by inflection (but not derivation).
Inflection

- **Lexeme**: An abstract entity; the set of all forms related by inflection (but not derivation).

- **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.
Inflection

- **Lexeme**: An abstract entity; the set of all forms related by inflection (but not derivation).

- **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.

- **Ending** – inflectional suffix

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

The forms of the Latin *insula* ‘island’

<table>
<thead>
<tr>
<th>Case</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominative</td>
<td><em>insula</em></td>
<td><em>insulae</em></td>
</tr>
<tr>
<td>accusative</td>
<td><em>insulam</em></td>
<td><em>insulae</em></td>
</tr>
<tr>
<td>genitive</td>
<td><em>insulae</em></td>
<td><em>insularum</em></td>
</tr>
<tr>
<td>dative</td>
<td><em>insulae</em></td>
<td><em>insulis</em></td>
</tr>
<tr>
<td>ablative</td>
<td><em>insula</em></td>
<td><em>insulis</em></td>
</tr>
</tbody>
</table>
The terminology is common, but not universally accepted, for example:

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

- 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.
Inflection × Derivation (cont.)

- 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.
  - *unhappy, unacceptable;*
Inflection × Derivation (cont.)

- 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.
  
  - *unhappy, unacceptable; unloosen = loosen, unnerve, unearth*
Inflection × Derivation (cont.)

- 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.
  - unhappy, unacceptable; unloosen = loosen, unnerve, unearth
  - sincere > sincerity, scarce > scarcity, curious > curiosity;
Inflection × Derivation (cont.)

- 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.
  
  - *unhappy, unacceptable; unloosen = loosen, unnerve, unearth
  - *sincere > sincerity, scarce > scarcity, curious > curiosity; fierce but *fiercity
 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.

- *unhappy, unacceptable; unloosen = loosen, unnerve, unearth
- *sincere > sincerity, scarce > scarcity, curious > curiosity; fierce but *fiercity

However, the boundary between derivation and inflection is often fuzzy and unclear.
## Inflection × Derivation (cont.)

<table>
<thead>
<tr>
<th></th>
<th>Derivational</th>
<th>Inflectional</th>
</tr>
</thead>
<tbody>
<tr>
<td>category-changing</td>
<td>often</td>
<td>generally not</td>
</tr>
<tr>
<td>paradigmatic</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>productivity</td>
<td>limited &amp; variable</td>
<td>highly productive</td>
</tr>
<tr>
<td>type of meaning</td>
<td>often lexical</td>
<td>often purely grammatical</td>
</tr>
<tr>
<td>semantic regularity</td>
<td>often unpredictable</td>
<td>regular</td>
</tr>
<tr>
<td>restricted to specific</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>syntactic env.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>position</td>
<td>central</td>
<td>peripheral</td>
</tr>
<tr>
<td>portmanteau forms</td>
<td>rarely</td>
<td>often</td>
</tr>
<tr>
<td>repeatable?</td>
<td>sometimes</td>
<td>never</td>
</tr>
</tbody>
</table>

(Based on R. Sproat’s course notes and (Kroeger 2005:253))
Morphological processes

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

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

Often, there are phonological/graphemic 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:
  - Afrikaans: *amper* ‘nearly’ – *amper-amper* ‘very nearly’; *dik* ‘thick’ – *dik-dik* ‘very thick’
  - Indonesian: *orang* ‘man’ – *orang-orang* ‘all sorts of men’ (Cf. *orangutan*)
Morphological processes (cont.)

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

- **Reduplication** – part of the word or the entire word is doubled:
  - Afrikaans: *amper* ‘nearly’ – *amper-amper* ‘very nearly’; *dik* ‘thick’ – *dik-dik* ‘very thick’
  - Indonesian: *orang* ‘man’ – *orang-orang* ‘all sorts of men’ (Cf. *orangutan*)
  - Tagalog: *basa* ‘read’ – *ba-basa* ‘will read’; *sulat* ‘write’ – *su-sulat* ‘will write’
  - Samoan:
    - *alofa* ‘love<sub>Sg</sub>’
    - *galue* ‘work<sub>Sg</sub>’
    - *la:po?a* ‘to be large<sub>Sg</sub>’
    - *tamo?e* ‘run<sub>Sg</sub>’
    - *a-lo-lofa* ‘love<sub>Pl</sub>’
    - *ga-lu-lue* ‘work<sub>Pl</sub>’
    - *la:-po-po?a* ‘to be large<sub>Pl</sub>’
    - *ta-mo-mo?e* ‘run<sub>Pl</sub>’
  - English: *humpty-dumpty*
  - American English (borrowed from Yiddish): *baby-schmaby, pizza-schmizza*

Jirka Hana & Anna Feldman

Basics of Phonetics, Phonology & Morphology
Morphological processes (cont.)

