

# HLT

## The Nature of Morphology

University of Malta

# Acknowledgements

- Richard Sproat, Morphology and Computation, MIT Press, ISBN 0-262-19314-0 (1992)
- Shuly Wintner, Lecture Notes, 2008

- 1 What is Morphology
- 2 Language Types
- 3 Morphological Function
- 4 Morphological Processes
- 5 Morphotactics: the order of morphemes
- 6 Orthography versus Phonology

# Outline

- 1 What is Morphology
- 2 Language Types
- 3 Morphological Function
- 4 Morphological Processes
- 5 Morphotactics: the order of morphemes
- 6 Orthography versus Phonology

# Outline

- 1 What is Morphology
- 2 Language Types
- 3 Morphological Function
- 4 Morphological Processes
- 5 Morphotactics: the order of morphemes
- 6 Orthography versus Phonology

# Outline

- 1 What is Morphology
- 2 Language Types
- 3 Morphological Function
- 4 Morphological Processes
- 5 Morphotactics: the order of morphemes
- 6 Orthography versus Phonology

# Outline

- 1 What is Morphology
- 2 Language Types
- 3 Morphological Function
- 4 Morphological Processes
- 5 Morphotactics: the order of morphemes
- 6 Orthography versus Phonology

# Outline

- 1 What is Morphology
- 2 Language Types
- 3 Morphological Function
- 4 Morphological Processes
- 5 Morphotactics: the order of morphemes
- 6 Orthography versus Phonology



## Example

<i>Nadie</i>	<i>puede</i>	<i>estar</i>	<i>contento</i>
Nobody	can	be	happy

- 1 As a word, *puede* is not atomic, but has internal structure.
- 2 *puede* is the **third person singular present indicative** of the verb *poder* (to be able).
- 3 The form *puede* is formed by **concatenating** the (second conjugation) suffix *-e* to the stem for the verb, *pued-*.
- 4 In the form *puede* **the stem vowel has changed** from *o* to *ue*

The change in (3) is **obligatory**, in the sense that *pode* is not well-formed.

- What are the smaller meaningful units from which words are built and what can they look like?
- What kinds of information do morphological forms convey?
- How are complex words assembled from such units?
- What are the constraints on the order of morphemes within words?
- What are the connections between constraints imposed by morphology (word structure) and phonology (sound structure) and orthography (writing system)?

# What is Morphology?

- **Morphology** is the area of linguistics which studies the structure of words.
- The structure varies according to your definition of a word.
  - A word is a sequence of graphemes (orthography)
  - A word is a sequence of phonemes (phonology)
  - A word is a sequence of morphemes (morphology)
- Rules govern the structures under each of these definitions.
- Sometimes phenomena in one domain affect behaviours in others.

# What is a Morpheme?

- Definition: smallest linguistic unit that has an *independent* meaning or grammatical function
  - **free morpheme**: morphemes that can stand alone as words.  
e.g. clock, sick
  - **bound morpheme**: morphemes that always attach to other morphemes, never existing as words themselves  
e.g. -ly, non-

# Words: Parts of Speech

- Words are traditionally classified into categories, known as parts of speech (POS) or word classes.
- The major parts of speech are noun, verb, adjective, adverb, pronoun, preposition, conjunction, article etc.
- Subcategorization of some major POS e.g. nouns
  - proper noun
  - common noun
- Open classes and closed classes
  - Content words vs. function words
- POS can shed light on the context in which a word can occur, its neighbours and even its pronunciation
  - “òbject” (noun) vs “objèct” (verb)
- The problematic nature of this classification
- POS tagging

# Why is Morphology Useful?

Almost all natural language applications require some processing of words.

- Dictionary tools
- Information Retrieval
- Spellchecking
- Machine Translation
- Speech Recognition
- Text-to-speech synthesis:
  - Word pronunciation often depends upon structure:
  - *Staatsprotokoll* would be pronounced with /sh/ for the underlined <s> if it were part of the same syllable as the following.
  - But it isn't because it is a separate morpheme which which happens to link two halves of a compound: *Staat* (state) and *Protokoll* (protocol).

