Basic Linguistic Concepts

Alexander Fraser
`fraser@cis.uni-muenchen.de`

CIS, Ludwig-Maximilians-Universität München

Morphology
2017
2017-05-22

Announcements (1 of 2)

- Today we will cover the basic linguistics of morphology. This material is from H&S Chapter 2, please read this chapter after class

- The exercise this week is cancelled (as always, please check the web page regularly)

Outline

Slides adapted from Guillou (LMU); Weller and Haselbach (IMS Stuttgart)

Originally based on: Haspelmath, M. & Sims, A. D. (2010): *Understanding Morphology* [2nd ed.], chapter 2 'Basic concepts', London: Hodder Education.

Outline

1. Introduction

2. Morphological relationships

3. Morphological building blocks

4. Allomorphy

Introduction
Segmentation

- Many words can be easily **segmented**,
  i.e. broken up into individually meaningful parts

  - *read   read-s   read-er   read-able*
    wash   wash-es   wash-er   wash-able
    write   write-s   writ-er   writ-able

  - *kind   kind-ness   un-kind*
    happy   happi-ness   un-happy
    friend-ly   friend-li-ness   un-friend-ly

- These meaningful parts are called **morphemes**

- Morphemes are the ultimate elements of morphological analysis;
  they are, so to speak, morphological atoms

Introduction
Morphemes

- **Morphemes** can be defined as
  the smallest meaningful constituents of a linguistic expression

- Example:

  Camilla met an unfriendly chameleon.

- Possible segmentations:
  - syntactic segmentation:
    Camilla | met | an | unfriendly | chameleon.
  - syntactic and morphological segmentation:
    Camilla | met | an | un|friend|ly | chameleon.

- Impossible segmentation:
  - *Camilla | met | an | un|friend|ly | *cha|meleon.*

    Neither *cha* or *meleon* are meaningful in isolation, nor do they share
    any aspect of meaning in other contexts, e.g. **cha|risma*

Introduction
What is a word?

- Contiguous sequence of letters (i.e. sounds).

- How many words are in this sentence?

  Our plan was to meet there in a few days' time,
  once our projects in Paris were concluded.

  (adapted from BNC)

- It depends on how you count.
  - 18 sequences of letters separated by blank spaces: **word tokens**
  - 16 different seq. of letters separated by blank spaces: **word types**
    (*our* and *in* occur twice)
  - 15 different 'dictionary words': **lexemes**
    ($\{was, were\} \rightarrow$ BE [verb lexeme])

Introduction
Contiguity of letters as a criterion for wordhood?

- In agglutinative languages, e.g. Turkish,
  'words' can correspond to sentences rather than to lexemes

    Ev-de-ydi-m.
    home-LOC-PAST-1SG
    'I was at home.'

  Agglutination: complex words are formed by stringing together
  morphemes without changing their spelling or phonetics

Outline

Morphological relationships
Lexeme

- A **lexeme** is a word in an abstract sense

- For example, the lexeme LIVE represents the core meaning shared by forms such as *live*, *lives*, *lived*, *living*

- In most languages, dictionaries are organised according to lexemes (cf. 'dictionary word')

Morphological relationships
Word-form

- A **word-form** is a word in a concrete sense. It is the combination of a lexeme and a set of <u>grammatical meanings</u> (or grammatical functions)

- For example the lexeme LIVE together with the grammatical meanings "third person, singular, present tense" yields the word-form *lives*

- Word-forms belonging to the same lexeme express different grammatical meanings, but the same core (semantic) concept

- **Test:** List some word-forms for the lexeme: READ

Morphological relationships
Paradigm

- The set of word-forms that belong to a lexeme is often called a **paradigm**
- Paradigm of the Modern Greek noun lexeme FILOS ('friend'):

	singular	plural
nominative	*fílos*	*fíli*
accusative	*fílo*	*fílus*
genitive	*fílu*	*fílon*