- **Reduplication** – part of the word or the entire word is doubled:
  - Afrikaans: *amper* ‘nearly’ – *amper-amper* ‘very nearly’; *dik* ‘thick’ – *dik-dik* ‘very thick’
  - Indonesian: *oran* ‘man’ – *oran-oran* ‘all sorts of men’ (Cf. *orangutan*)
  - Tagalog: *basa* ‘read’ – *ba-basa* ‘will read’; *sulat* ‘write’ – *su-sulat* ‘will write’
  - Samoan:
    - *alofa* ‘love<sub>Sg</sub>’  
    - *galue* ‘work<sub>Sg</sub>’  
    - *la:po?a* ‘to be large<sub>Sg</sub>’  
    - *tamo?e* ‘run<sub>Sg</sub>’  
  - English: *humpty-dumpty*
  - American English (borrowed from Yiddish): *baby-schmaby*, *pizza-schmizzza*
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
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:

- lomad ‘learn\textsubscript{masc}’
- lamad ‘learned\textsubscript{masc.sg.3rd}’
- limed ‘taught\textsubscript{masc.sg.3rd}’
- lumad ‘was-taught\textsubscript{masc.sg.3rd}’
- shotek ‘be-quiet\textsubscript{pres.masc}’
- shatak ‘was-quiet\textsubscript{masc.sg.3rd}’
- shitek ‘made-sb-to-be-quiet\textsubscript{masc.sg.3rd}’
- shutak ‘was-made-to-be-quiet\textsubscript{masc.sg.3rd}’
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\textsubscript{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)
  - Czech: *kráva* ‘cow*<sub>nom</sub>*’ – *krav* ‘cows*<sub>gen</sub>*’,
    *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\textsubscript{imperf}’ \(\rightarrow\) \(hi\) ‘walking\textsubscript{perf}’
    - \(hihim\) ‘walking\textsubscript{pl.imperf}’ \(\rightarrow\) \(hihi\) ‘walking\textsubscript{pl.perf}’

- French feminine adjective \(\rightarrow\) masculine adj. (much less clear)
  - \(grande\) \[gr\textasciitilde Ad\] ‘big’ \(\rightarrow\) \(grand\) \[gr\textasciitilde A\] ‘big’
  - \(fausse\) \[fos\] ‘false’ \(\rightarrow\) \(faux\) \[fo\] ‘false’
Morphological processes (cont.)

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

  - Papago (a native American language in Arizona)
    - imperfective → perfective
    - \( \text{him} \) ‘walking}_{\text{imperf}} \rightarrow \text{hi} \) ‘walking}_{\text{perf}}
    - \( \text{hihim} \) ‘walking}_{\text{pl.imperf}} \rightarrow \text{hihi} \) ‘walking}_{\text{pl.perf}}

  - French
    - feminine adjective → masculine adj. (much less clear)
    - \( \text{grande} \) [grād] ‘big\(_f\)’ \rightarrow \text{grand} [grā] ‘big\(_m\)’
    - \( \text{fausse} \) [fos] ‘false\(_f\)’ \rightarrow \text{faux} [fo] ‘false\(_m\)’
Word formation: some examples

- **Affixation** – words are formed by adding affixes.
  - V + -able $\rightarrow$ Adj: *predict-able*
  - V + -er $\rightarrow$ N: *sing-er*
  - *un* + A $\rightarrow$ A: *un-productive*
  - A + -en $\rightarrow$ V: *deep-en, thick-en*
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
Word formation (cont.)

- **Acronyms** – like abbreviations, but acts as a normal word
  - *laser* – light amplification by simulated emission of radiation
  - *radar* – radio detecting and ranging
Word formation (cont.)

