Computational Morphology: Introduction

Yulia Zinova

1 – 5 August 2016

Plan

1. 1 August – Introduction to theoretical and computational morphology, solving some morphological problems (pre-formally)
2. 2 August – Finite state automata and transducers: theory and paper practice
3. 3 August – xfst
4. 4 August – lexc
5. 5 August – practice xfst+lexc, finishing exercises from the previous days, discussing APs.

Requirements for BNs and APs

- Attendance:
  - I usually don't care about attendance itself, but as this is an intensive course, I think attendance is important;
  - attendance sheets will be passed twice a day;
  - if you are absent in some class you can expect that I will ask you some questions about the topic we discussed during that time when I check your exercises.

Requirements for BNs and APs

- For both BN and AP:
  - at the end of the class you should have solutions to all the exercises we have done during the class (together and on your own);
  - for each exercise that includes writing a script you should be able to explain what any line of the script means;
  - you should show general understanding of the material discussed in class.

For an AP: organizational

- Please bring the AP forms to sign within this week;
- you will have to describe a piece of morphology using one of the frameworks we will be working with;
- each student doing an AP should be describing a separate piece of morphology;
- the area covered by your program should be something that takes around 70 optimal rules;
- to find such a piece, go to the library and study the shelves with grammars of languages you don't know;
- you have to tell (show) me the material you want to work with and receive my approval (please do it within this week).

For an AP: results

- As a result of you work I expect to receive a script, a set of test examples (with the corresponding set of outputs), and a paper.
- The script has to work for all the cases described by the piece of morphology you aim to cover.
- Your set of test examples should be representative of the data you aim to cover, be sure to check that all the important cases are included and you are not testing exactly the same combination of rules multiple times (unless you provide an automated testing script that checks the output).
- In the paper you should describe the facts that you are modeling, the choices you had to make while writing the program (e.g., the ordering of rules and the selection of the formalism), the testing phase, and (optional) the material that you are aware of, but your program does not cover for good reasons.

AP – Grades

- The description part is worth 30 points, the script part – 60 points, the set of testing examples – 10 points;
    - 1.0: 95 – 100
    - 1.3: 91 – 94
    - 1.7: 87 – 90
    - 2.0: 83 – 86
    - 2.3: 80 – 82
    - 2.7: 75 – 79
    - 3.0: 70 – 74
    - 3.3: 65 – 69
    - 3.7: 60 – 65
    - 4.0: 50 – 59

Computational Morphology

- Theoretical knowledge of morphology
  - speaker's intuition
  - language grammar
- Programming skills
  - mastery of the tools
  - designing the program
  - problem solving (decomposition of complex rules)

Morphology

- Morphology: "study of shape" (Greek)
- Morphology in different fields:
  - Archaeology: study of the shapes or forms of artifacts;
  - Astronomy: study of the shape of astronomical objects such as nebulae, galaxies, or other extended objects;
  - Biology: the study of the form or shape of an organism or part thereof;
  - Folkloristics: the structure of narratives such as folk tales;
  - River morphology: the field of science dealing with changes of river platform;
  - Urban morphology: study of the form, structure, formation and transformation of human settlements;
  - Geomorphology: study of landforms

Morphology in linguistics

- The study of the internal structure and content of word forms;
- First linguists were studying morphology:
  - ancient Indian linguist Pānini formulated 3,959 rules of Sanskrit morphology in the text Astādhyāyī;
  - The Greco-Roman grammatical tradition was also engaged in morphological analysis.
  - Studies in Arabic morphology: Marāḥ al-arwāḥ and Ahmad b. ‘alī Mas‘ūd, end of XIII century;
  - Well-structured lists of morphological forms of Sumerian words: written on clay tablets from Ancient Mesopotamia; date from around 1600 BC.