- Paradigm of the Latin noun INSULA ('island'); (note: 7 different sequences of sound, but 10 word-forms)

	singular	plural
nominative	*insula*	*insulae*
accusative	*insulam*	*insulās*
genitive	*insulae*	*insulārum*
dative	*insulae*	*insulīs*
ablative	*insulā*	*insulīs*

Morphological relationships
Word family

- Different lexemes may also be related to each other. A set of related lexemes is sometimes called a **word family**

- Two English word families:
  - READ, READABLE, UNREADABLE, READER, READABILITY, REREAD
  - LOGIC, LOGICIAN, LOGICAL, ILLOGICAL, ILLOGICALITY

- **Test**: suggest a set of lexemes that constitute a word family

Morphological relationships
Word family

- Each member of a word family is given its own dictionary entry
  - Complex lexemes generally denote new concepts that are different from the concepts of the corresponding simple lexemes
    (e.g. *read* denotes activity, *reader* denotes individual)
  - Complex lexemes are normally less predictable then word-forms
    (e.g. a specialist in logic is a *logician* rather than a **logicist*)

- In a word family there may be different parts of speech (V, N, ...), this cannot be the case within a paradigm

Morphological relationships
Inflection and derivation

- **Inflection** (= inflectional morphology):
  The relationship between word-forms of a lexeme

- A lexeme <u>inflects for</u> (or: <u>is inflected for</u>) grammatical features,
  e.g. the Latin lexeme INSULA inflects for case and number

- **Derivation** (= derivational morphology):
  The relationship between lexemes of a word family

- A lexeme can <u>derive from</u> (or: can <u>be derived from</u>) another lexeme,
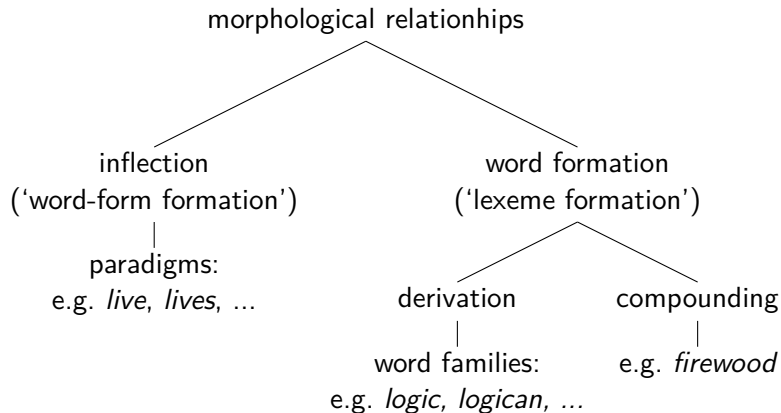 e.g. the lexeme READER is derived from the lexeme READ

Morphological relationships
Compounding

- Some morphologically complex words belong to two or more word families simultaneously

- For example, the lexeme FIREWOOD belongs both in the word family of FIRE and in the word family of WOOD

- Such relationships are called **compounding** and the resulting lexemes are called **compound lexemes** or just **compounds**

Morphological relationships
Subdivisions of Morphology

morphological relationhips

inflection
('word-form formation')

paradigms:
e.g. *live*, *lives*, ...

word formation
('lexeme formation')

derivation

word families:
e.g. *logic*, *logican*, ...

compounding

e.g. *firewood*

Outline

Morphological building blocks
Abstractness of meaning of morphemes

- In both inflection and derivation, morphemes have various kinds of meanings

- Some meanings are very **concrete** and can be described easily, e.g. the meanings of the morphemes *wash*, *logic*, *chameleon*, *un-*

- Other meanings are **abstract** and more difficult to describe, e.g. *-ity* in *readabil-ity* can be said to mean 'the *quality* of being readable'

- Some meanings are so abstract that they can hardy be called meanings, e.g. *-s* in English *read-s* is needed when the subject is a third person singular noun phrase. However it is difficult to say what the meaning of *-s* is. Instead we say that these kinds of morphemes have certain *grammatical functions*

Morphological building blocks
Affix and base

- Word-forms in an inflectional paradigm generally share (at least) one longer morpheme with a concrete meaning

- An **affix** attaches to a word or a main part of a word, i.e. to the **base**. The affix usually has an abstract meaning and cannot occur by itself.