- **Acronyms** – like abbreviations, but acts as a normal word
  - *laser* – *light amplification by simulated emission of radiation*
  - *radar* – *radio detecting and ranging*
- **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* – light amplification by simulated emission of radiation
  - *radar* – radio detecting and ranging
- **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:
  
  (7) a. John wrote/has written a letter. (complete)
  
  b. John was writing a letter. (process)

- Russian: Aspect is marked mostly by prefixes:
  
  (8) a. John napisal pis’mo. (the action is complete)
  
  b. John pidal pis’mo. (process)
Morphological types of languages (cont.)

There are two basic morphological types of languages:

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

  (9) Vietnamese:

  khi tôi đến nhà bạn tôi, chúng tôi bắt đầu
  when I come house friend I, PLURAL I begin do
  làm bài
  homework

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

- **Synthetic** – both free and bound morphemes. Affixes are added to roots.
Morphological types of languages – synthetic

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:

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

‘house’
Morphological types of languages – synthetic

- **Fusional** – like agglutinating, but affixes tend to “fuse together”, one affix has more than one function.
  - E.g., Slavic, Romance languages, Greek
    - Serbian/Croatian: the number and case of nouns is expressed by one
      - suffix:
        - 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
        - local: ovc-o, ovc-e
        - instrumental: ovc-om, ovc-ama

      It is not possible to isolate separate singular or plural or nominative or accusative morphemes.
### Morphological types of languages – synthetic – fusional

#### Homonymy of the *a* ending in Czech:

<table>
<thead>
<tr>
<th>form</th>
<th>lemma</th>
<th>gloss</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>měst-a</td>
<td>město</td>
<td>town</td>
<td>noun neut sg gen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NP1 (5) noun neut pl nom (voc)</td>
</tr>
<tr>
<td>tém-a</td>
<td>téma</td>
<td>theme</td>
<td>noun neut sg nom (voc)</td>
</tr>
<tr>
<td>žen-a</td>
<td>žena</td>
<td>woman</td>
<td>noun fem sg nom</td>
</tr>
<tr>
<td>pán-a</td>
<td>pán</td>
<td>man</td>
<td>noun masc anim sg gen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>noun masc anim sg acc</td>
</tr>
<tr>
<td>ostrov-a</td>
<td>ostrov</td>
<td>island</td>
<td>noun masc inanim sg gen</td>
</tr>
<tr>
<td>předsed-a</td>
<td>předseda</td>
<td>president</td>
<td>noun masc anim sg nom</td>
</tr>
<tr>
<td>vidě-l-a</td>
<td>vidět</td>
<td>see</td>
<td>verb past fem sg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>verb past neut pl</td>
</tr>
<tr>
<td>vidě-n-a</td>
<td></td>
<td></td>
<td>verb passive fem sg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>verb passive neut pl</td>
</tr>
<tr>
<td>vid-a</td>
<td></td>
<td></td>
<td>verb transgressive masc sg</td>
</tr>
</tbody>
</table>

Jirka Hana & Anna Feldman

Basics of Phonetics, Phonology & Morphology
**Morphological types of languages – synthetic – fusional**

**Ending -e and noun cases in Czech:**

<table>
<thead>
<tr>
<th>case</th>
<th>form</th>
<th>lemma</th>
<th>gender</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>nom</td>
<td>kuř-e</td>
<td>kuře</td>
<td>neuter</td>
<td>chicken</td>
</tr>
<tr>
<td>gen</td>
<td>muž-e</td>
<td>muž</td>
<td>masc.anim.</td>
<td>man</td>
</tr>
<tr>
<td>dat</td>
<td>mouš-e</td>
<td>moucha</td>
<td>feminine</td>
<td>fly</td>
</tr>
<tr>
<td>acc</td>
<td>muž-e</td>
<td>muž</td>
<td>masc.anim.</td>
<td>man</td>
</tr>
<tr>
<td>voc</td>
<td>pan-e</td>
<td>pán</td>
<td>masc.anim.</td>
<td>mister</td>
</tr>
<tr>
<td>loc</td>
<td>mouš-e</td>
<td>moucha</td>
<td>feminine</td>
<td>fly</td>
</tr>
<tr>
<td>inst</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Morphological types of languages – synthetic

- **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.)
Morphological types of languages (cont.)

- 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)
- Czech (similarly other Slavic lgs) – 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*
- The distinction between analytic and (poly)synthetic languages is a continuum.
### Morphological types of languages (cont.)

