Morphology is the area of linguistics which studies the structure of words.

Almost all natural language applications require some processing of words: lexicon lookup, morphological analysis and generation, part-of-speech determination etc.

In order to implement such functions, it is necessary to understand which morphological processes take place in a variety of languages.

Why look at many languages?
Example

حة דיבר כל הלילה

Observations:

- דיבר is third person, plural, past form of the verb דיבר
- this form is obtained by concatenating the suffix 1 to the base דיבר
- in the inflected form [dibru], the vowel [e] of the base [diber] is reduced to a schwa. This reduction is mandatory, as [diberu] is ungrammatical.
These simple observations shed light on a variety of issues:

- **What information is encoded by morphology?**
  In the example, morphology encodes details such as person, number and tense.

- **How does morphology encode information?**
  In the example, the final form is obtained by concatenating an affix (which is not a word) to the end of a base (which might be a word).

- **Interaction of morphology and phonology**
  In the example, the vowel [e] is shortened to a schwa.
Structure of this part of the course

- Parts of speech
- Typology of languages
- Inflection and derivation
- What information is encoded by morphology
- How morphology encodes information
  - concatenation, infixation, circumfixation, root and pattern, reduplication
- Interaction of morphology and phonology
Parts of speech

- Words are traditionally classified into categories, known as parts of speech or word classes.
- The major parts of speech are noun, verb, adjective, adverb, pronoun, proper name, preposition, conjunction, article etc.
- Subcategorization
- Open classes and closed classes
- Content words vs. function words
- Parts of speech can shed light on the context in which a word can occur, its neighbors and even its pronunciation: complex
- The problematic nature of this classification
- POS tagging
**Typology of languages**

- **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

No bound forms. Example: Mandarin Chinese

Example:

<table>
<thead>
<tr>
<th>gǒu</th>
<th>bú</th>
<th>ài</th>
<th>chī</th>
<th>qīngcài</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>not</td>
<td>like</td>
<td>eat</td>
<td>vegetable</td>
</tr>
</tbody>
</table>

Can mean any of the following (inter alia):

- 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 don’t like to eat vegetables
Agglutinative languages

Beads on a string. Example: Turkish
çöplüklerimizdekiledenmiydi

Example:

<table>
<thead>
<tr>
<th>Trash</th>
<th>Aff</th>
<th>Pl</th>
<th>1p/Pl</th>
<th>Loc</th>
<th>Rel</th>
<th>Pl</th>
<th>Abl</th>
<th>Int</th>
<th>Aux</th>
<th>Past</th>
</tr>
</thead>
<tbody>
<tr>
<td>çöp</td>
<td>lük</td>
<td>ler</td>
<td>imiz</td>
<td>de</td>
<td>ki</td>
<td>ler</td>
<td>den</td>
<td>mi</td>
<td>y</td>
<td>di</td>
</tr>
</tbody>
</table>

“was it from those that were in our garbage cans?”

? המישב_printf?
Polysynthetic languages

Morphology encodes units that are usually considered syntactic (as in noun incorporation). Example: Yupik

qayá:liy’u:l’u:n’i

Example:

qayá: li y’u: l’u: n’i
kayaks make excellent he Past
“he was excellent at making kayaks”

“The grammar is in the morphology”
Inflectional languages

Portmanteau morphemes: a single morpheme can encode various bits of information. Example: Latin

\textit{am\textbar{o}}

Example:

\begin{table}[h]
\centering
\begin{tabular}{ll}
\texttt{am} & \texttt{\textbar{o}} \\
love & 1p/Sg/Pres/Indicative/Active \\
& “I love”
\end{tabular}
\end{table}
Inflections and derivations

- *Inflectional* morphology takes as input a word and outputs a form of the same word appropriate to a particular context.

Example:

\[ \text{[dibber]} \Rightarrow \text{[dibbru]} \]

- The output is appropriate to a context in which the subject is third person plural and the tense is past.
- Hence: words have *paradigms*, defining all possible inflected forms of a word. Words which belong to the same paradigm are all *inflected forms* of a single *lexeme*. 
Derivational morphology takes as input a word and outputs a different word that is derived from the input. This is also called word formation.