- Russian case inflection (sg. forms, *ruk-* 'hand')

nominative	*ruk-a*
accusative	*ruk-u*
genitive	*ruk-i*
dative	*ruk-e*
locative	*ruk-e*
instrumental	*ruk-oj*

  $\rightarrow$ affix indicates case

Morphological building blocks
Affix and base

- Classical Nahuatl possessor inflection

1st sg.	*no-cal*	'my house'
2nd sg.	*mo-cal*	'your (SG) house'
3rd sg.	*i-cal*	'her/his house'
1st pl.	*to-cal*	'our house'
2nd pl.	*amo-cal*	'your (PL) house'
3rd pl.	*in-cal*	'their house'

$\rightarrow$ affix indicates possessor

Morphological building blocks
Types of affixes

- Affixes can be characterised by their position within the word
- Types of affixes

Suffix	follows the base	Russian -a in ruk-a 'hand'
		English -ful in event-ful
Prefix	precedes the base	Classical Nahuatl no-
		in no-cal 'my house'
		English un- in un-happy
Infix	occurs inside the base	Arabic -t- in (i)š-t-aġala
		'be occupied' (base: šaġala)
		Tagalog -um- in s-um-ulat
		'write' (base: sulat)
		Also: English swear words
Circumfix	occurs on both sides	German ge-...-t in ge-mach-t
	of the base	'made' (base: mach)

Morphological building blocks
Test

- Suggest some affixes to attach to the base "like"
- Give positive and negative examples

Morphological building blocks
Root

- The **base** is a relative notion that is defined wrt. the notion **affix**

- A base that cannot be analyzed any further into constituent morphemes is called a **root**
    - E.g. Readable → read (root and base)
    - E.g. Readability → read (root); readable (base)

Morphological building blocks
Stem

- A base is also sometimes called a **stem**, especially in inflection

- A base may or may not be able to function as a word-form
  - Bases that can also function as word-forms are called **free stems**, e.g. *cat* is both the base of the inflected word-form *cats* and itself a word-form
  - Bases that cannot also function as word-forms are called **bound stems**, e.g. the Italian word-form *gatti* can be broken up into the plural suffix *-i* and the base *gatt-*, but both aren't word-forms by themselves. Italian nouns inflect must for number, even in the singular (e.g. *gatt-o* 'cat', *gatt-i* 'cats').

Morphological building blocks
Borderline cases

- Roots and affixes can generally be distinguished quite easily, but sometimes there are problems

- For example, the morphemes *bio-* and *-crat* could be regarded as affixes because they do not occur as independent lexemes, but their very concrete meanings suggest that they should be regarded as bound stems (i.e. roots) that only occur in compounds
    - *biogeography*, *bioethics*, *bioengineering*, *biorhythm*, *bioterrorism*, *biomedicine*, *biochip*, etc.
    - *artistocrat*, *autocrat*, *democrat*, *Eurocrat*, *plutocrat*, *technocrat*, *theocrat*, etc.

Outline

Allomorphy
Allomorphs

- Morphemes may have different phonological shapes under different circumstances: **allomorph**
- For example, the plural morpheme/affix in English -s has various pronunciations:
  - [s] as in *cats* [kæts]
  - [z] as in *dogs* [dɒgz]
  - [əz] as in *faces* [feisəz]
- Roots can also have various shapes:
  - Rule: [iː] → [ɛ] (present-tense → past-tense form)
  - Instances: *sleep*/*slep-t*, *keep*/*kep-t*, *feel*/*fel-t*, *mean*/*mean-t*
- Allomorphs of one morpheme occur in different environments in **complementary distribution**. E.g. indefinite articles *a* and *an*:
  - an aardvark / * an bear
  - * a aardvark / a bear

