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

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

Phonological Rules

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

(1) a. /pɪt/ → [p^hɪt]
b. /spɪt/ → [spɪt]

Phonological Rules

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

(1) a. /pɪt/ → [p^hɪt]
b. /spɪt/ → [spɪt]

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

Phonological Rules

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

(1) a. /pɪt/ → [p^hɪt]
b. /spɪt/ → [spɪt]

(2) /p/ → [p^h] 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/ → [p^hɪt]
b. /spɪt/ → [spɪt]

(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

Phonological Rules

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

(1) a. /pɪt/ → [p^hɪt]
b. /spɪt/ → [spɪt]

(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] / # _

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, ʊ, o, ɔ]),
- nasals ([m, n, ŋ]),
- 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* [ɱf], *input* [mp]
  - voicing assimilation *talks* [ks] vs. *dogs* [gz]
  - *bit* [ɪ] vs. *bin* [ɪ̃] – /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

badu	'he goes away'	*inĝen*	'he went'
baddun	'I go away'	*inĝenen*	'I went'
bašidu	'he goes away to him'	*inšiĝen*	'he went to him'
bašiduun	'I go away to him'	*inšiĝenen*	'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).

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.

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

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.

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.

(5) 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*

Order of Morphemes

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

Bound × Free Morphemes (cont.)

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 × 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 × 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
  - 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): *basa* 'read' *b-um-asa* 'read$_{past}$'
  sulat 'write' *s-um-ulat* 'wrote'

  Khmer (Cambodia): *leun* 'fast' – *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$_{past}$'
  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$_{past}$'
  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 & Janda 1988:212).

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):
  baddang 'help', *ka-baddang-an* 'helpfulness', **ka-baddang*, **baddang-an*;

  Czech *po+. . . +í*:
  Vltava → *Po-vltav-í* 'Vltava river area' (**povltava*, **vltaví*);
  Pobaltí 'Baltics', *pohraničí* 'border region', *pobřeží* 'sea shore', *pohoří*, *potrubí*, *polesí*

Affixing

- Suffixing is more frequent than prefixing and far more frequent than infixing/circumfixing (Sapir 1921, Greenberg 1957, Hawkins & Gilligan 1988).

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 (see Hana & Culicover 2008 for an overview):
  - processing arguments (Cutler et al 1985, Hawkins & Gilligan 1988),
  - historical arguments (Givòn 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* (3^{rd} sg)

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 word
  bring, *brought*, *brings*, *bringing*
- **Derivation**: creates new words
  logic, *logical*, *illogical*, *illogicality*, *logician*, etc.

Inflection

- **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'

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

Complications with terminology

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 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.

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 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*;

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*
  - *sincere* > *sincerity*, *scarce* > *scarcity*, *curious* > *curiosity*; *fierce* but **fiercity*

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*
  - *sincere* > *sincerity*, *scarce* > *scarcity*, *curious* > *curiosity*; *fierce* but **fiercity*
- 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))

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: *oraŋ* 'man' – *oraŋ-oraŋ* '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: *oraŋ* 'man' – *oraŋ-oraŋ* '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: *oraŋ* 'man' – *oraŋ-oraŋ* 'all sorts of men' (Cf. *orangutan*)
  - Tagalog: *basa* 'read' – *ba-basa* 'will read'; *sulat* 'write' – *su-sulat* 'will write'
  - 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$_{Pl}$'

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: *oraŋ* 'man' – *oraŋ-oraŋ* 'all sorts of men' (Cf. *orangutan*)
  - Tagalog: *basa* 'read' – *ba-basa* 'will read'; *sulat* 'write' – *su-sulat* 'will write'
  - 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

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$_{pl.perf}$'

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*

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 **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 languages:

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

  (9) Vietnamese:

  khi tôi đên nhà bạn tôi, chúng tôi bat dăù
  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:

	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 – 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

		singular	plural	
	nominative	ovc-a	ovc-e	'*ovca* 'sheep'
	genitive	ovc-e	ovac-a	
suffix:	ovac+0 ?? 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:

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

Morphological types of languages – synthetic – fusional

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	–	–		

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.)

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)

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:

dʒo-i	'my head'
dʒo-k	'your (masc.) head'
dʒo	'your (fem.) head'
dʒo-f	'his head'
dʒo-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	bag**òu**	boats
bagig	little boat	bag**òu**ig**òu**	little boats

Complications – Some more

Breton diminutive plurals:

bag	boat	bag**òu**	boats
bagig	little boat	bag**òu**ig**òu**	little boats

English:
pick up – picker upper, *tuck in – tuck***er**-*inn***er**
(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

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.

Clitics (cont.)

The exact mix of these properties varies considerably across languages.

Clitics (cont.)

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*)