An ancient example

- Well-structured lists of morphological forms of Sumerian words: written on clay tablets from Ancient Mesopotamia; date from around 1600 BC;

badu	‘he goes away’	ingĝen	‘he went’
baddun	‘I go away’	ingĝenen	‘I went’
bašidu	‘he goes away to him’	inšigĝen	‘he went to him’
bašiduun	‘I go away to him’	inšigĝenen	‘I went to him’

(see Jacobsen, 1974, 53-4)

Questions that morphological theory answers

- What is the past tense of the English verb *sing*?
- Do Greek nouns have dual formas?
- How are causative verbs formed in Finnish?
- What word form in Latin is *amavissent*?

Terminology

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

Terminology

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

Example

- The paradigm of the Latin lexeme insula ‘island’

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

Terminology: Complications

- The terminology is not universally accepted, for example:
  - lemma and lexeme are often used interchangeably (and so will we use it too);
  - sometimes lemma is used to denote all forms related by derivation;
  - paradigm can stand for the following:
    1. set of forms of one lexeme;
    2. a particular way of inflecting a class of lexemes (e.g. plural is formed by adding -s);
    3. a mixture of the previous two: set of forms of an arbitrarily chosen lexeme, showing the way a certain set of lexemes is inflected (language textbooks).

Morpheme

- Morphemes are the smallest meaningful constituents of words;
- e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme;
- words are composed of morphemes (one or more).

Morpheme

- Morphemes are the smallest meaningful constituents of words;
- e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme;
- words are composed of morphemes (one or more).
- Your examples?
  1. a word with 1 morpheme?

Morpheme

- Morphemes are the smallest meaningful constituents of words;
- e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme;
- words are composed of morphemes (one or more).
- Your examples?
  1. a word with 1 morpheme?
  2. 2 morphemes?

Morpheme

- Morphemes are the smallest meaningful constituents of words;
- e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme;
- words are composed of morphemes (one or more).
- Your examples?
  1. a word with 1 morpheme?
  2. 2 morphemes?
  3. 3 morphemes?

Morpheme

- Morphemes are the smallest meaningful constituents of words;
- e.g., in *books*, both the suffix -*s* and the root *book* represent a morpheme;
- words are composed of morphemes (one or more).
- Your examples?
  1. a word with 1 morpheme?
  2. 2 morphemes?
  3. 3 morphemes?
  4. 4 morphemes?

Morpheme

- Morphemes are the smallest meaningful constituents of words;
- e.g., in *books*, both the suffix *-s* and the root *book* represent a morpheme;
- words are composed of morphemes (one or more).
- Your examples?
  1. a word with 1 morpheme?
  2. 2 morphemes?
  3. 3 morphemes?
  4. 4 morphemes?
  5. 5 and more morphemes?

Morphs and allomorphs

- The term *morpheme* is used both to refer to an abstract entity and its concrete realization(s) in speech or writing.
- When there is a need to make a 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;
- Allomorphs have the same function but different forms. Unlike the synonyms they usually cannot be replaced one by the other.
- Examples?

Examples of allomorphs

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

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

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

Morphemes

- The order of morphemes/morphs matters:

(2) a. talk-ed ≠ *ed-talk
b. re-write ≠ *write-re
c. un-kind-ly ≠ *kind-un-ly

- Complications: how would you decompose *cranberry* into morphemes?

Morphemes

- The order of morphemes/morphs matters:

(2) a. talk-ed ≠ *ed-talk
b. re-write ≠ *write-re
c. un-kind-ly ≠ *kind-un-ly

- Complications: how would you decompose *cranberry* into morphemes?
- The *cran* is unrelated to the etymology of the word *cranberry* (crane (the bird) + berry).

(3) cranberry = crane + berry ≠ cran + berry

- Zero-morphemes, empty morphemes.

Types of morphemes: bound/free

- Bound morphemes cannot appear as a word by itself.
- Examples?