# Four Broad Types of Language (Bloomfeld 1933)

- Isolating : no bound forms. Example: Mandarin Chinese
- Agglutinative : bound forms occur and are arranged in the word like beads on a string. Example: Turkish
- Polysynthetic : elements that often occur as separate words in other languages (such as arguments of the verb) are expressed morphologically. Example: Yupik (central Alaska)
- Inflectional : distinct features are merged into a single bound form. Example: Latin

# Isolating Languages

- Isolating languages have no *bound* forms such as affixes that must be attached to words
- Mandarin Chinese lacks morphological marking for tense and number

## Example

<i>gou</i>	<i>bu</i>	<i>ai</i>	<i>chi</i>	<i>qingcai</i>
dog	not	like	eat	vegetables

- the dog doesn't like to eat vegetables
- the dog didn't like to eat vegetables
- the dogs don't like to eat vegetables
- the dogs didn't like to eat vegetables
- dogs doesn't like to eat vegetables



# Agglutinative Languages

## Turkish

- Not only is Turkish morphology exclusively concatenative, but in addition, all affixes are suffixes
- Turkish words are of the form `stem affix*`
- *copluklerimizdekiledenmiydi*

### Example

<i>cop</i>	<i>luk</i>	<i>ler</i>	<i>imiz</i>	<i>de</i>	<i>ki</i>	<i>ler</i>	<i>den</i>	<i>mi</i>	<i>y</i>	<i>di</i>
garbage	aff	pl	1ppl	loc	rel	pl	abl	int	aux	past

- Was it from those that were in our garbage cans?

- The grammar is in the morphology
- Morphology encodes units that are usually considered syntactic (as in noun incorporation). Example: Central Alaskan Yupik
- *qaya:liyu:lu:ni* : He was excellent at making kayaks
- The verb *li* (make) and its object *qaya* (kayak) are incorporated into the word

- The archetypal inflectional language is Latin, just like our original Spanish example.
- For example, the suffix -o in *amo* (I love) expresses several grammatical categories: first person, singular, present, indicative, active
- the -o is called a *portmanteau* morpheme

# Two Different Kinds of Morphology

Morphology can be broadly divided into two different classes:

- **Inflectional Morphology**

- Input: a word
- Output: a different form of the *same word*
- Example: *wasal* → *waslu*

- **Derivational Morphology**

- Input: a word
- Output: a *different word* that is *derived* the input word.
- Example: *important* → *unimportant*

- **issue**: What is meant by a different word?

# Differences between Inflectional and Derivational Morphology

- Inflectional Morphology
  - Does not change part of speech.
  - Particular inflections may be required in particular syntactic contexts such as subject and object positions:  
“I like her” versus “me like she”.
  - Productive: tends to apply across the board - especially to new words e.g. “I was googling all day”.
  - Predictable semantics: +s applies to most nouns and almost always means plural.
- Derivational Morphology
  - Changes part of speech
  - Syntactic context never requires particular derivation.
  - Not very productive: “dislike”, “dishate\*"
  - Unpredictable semantic effect: “business” versus “happiness”.

# What sort of thing can morphology mark in different languages?

The answer depends to some extent on the part of speech concerned.

- Verb Morphology
- Noun Morphology
- Adjective Morphology

And also to whether we are talking about derivational or inflectional morphology

# Verb Morphology: Agreement

## Inflectional

- In many languages the verb must agree on person, number, gender or other features with one or more of its arguments.
- *Les filles sont arrivées* - the girls arrived.

### Example

arriv	+	é	+	e	+	s
arrive	+	past	+	fem	+	plur

# Verb Morphology: Tense and Aspect

## Inflectional

- Verbs are commonly marked with indications of the time at which the situations denoted by them occurred (tense):
  - *laud + a + b + o* - he will praise
  - *laud + a + v + it* - he has praised
- or the state of completion of the action denoted (aspect).
- This is the distinction we observe in
  - I waited for three hours (perfective aspect)
  - I have been waiting for three hours (imperfective aspect)
- Again in English the distinction is not marked entirely by morphology (as part of the word) but by a syntactic construction that goes beyond the word.
- In Chinese it is e.g. *tashui+le* = he sleep + PERFECTIVE



# Verb Morphology: Voice, Transitivity

## Inflectional and Derivational

- Transitivity: the number of arguments a verb takes.
- Voice: Active vs. Passive
  - Active: John brings the book down
  - Passive: The book is brought down by John
- In English active/passive is marked *syntactically* i.e at sentence level.
- In many other languages, it is marked *morphologically* i.e. at word level
- Maltese example derived from *nizel* (to descend)
  - *nizel* (he descended) → *nizzel* (he brought down)
  - *nizzel* (he brought down) → *tnizzel* (be brought down)

Inflectional categories for nouns (and pronouns) include

- Number (singular, plural, dual)  
fortizz+a, fortizz+i, għajn+ejn
- Case (marking various kinds of syntactic/semantic functions)
  - Nominative, accusative, as seen in
    - English pronouns: he, him
    - Latin nouns tabul+a, tabul+a+m
  - German has four cases: nominative, genitive, dative, accusative but only genitive is marked on the noun.
  - Latin has six
  - Finnish has fourteen!
- Gender (feminine, masculine, neuter)
  - In some languages, such as the Bantu languages, more detailed gender classes exist.