Allomorphy
Allomorphs in German

- Plural in German nouns is also characterized by allomorphy
    - For instance: "-s" (Auto-s), "-er" (Kind-er), "-n" (Tante-n)
- For some German nouns, multiple allomorphs are OK:
    - "Fazit-e" / "Fazit-s" (conclusion); "Komma-s" / "Komma-ta"
    - In these cases the allomorphs:
      "-e", "-s" und "-ta"
      are called "freely varying allomorphs"
    - Consider also: German genitive singular (des) "Recht-s" / "Recht-es"
      (... of law).

Allomorphy
Examples of allomorphy in affixes

- Korean accusative suffix (marker of direct object): two allomorphs

-ul	ton	'money'	ton-ul	'money-ACC'
	chayk	'book'	chayk-ul	'book-ACC'
-lul	tali	'leg'	tali-lul	'leg-ACC'
	sakwa	'apple'	sakwa-lul	'apple-ACC'

(-ul after a consonant; -lul after a vowel)

Allomorphy
Examples of allomorphy in affixes

- Turkish first person possessive suffix: five allomorphs

-im	ev	'house'	ev-im	'my house'
	dil	'language'	dil-im	'my language'
-üm	köy	'village'	köy-üm	'my village'
	gün	'day'	gün-üm	'my day'
-um	yol	'way'	yol-um	'my way'
	tuz	'salt'	tuz-um	'my salt'
-ım	ad	'name'	ad-ım	'my name'
	kız	'girl'	kız-ım	'my daughter'
-m	baba	'father'	baba-m	'my father'

Allomorphy
Examples of allomorphy in roots

- German: a voiced obstruent (speech sound formed by obstructing air flow [consonants]) becomes voiceless (does not apply) in syllable-final position

Tag	[taːk]	'day'	*Tage*	[taːgə]	'days'
Hund	[hʊnt]	'dog'	*Hunde*	[hʊndə]	'dogs'
Los	[loːs]	'lot'	*Lose*	[loːze]	'lots'

Allomorphy
Examples of allomorphy in roots

- Russian: when the stem is followed by a vowel-initial suffix,
  the vowel *o/e* is often dropped if it is the last vowel in the stem

zamok	'castle'	*zamk-i*	'castles'
kamen'	'stone'	*kamn-i*	'stones'
nemec	'German'	*nemc-y*	'Germans'
nogot'	'nail'	*nogt-i*	'nails'

Allomorphy
Phonological allomorphy

- Allomorphs that are phonologically similar are **phonological allomorphs**: small differences in the shapes of morphemes that can be regarded as mere differences in pronunciation

- Formal relation between two or more phonological allomorphs: **alternation**

- Each phonological allomorph is called an **alternant**

Allomorphy
Morphophonological rules

- Alternations described using **morphophonological rules**:
  A **morphophonological rule** can manipulate
  an **underlying representation** under certain conditions and yields
  a **surface representation** (i.e. what is pronounced)

- E.g. Russian: when the stem is followed by a vowel-initial suffix,
  the vowel *o/e* is often dropped if it is the last vowel in the stem

Morphophonological rule:			
"*o/e* in the final stem syllable disappears when the stem is followed by a vowel-initial suffix"			
underlying:	[zamok] 'castle-SG'	underlying:	[zamok-i] 'castle-PL'
application:	no	application:	yes ([zamok-i] $\rightarrow$ [zamk-i])
surface:	[zamok] 'castle-SG'	surface:	[zamk-i] 'castle-PL'

Allomorphy
Suppletion

- **Suppletion**: the use of one word as the inflected form of another word when the two words are not cognate (i.e. share different roots)
- Morphemes may also have allomorphs that are not similar (i.e. irregular) in pronunciation: **suppletive allomorphs**
- **Strong suppletion**: allomorphs exhibit no similarity at all

go		wen-t		English
good		bett-er		
čelovek	'human being (person)'	ljud-i	'people'	Russian

