Introduction to Morphology

Linguistics for Computer Scientists
Session 4

Antske Fokkens

Department of Computational Linguistics
Saarland University

11 October 2007

Today's lecture

- What is morphology?
- Subdomains of Morphology
- Morphological Properties
- Morphological Processes
- Automata

Introduction to Morphology

1. A definition of Morphology
2. A simple model of language
3. Morphemes and Morphology, basic vocabulary
4. Types of morphemes
5. Subdomains of Morphology
6. Morphological properties

What is morphology?

Morphology is the study of form and structure.

In linguistics, it generally refers to the study of form and structure of words.

Words and morphemes

There are two main usages of the term *word*:

1. Surface form (spoken or written represenation)
2. Abstract form (lemma or dictionary entry,
   e.g. bare infinitives in English, nominative single form of nouns in Latin)

   The class of forms representing a word in different contexts is called a **lexeme**
   e.g. sing = {*sing, sings, sang, sung, singing*}

A definition of words?

Words can be described as units of language (either sequences of sounds, or signs) that function as meaning bearers. But this is a fuzzy notion, e.g.:

- *sang* expresses both “singing” and past tense.
- Is *more or less* one word, or are there three words?

A structuralist solution: **morphemes**

A language:

11-112 phonemes

4,000-10,000 morphemes

An infinite number of sentences

Morphemes and Morphological analysis

- **Morphemes**
  - Morphemes are minimal meaning-bearing units:
    e.g. *talked* contains two morphemes: *talk* and *-ed* (past).
  - Form-function pairs (sound/sign-meaning)
  - Basic units of morphology
  - The realisations of morphemes are called *morphs*:
    e.g. English plural morpheme:
    [NUMBER pl]: -s, -es, -en, -∅
    boy-s, box-es, ox-en, sheep
  - These different realisations of the same morpheme are called **allomorphs**.
- **Morphological analysis**
  - Segmentation of expressions into basic units (mostly starting from word-level).
  - Classification of these basic units according to function.

Types of morphemes

- **Free Morphemes**
  Free morphemes can occur independently. Free morphemes are common in both English and German.

  e.g. *boy*, *sing*
- **Bound Morphemes**
  Bound morphemes must be attached to another morpheme, and cannot be used independently.

  e.g. [NUMBER pl] *-s rightarrow boys*

  Typical bound morphemes are:
  - **affixes** (*boy+s*, *talk+ed*)
  - **clitics** (French: *je ne sais pas*, *je* and *ne* cannot occur without a verb)
  - **roots** (Spanish *habl-* needs an ending indicating person, number, mode, etc.)

Formatives and pseudo-morphemes

Morphemes are form-meaning pairs, but not all segmentable forms have an identifiable meaning:

- **Formatives** are forms without identifiable meaning

  e.g. Linking elements in German compounds:
  *Geburt+**s**+tag* (Birthday), *Schwan+**en**+hals* (swan neck).
- **Pseudo-morphemes** or **cranberry morphemes** are special cases of formatives.
  They are segmentable part of a complex word, but do not have an independent meaning:

  e.g.
  - ***cran**+berry*, ***rasp**+berry*
  - *re+**ceive***, *con+**ceive***

What is morphology? (follow up)

Morphology can refer to three different things

a. Description of the behaviour of morphemes and how they are combined.
b. Derivational, inflectional and compositional processes of word formation occurring in a specific language.
   e.g. “German has a richer morphology than English”
c. Description of such word formation processes.

Root, base and stem

- **Root**: an unanalysable form, expressing the basic lexical content of a word. Also defined as 'what is left of a complex form when all affixes are stripped'.
- **Stem**: consists of at least a root.
  It can contain (an) derivational affix(es).
  In inflectional morphology, *stem* is generally defined as the root + a thematic vowel.
- **Base**: a form to which an affix may be added. A base may be simplex (root) or complex (root + affixes).