- In many languages adjectives, nouns, pronouns and verbs must “agree”
- Agreement features include number, gender, case and possibly other features.
- Example: Swahili has inflection affixes for humans, thin objects, paired things, instruments and extended body parts, inter alia.

- Many languages express *comparison of adjectives* morphologically.
- English
  - Hard (unmarked - stem)
  - Harder (comparative + er)
  - Hardest (superlative + est)
  - Sometimes the e is omitted.
- In English comparison is sometimes expressed with syntax.
  - “more different” not “differenter”

# Other Examples of Derivations

- Recall that derivations are category-changing
- Typical examples involving nouns, adjectives and verbs
  - Nominalization (verb → noun):
    - destroy → destruc+tion
    - catch → catch+er
  - Deverbal adjectives (verb → adjective)
    - The English suffix -able attaches to transitive verbs *x* and means *able to be x'ed*
    - Example drink → drink+able;
  - Nominalized adjective (noun → adjective)
    - The English suffix -less attaches to nouns *x* and means *something not possessing x*
    - Example brain → brain+less;

- Linear Concatenation, where a morphologically complex word can be analyzed as a series of morphemes concatenated together as with prefixes and suffixes: en + large + ment.
- Non-Linear Concatenation
  - infix
    - Bontoc (Philippines)
    - *fikas* strong; *fi+um+ikas* be strong
    - *kilad* red; *fi+um+ikas* be red
  - circumfix:
    - German *ge* + stem + *t*
    - e.g. *sagen*, *gesagt*
    - Maltese negative: *ma* + *naf* + *x*

# Morphological Processes

- Reduplication, e.g. in Indonesian  
*orang* man → *orang+orang* (men)
- Note that although reduplication is concatenative, it is context-dependent. What is inserted depends on what comes before.
- Vowel change
  - swim/swam
- Consonant change
  - send/sent

# Morphological Processes

- Interdigitation: a basic characteristic of Semitic languages (Maltese, Arabic, Hebrew, Akkadian, Syriac...)
  - Input radicals + vocalism
  - Output stem
  - Example  $k \ t \ b + i \ e \rightarrow \text{kiteb}$
- Interdigitation is an example of a *non-concatenative* operation.
- The output stem is then used as a basis for further morphological operations
- $n + \text{kiteb} + u \rightarrow \text{niktbu}$
- Note that the end result is not  $\text{nkitebu}$



# Compounding

- In contrast to derivations and inflections, where affixes are attached to a stem, in compounding two or more **lexemes** are joined together.
- Both lexemes might undergo modification in the process.
- In German, the concatenation is expressed in the orthography:

## Example

*lebensversicherungsgesellschaftsangestellter*

(life insurance company employee)

leben	s	versicherung	s	gesellschaft	s	angestellter
life		insurance		company		employee

- Morphotactics investigates the constraints imposed on the order in which morphemes are combined.
- Various kinds of such constraints are known.
- Constraints on the type of affix
  - “un” is a prefix
  - “tion” is a suffix
- Syntactic constraints
  - the suffix -able applies to verbs to yield an adjective
- Other constraints:
  - in English, “Latin” affixes are attached before “native” ones:  
non+im+partial non+il+legible  
in+non+partial in+non+legible

- Ideally, the task of a morphological analysis system would be to break the word down to its component morphemes and determine the meaning of the resulting decomposition.
- Most computational analyses assume *written* input.
- Therefore, strictly speaking, such systems embody orthographic rules, not phonological rules.
- This does not matter for languages where the standard orthography is a good phonemic transcription
  - Finnish - excellent
  - Maltese - not bad
  - English: cough, bough, rough, through, though. [terrible]
- For many languages, where the orthography does not reflect the phonology, a great deal of the effort in constructing computational models of morphology is spent on developing techniques for dealing with phonological rules.

# Phonological Rules

- Phonological rule: plosive consonant takes on the voicing of the next consonant.
  - Example: *kbir* (big) is pronounced *gbir*.
- The phonological rule is not reflected in the orthography (the word is written with “k”).
- Morphological rule
  - Example: city + s → cities
- The orthographic rule is not reflected in the phonology

- Morphological structure conveys important information of different types
- Morphological structure is manifest in different ways.
- Morphological structure is governed by rules.
- Morphological analysis attempt to discover that structure