Types of morphemes: bound/free

- Bound morphemes cannot appear as a word by itself.
- Examples?
- -s (dog-s), -ly (quick-ly), -ed (walk-ed)
- Free morphemes can appear as a word by itself; often can combine with other morphemes too.
- Examples?

Types of morphemes: bound/free

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

Types of morphemes: bound/free

- The property of being bound or free is language-dependent: past tense morpheme is a bound morpheme in English (-ed) but a free morpheme in Mandarine Chinese (le)

(4) 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.'

Types of morphemes: content/functional

- **Content** morphemes carry some semantic content;
- **Functional** morphemes provide grammatical information;
- Examples?

Morphemes: Root

- Root is the nucleus of the word that affixes attach too.
- In English, most of the roots are free.
- In some languages that is less common: in Russian, noun and verbal roots are bound morphemes, sometimes with zero affixes;
- Some words (compounds) contain more than one root: *homework*.

Morphemes: Affixes (1)

- **Affix** is a morpheme that is not a root; it is always bound;
- **Suffix** follows the root;
- Suffixes in English: -ful in event-ful, talk-ing, quick-ly, neighbor-hood
- **Prefix** precedes the root;
- Prefixes in English: un- in unhappy, pre-existing, re-view;
- **Infix** occurs inside the root;
- Infixes in Khmer: -b- in lbeun ‘speed’ from leun ‘fast’;
- Infixes in Tagalog: -um- in s-um-ulat ‘write’

Morphemes: Affixes (2)

- **Circumfix** occurs on both sides of the root
- Circumfixes in Tuwali Ifugao: baddang ‘help’, ka-baddang-an ‘helpfulness’, *ka-baddang, *baddang-an;
- Circumfixes in Dutch:
  - berg ‘mountain’ – ge-berg-te ‘mountains’, *geberg, *bergte;
  - vogel ‘bird’, ge-vogel-te ‘poultry’, *gevogel, *vogelte

Typology of affixation

- Suffixing is more frequent than prefixing;
- Infixing/circumfixing are very rare (Sapir, 1921; Greenberg, 1957; Hawkins and Gilligan, 1988);
- Postpositional and head-final languages use suffixes and no prefixes;
- 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).

Typology of affixation

- Several attempts to explain the asymmetry between suffixing and prefixing (Hana and Culicover, 2008):
  - processing arguments (Cutler *et al.*, 1985; Hawkins and Gilligan, 1988)
  - historical arguments (Givón, 1979)
  - combinations of both (Hall, 1988)

Derivation and Inflection

Two different kinds of morphological relations among words:

- **Inflection**: creates new forms of the same lexeme.
  E.g., *bring, brought, brings, bringing* are inflected forms of the lexeme *bring*.
- **Derivation**: creates new lexemes E.g., *logic, logical, illogical, illogicality, logician*, etc. are derived from *logic*, but they all are different lexemes.
- Inflectional suffix is often called **ending**
- A word without its inflectional affixes (root + all derivational affixes) is called **stem**.

Derivation and Inflection

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

Derivation and Inflection: Properties (Kroeger, 2005)

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

Morphological processes: Concatenation

- Concatenations is adding continuous affixes, without splitting the stem
- The most common process
  hope+less, un+happy, anti+capital+ist+s
- Often, there are phonological changes on morpheme boundaries:
  book+s [s], shoe+s [z] happy+er → happi+er

Morphological processes: Reduplication

- Reduplication – part of the word or the entire word is doubled:
  - Tagalog: *basa* ‘read’ – *ba-basa* ‘will read’; *sulat* ‘write’ – *su-sulat* ‘will write’
  - Afrikaans: *amper* ‘nearly’ – *amper-amper* ‘very nearly’; *dik* ‘thick’ – *dik-dik* ‘very thick’
  - Indonesian: *oraŋ* ‘man’ – *oraŋ-oraŋ* ‘all sorts of men’
  - 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, hocus-pocus*
  - American English (borrowed from Yiddish): *pizza-schmizza*

Morphological processes: Templates

- Template morphology: both roots and affixes are discontinuous.
- Found in Semitic languages (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: Suppletion

- Suppletion: ‘irregular’ relation between the words
- English:
  be – am – is – was,
  go – went,
  good – better
- German?