Areas of morphology

We distinguish:

- **Word forming**:
  - Derivational morphology
  - Compounding
- **Inflection**

Derivational Morphology

- allows to build complex words by combining bound and free morphemes.
- Derivational operations are per definition optional, i.e. not required by syntactic criteria.
- They change
  a. semantics,
     e.g. [*clear*] $\rightarrow$ [*un*+[*clear*]] = unclear
  b. syntactic category,
     e.g. $[derive]_V \rightarrow [[[derive]_V + ation]_N + al]_{Adj}$ = derivational
  c. valency of a verb,
     e.g. [*qaw*] 'it breaks' $\rightarrow$ [*t*+[*qaw*]] 'he breaks it' (Havasupai)
  d. several from the above, e.g. $[understand]_V \rightarrow$ [[understand]$_V$+able] = understandable

Compounding

- allows to build complex words by juxtaposition of free morphemes.
  [[*sale*]+s+[*man*]], [[*dish*]+[*washer*]].
- Productive compounding results in an infinite lexicon.
  $$\left\{\begin{array}{l}\textit{English}\\ \textit{German}\\ \textit{Havasupai}\end{array}\right\}\left\{\begin{array}{l}\textit{phonetics}\\ \textit{phonology}\\ \textit{morphology}\end{array}\right\}\left\{\begin{array}{l}\textit{teacher}\\ \textit{researcher}\\ \textit{student}\end{array}\right\}$$
- Compounds are “referential islands”.

Inflectional Morphology

- Inflection is required by syntactic criteria, e.g. an English verb must have tense.
- It marks grammatical (=morphosyntactic) distinctions:
  - Conjugation (verbal categories):
    1. person, number, gender
    2. tense, aspect, mood, agreement
  - Declination (nominal categories)
    - case, number, gender, degree, definiteness
- Meaning or, at least, the general concept is (generally) not changed, though *when, who* or *what* and sometimes *where, how* and *whether* may be specified by inflectional morphemes.
- There are bound and free inflectional morphemes:
  go [TENSE past]: *went*
  go [TENSE future]: *will go*

Inflection — paradigm

Inflectional morphology is typically organised in **paradigms**.

Paradigm

"A set of forms having the same root/stem, one of which must be selected in a certain syntactic environment" (definition based on Crystal (1997:277) and Payne (1997: 26)

For instance, German conjugation:

present	NUMBER		*past*	NUMBER	
	singular	*plural*		*singular*	*plural*
1.	dehn-e	dehn-en	1.	dehn-te	dehn-te-n
2.	dehn-st	dehn-t	2.	dehn-te-st	dehn-te-t
3.	dehn-t	dehn-en	3.	dehn-te	dehn-te-n

Paradigm — An example

Latin declination of a noun of the first declination:

case	NUMBER	
	singular	*plural*
NOM	puella	puellae
GEN	puellae	puellarum
DAT	puellae	puellis
ACC	puellam	puellas
ABL	puella	puellis

Syncretism/exponence

We observe both:

- **syncretism**: the same form is used to express different feature combinations.
  Here: *-ae*: GEN or DAT singular, or NOM plural, *-a* NOM or ABL singular, *-is*: DAT or ABL plural.
- **exponence**: the relation between form and function is **m:n**:
  - **multi-exponence** (cumulation): one form expresses several functions.
    Here: *-am* expresses both accusative and singular
  - **Extended exponence**: in *ge-dehn-t*, *ge-* and *-t* express one function together.

Morphological Properties — Synthesis

Synthesis: the number of morphemes that tend to occur within a word.

- In **isolating** languages words tend to consist of only one morpheme. (e.g. Chinese languages)
- **Polysynthetic** languages are known for the large number of morphemes that may occur in a single word. For instance, the Quechua and Inuit languages. The following example is from Yup'ik:

(1) tuntussuqatarniksaitengqiggtuq
tuntu-ssur-qatar-ni-ksaite-ngqiggte-uq
reindeer-hunt-FUT-say-NEG-again-3gg-IND
'He had not yet said again that he was going to hunt reindeer'

(Payne, 1997:28)

Morphological Properties — Fusion

Fusion: the number of meaning units that are found in one morphological shape:

- **Agglutinative** languages have little fusion: each meaning component is represented by its own morpheme (e.g. Turkish).
- **Fusional** languages have morphemes that express many meaning units: e.g. *-ó* in Spanish *habló* expresses indicative mode, 3rd person, singular, past tense and perfect aspect.

In English, both examples of agglutinative morphemes, and fusional ones can be found:

- **agglutinative**: anti+dis+establish+ment+arian+ism
- **fusion**: vowel change in plural forming (*goose/geese*) and strong verbs (*sing/sang*).
  Individual morphemes (root and number/tense) cannot be segmented in chunks, therefore these forms are fusional.

Morphology in Computational Linguistics

Morphology related applications in computational linguistics are:

1. Analysing complex words, defining their component parts:

   anti+dis+establish+ment+arian+ism

2. Analysis of grammatical information, encoded in words:

   sings
   sing[PERSON 3, NUMBER singular,TENSE present]