- **Weak suppletion**: allomorphs exhibit some similarity, but this cannot be described by phonological rules

buy	[baɪ]	bough-t	[bɔːt]
catch	[kætʃ]	caugh-t	[kɔːt]
teach	[tiːtʃ]	taugh-t	[tɔːt]

Allomorphy
Types of allomorphy

Phonological allomorphy	Alternation could be described by a rule of pronunciation	English plural [s], [z], [əz]
Weak suppletive allomorphy	Allomorphs ehibit some similarity, but this cannot be described by phonological rules	English *buy*/*bough-t*, *catch*/*caugh-t*, etc.
Strong suppletive allomorphy	Allomorphs exhibit no similarity at all	English *good*/*bett-er*, *go*/*wen-t*, etc.

- Note that it is often hard to distinguish between weak suppletive allomorphy and phonological allomorphy

Allomorphy
Phonological conditioning

- Phonological allomorphs typically have **phonological conditioning**: the phonological context determines the choice of allomorph

- Example: English plural *-s*
  - [əz] after a sibilant (i.e. [s], [z], [ʃ], [ʒ], [tʃ], or [dʒ]), e.g. *face-s*, *maze-s*, *bush-es*, *garage-s*, *church-es*, *badge-s*
  - [s] after a voiceless non-sibilant obstruent, e.g. *cat-s*, *book-s*, *cliff-s*
  - [z] elsewhere, e.g. *bag-s*, *bell-s*, *key-s*

Allomorphy
Morphological conditioning

- Stem suppletion usually has **morphological conditioning**:
  the morphological context (usually, grammatical function) determines
  the choice of allomorph

- Example: Spanish verb *ir* 'go'
  - *ir* in the infinitive and future tense
  - *va-* in the present and imperfective past tense
  - *fu-* in the perfective past tense

Allomorphy
Lexical conditioning

- Sometimes the choice of a suppletive affix allomorph is dependent on other properties of the base (e.g. semantic properties): **lexical conditioning**

- Persian plural marking: human nouns -*an*, non-human nouns -*ha* (i.e. a semantic property)

-*an*	*mœrd*	'man'	*mœrd-an*	'men'
	geda	'beggar'	*geday-an*	'beggars'
-*ha*	*gorbe*	'cat'	*gorbe-ha*	'cats'
	ettefaq	'incident'	*ettefaq-ha*	'incidents'

- English past participle suffix -*en*
  (speakers must simply learn which verbs take -*en* and not the more common suffix -*ed*)
  eat/*eat-en*, *write*/*writt-en*, *ride*/*ridd-en*

Allomorphy
Types of conditioning

Phonological conditioning	Choice of allomorphs depends on phonological context	English plural depends on final sound in stem (e.g. cat-s [s] / face-s [əz])
Morphological conditioning	Choice of allomorphs depends on the morphological context	Spanish *ir*, *va-*, or *fu-*, depending on tense
Lexical conditioning	Choice of allomorphs depends on the individual lexical item	English past participle *-en*/*-ed* is unpredictable and depends on individual verbs

Summary

- **Morphemes**: smallest meaningful constituents
- **Lexeme**: dictionary words (is, was, were $\rightarrow$ BE)
- **Word-form**: lexeme + grammatical meanings
    - e.g. LIVE + "third person, singular, present tense" = lives
- **Paradigm**: set of word-forms belonging to a lexeme
- Morphological relationships:
    - **Inflection**: lexeme is inflected for grammatical features, e.g. INSULA + num/gen
    - **Derivation**: e.g. lexeme READER is derived from the lexeme READ
    - **Compounding**: e.g. FIRE + WOOD = FIREWOOD
- **Affixes** attach to **base**: e.g. read-able; A **Root** is a **base** with no affixes
- **Allomorph**: Morpheme with different phonological shapes under different circumstances. E.g. dog takes [z] / cat takes [s]

Thank you for your attention