Morphological processes: Ablaut

- Morpheme internal changes (apophony, ablaut): the word changes internally
- English: *sing* – *sang* – *sung*, *man* – *men*, *goose* – *geese* (not productive)
- German? Productivity?

Morphological processes: Substraction

- 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$_{pl.perf}$

- Another possible analysis for this example?

Word formation: Examples (1)

- 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
- 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: Examples (2)

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

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:

(5) a. John wrote (AE)/ has written a letter. (the action is complete)
b. John was writing a letter (process).

- Russian: Aspect is marked mostly by prefixes:

(6) a. Vasja napisal pis'mo. (the action is complete)
b. Vasja pisal pis'mo. (process)

Types of languages: analytic and synthetic

- Two basic morphological types of language structure: analytic and synthetic
- Analytic languages have only free morphemes, sentences are sequences of single-morpheme words (Vietnamese)
- Synthetic languages have both free and bound morphemes. Affixes are added to roots.

Subtypes of synthetic languages (1)

- Agglutinating languages: each morpheme has a single function, it is easy to separate them.
- Examples: Uralic languages (Estonian, Finnish, Hungarian), Turkish, Basque, Dravidian languages (Tamil, Kannada, Telugu), Esperanto
- Turkish (paradigm for 'house':

	singular	plural
nom.	ev	ev-ler
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

Subtypes of synthetic languages (2)

- Fusional languages: like agglutinating, but affixes tend to "fuse together", one affix has more than one function.
- Examples: Indo-European, Semitic, Sami
- Czech *matk-a* 'mother' – *-a* means the word is a noun, feminine, singular, nominative.
- Serbian/Croatian: the number and case of nouns is expressed by one suffix (paradigm for *ovca*'sheep'):

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

Subtypes of synthetic languages (3)

- Polysynthetic languages: 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.)

Types of languages: continuum

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

Degree of synthesis (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

Computational Morphology

- **Computational morphology** deals with developing techniques and theories for computational analysis and synthesis of word forms.
- Applications?

Computational Morphology

- **Computational morphology** deals with developing techniques and theories for computational analysis and synthesis of word forms.
- Applications?
- Spelling correction
- Search engines
- Machine translation
- Text generation
- Text-to-speech

Applications that do not belong to morphology

- **Tokenization**: split the input into words, punctuation marks, digit groups, etc. *Before* morphological analysis.
- **Part-of-speech (POS) tagging**: resolve ambiguities with respect to POS tagging. *After* morphological analysis.
- **Stemming/lemmatization**: find out the lemma of a word, but ignore the morphological tags. *Instead of* morphological analysis.

Basic morphological processing

- **Analysis**: given a word, find its form description.
- Form description is **lemma** followed by **tags**
- **Synthesis**: given a verb description, find the resulting string

word	lemma	tags
play	play	+N +Sg +Nom
	play	+V +Inf
plays	play	+N +Pl +Nom
	play	+V +IndPres3sg

Mathematical view on morphology

- Morphology is a relation `M` between words `W` and their form descriptions `D`:
  `M : P(W x D)`
- A morphological analyzer is a function
  `f : W → P(D) such that d : f(w) iff (w, d) : M`
- A morphological synthesizer is a function
  `g : D → P(W) such that w : g(w) iff (w, d) : M`

Finite-state morphology

- Common assumption: `M` is a **regular relation**.
- This implies that
  - `M` can be defined using **regular expressions**
  - word-description pairs in `M` can be recognized by a **finite-state automaton (transducer)**

Finite-state morphology

- In most computational systems M is finite.
- This holds if one assumes that
  - the language (at a given moment) has a finite number of words
  - each word has a finite number of forms
- A finite morphology M is trivially a regular relation