<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>

**Table**: The degree of synthesis of some languages (Haspelmath 2002)
Complications – Cranberry morphemes

- What is the meaning of cran in cranberry?
Complications – Cranberry morphemes

- What is the meaning of *cran* in *cranberry*?
- Etymologically from *kraan* ‘crane’, but synchronically, there is no connection.
Complications – Cranberry morphemes

- What is the meaning of *cran* in *cranberry*?
- Etymologically from *kraan* ‘crane’, but synchronically, there is no connection.
- Similarly, *mul* exists only in *mulberry*.
Complications – Cranberry morphemes

- What is the meaning of cran in cranberry?
- Etymologically from kraan ‘crane’, but synchronically, there is no connection.
- Similarly, mul exists only in mulberry.
- boysenberry?
Complications – Cranberry morphemes

- What is the meaning of *cran* in *cranberry*?
- Etymologically from *kraan* ‘crane’, but synchronically, there is no connection.
- Similarly, *mul* exists only in *mulberry*.
- boysenberry? After Mr. Boysen who developed the hybrid of blackberry and raspberry.
- Similarly:
  - -ceive: conceive, receive, perceive, deceive
  - -mit: commit, permit, remit, submit, transmit, admit
  - inept (*ept), ungainly (*gainly), discern (*cern), nonplussed (*plussed), upholster (*downholster), outlandish (*landish) …
Complications – Cranberry morphemes

- What is the meaning of *cran* in *cranberry*?
- Etymologically from *kraan* ‘crane’, but synchronically, there is no connection.
- Similarly, *mul* exists only in *mulberry*.
- Boysenberry? After Mr. Boysen who developed the hybrid of blackberry and raspberry.
- Similarly:
  - -ceive: conceive, receive, perceive, deceive
  - -mit: commit, permit, remit, submit, transmit, admit
  - inept (*ept), ungainly (*gainly), discern (*cern), nonplussed (*plussed), upholster (*downholster), outlandish (*landish) . . .
- But: resurrection > resurrect, editor > edit
Complications – Cranberry morphemes

What is the meaning of *cran* in *cranberry*?
- Etymologically from *kraan* ‘crane’, but synchronically, there is no connection.
- Similarly, *mul* exists only in *mulberry*.
- boysenberry? After Mr. Boysen who developed the hybrid of blackberry and raspberry.

Similarly:
- -ceive: conceive, receive, perceive, deceive
- -mit: commit, permit, remit, submit, transmit, admit
- inept (*ept), ungainly (*gainly), discern (*cern), nonplussed (*plussed), upholster (*downholster), outlandish (*landish) . . .
- But: resurrection > resurrect, editor > edit
- cranapple juice
Complications – Zero morphemes

- Zero morpheme
  - Coptic:
    - d3o-i ‘my head’
    - d3o-k ‘your (masc.) head’
    - d3o ‘your (fem.) head’
    - d3o-f ‘his head’
    - d3o-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.
Complications – 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?
Complications – Some more

Breton diminutive plurals:

bag    boat    bagòù  boats
bagig  little boat  bagòuigòù  little boats
Complications – Some more

Breton diminutive plurals:
bag  boat  bagôu  boats
bagig  little boat  bagôuigôu  little boats

English:
*pick up – picker upper, tuck in – tucker-inner*
(notice the regular 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*
Clitics are units that are transitional between words and affixes, having some properties of words and some properties of affixes.
Clitics vs. Words

Unlike words, clitics:

- Placement of clitics is more restricted.
- Cannot stand in isolation.
- Cannot bear contrastive stress.
- etc.
Clitics vs. Affixes

Unlike affixes, clitics:
- Are less selective to which word (their host) they attach, e.g. host’s part-of-speech may play no role.
- Phonological processes that occur across morpheme boundary do not occur across host-clitic boundary.
- etc.
The exact mix of these properties varies considerably across languages.
The way clitics are spelled also varies within a single language:

- **written as affixes of their host**
  - English: *don’tcha*
  - Czech: *Cos* ‘what+refl’, *proň* ‘for him’

- **separated by punctuation**
  - English: possessive ‘s
  - Czech: *-li* ‘if’ (*Viděl-li auto . . . ‘If he saw a car . . .’)*

- **written as separate words**
  - English: *her*: *He sees her* [Hi siz hr] (word) vs. [Hi sizr] (clitic)
  - French: *le* ‘him’ vs. *lui* (*Je le vois ‘I see him’ vs. *Je lui vois*)