Formats for a finite morphology

- **Full-form lexicon:** list of all words with their descriptions
- **Morphological lexicon:** list of all lemmas and all their forms in canonical order

  ```
  play N: play, plays, play's, plays'
  player N: player, players, player's, players'
  ```
- It is easy to transform a morphological lexicon to a full-form lexicon

Analyzing with a full-form lexicon

- It is easy to compile a full-form lexicon into a **trie** – a **prefix tree**
- A trie has transitions for each symbol, and it can return a value (or several values) at any point.
- A trie is also a special case of a finite automaton - an **acyclic deterministic finite automaton**.

Models of morphological description (Hockett, 1954)

- **Item and arrangement**: inflection is concatenation of morphemes (stem + affixes).
  `dog +Pl → dog s → dogs`
- **Item and process**: inflection is application of rules to the stem (one rule per feature).
  `baby +Pl → baby(y → ie / _s) s → babie s → babies`
- **Word and paradigm**: inflection is association of a model inflection table to a stem
  `{Sg:fly, Pl:flies}(fly := baby) → {Sg:baby, Pl:babies}`

Paradigms, mathematically

- For each part of speech `C` (“word class”), associate a finite set `F(C)` of inflectional features.
- An **inflection table** for `C` is a function of type `F(C)` → `Str`.
- Type `Str`: lists of strings (some lists may be empty).
- A **paradigm** for `C` is a function of type `String` → `F(C)` → `Str`.
- Thus there are different paradigms for nouns, adjectives, verbs, etc.

Inflectional table: Example

- `F(N) = Number x Case`, where
  `Number = {Sg, Pl}, Case = {Nom, Gen}`
- The word dog has the inflection table (using GF notation)

```
table {
  <Sg,Nom> => "dog" ;
  <Sg,Gen> => "dog's" ;
  <Pl,Nom> => "dogs" ;
  <Pl,Gen> => "dogs'"
}
```

Paradigm: Example

- regN, the regular noun paradigm, is the function (of variable x)

```
\x → table {
  <Sg,Nom> => x ;
  <Sg,Gen> => x+ "’s" ;
  <Pl,Nom> => x+ "s" ;
  <Pl,Gen> => x+ "s’"
}
```

Example problem: consonant reduplication

(7) I am swimming

- There is a lexeme ‘to swim’
- The +ing portion tells us that this event is taking place at the time the utterance is referring to.
- Why there is an extra **m**?

Problem: zero mophemes

- 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'

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

Problem: empty mophemes

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

case	'bear'	'elephant'	(male name)
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

- 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.
- Alternative analysis?

References:

Cutler, A., Hawkins, J. A., and Gilligan, G. (1985). The suffixing preference: a processing explanation. *Linguistics*, **23**(5), 723–758.

Givón, T. (1979). *Discourse and syntax*. Academic Press New York.

Greenberg, J. H. (1957). The nature and uses of linguistic typologies. *International journal of American linguistics*, pages 68–77.

Hall, C. J. (1988). Integrating diachronic and processing principles in explaining the suffixing preference. *Explaining language universals*, pages 321–349.

Hana, J. and Culicover, P. W. (2008). Morphological complexity outside of universal grammar. *Ohio State dissertations in linguistics*, page 85.

Haspelmath, M. (2002). *Understanding Morphology*. Arnold Publishers.

Hawkins, J. A. and Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua*, **74**(2-3), 219–259.

Hockett, C. F. (1954). Two models of grammatical description. *Word*, **10**(2-3), 210–234.

Jacobsen, T. (1974). Very ancient texts: Babylonian grammatical texts. *Studies in the history of linguistics: traditions and paradigms*, page 41.

Kroeger, P. R. (2005). *Analyzing grammar: An introduction*. Cambridge University Press.

Sapir, E. (1921). An introduction to the study of speech. *Language*.