Example:
establish+ment+ary+an+ism

Example:
הलשימת → החלש → החלש → החליש
Inflections and derivations: distinctive criteria

- Inflection does not change the part-of-speech, derivation might.
- Inflection is sometimes required by the syntax, derivation never is.
- If a language marks an inflectional category, it marks it on all appropriate words. In other words, the relation denoted by inflectional morphology is *productive*.
Verbs specify the number (and type) of arguments they may take. In many languages, morphological devices modify these lexically specified markings. Example: passivization (Latin)

Example:

puer Cicerōnem laudat
boy Cicero praise/3/Sg/Pres/Ind/Act
“the boy praises Cicero”

Cicerō laudātur
Cicero praise/3/Sg/Pres/Ind/Pass
“Cicero is praised”
Verbal morphology

Example: causativization

Example:

荷花 → >({荷花}; uml ) uml
Verbal morphology

Verbs are commonly marked with indications of the time at which the situations denoted by them occurred, or the state of completion of the situation. Such markers encode *tense* and *aspect*, respectively.

Example: Latin

<table>
<thead>
<tr>
<th>Latin</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>vir Cicerōnem laudābō</em></td>
<td>&quot;the man will praise Cicero&quot;</td>
</tr>
<tr>
<td><em>man Cicero praise/3/Sg/Future/Ind</em></td>
<td></td>
</tr>
<tr>
<td><em>vir Cicerōnem laudāvit</em></td>
<td>&quot;the man has praised Cicero&quot;</td>
</tr>
<tr>
<td><em>man Cicero praise/3/Sg/Perf/Ind</em></td>
<td></td>
</tr>
</tbody>
</table>
Verbal morphology

In many languages the verb must *agree* on person, number, gender or other features with one or more of its arguments.

Example:

<table>
<thead>
<tr>
<th>The princess</th>
<th>kisses</th>
<th>the</th>
<th>frog</th>
</tr>
</thead>
<tbody>
<tr>
<td>*The princess</td>
<td>kiss</td>
<td>the</td>
<td>frog</td>
</tr>
</tbody>
</table>

In some languages (e.g., Georgian and Chichewa) verbs agree not only with their subjects but also with their objects.
Nominal morphology

- Inflectional categories for nouns (and adjectives) include:
  - number (singular, plural, dual)
  - case (marking various kinds of syntactic/semantic functions)
  - gender (feminine, masculine, neuter)
- Latin has five cases: nominative, genitive, dative, accusative, ablative.
- Finnish has fourteen different cases!
- Example: the inflection paradigm of the adjective *magnus* (big) in Latin.
The inflection paradigm of Latin *magnus*

<table>
<thead>
<tr>
<th>Case</th>
<th>Masculine</th>
<th>Feminine</th>
<th>Neuter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>singular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nominative</td>
<td>magnus</td>
<td>magna</td>
<td>magnum</td>
</tr>
<tr>
<td>genitive</td>
<td>magni</td>
<td>magna</td>
<td>magni</td>
</tr>
<tr>
<td>dative</td>
<td>magni</td>
<td>magna</td>
<td>magni</td>
</tr>
<tr>
<td>accusative</td>
<td>magnum</td>
<td>magnam</td>
<td>magnum</td>
</tr>
<tr>
<td>ablativus</td>
<td>magni</td>
<td>magni</td>
<td>magni</td>
</tr>
<tr>
<td><strong>plural</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nominative</td>
<td>magnorum</td>
<td>magnarum</td>
<td>magnorum</td>
</tr>
<tr>
<td>genitive</td>
<td>magnis</td>
<td>magnis</td>
<td>magnis</td>
</tr>
<tr>
<td>dative</td>
<td>magnis</td>
<td>magnis</td>
<td>magnis</td>
</tr>
<tr>
<td>accusative</td>
<td>magnos</td>
<td>magnas</td>
<td>magnis</td>
</tr>
<tr>
<td>ablativus</td>
<td>magnis</td>
<td>magnis</td>
<td>magnis</td>
</tr>
</tbody>
</table>
Nominal morphology

- Many languages distinguish between two or three grammatical genders: feminine, masculine and neuter.
- In some languages, such as the Bantu languages, more detailed gender classes exist.
- 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.

### Example: Welsh

<table>
<thead>
<tr>
<th>gwyn</th>
<th>gwynn+<strong>ed</strong></th>
<th>gwynn+<strong>ach</strong></th>
<th>gwynn+<strong>af</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>as white</td>
<td>whiter</td>
<td>whitest</td>
</tr>
<tr>
<td>teg</td>
<td>tec+<strong>ed</strong></td>
<td>tec+<strong>ach</strong></td>
<td>tec+<strong>af</strong></td>
</tr>
<tr>
<td>fair</td>
<td>as fair</td>
<td>fairer</td>
<td>fairest</td>
</tr>
</tbody>
</table>
Derivational morphology

- In general, derivational morphology is not as productive as inflectional morphology.
- Nominalization: destroy → destruction; ביום שמור → המבס מקים פיהו
- Deverbal adjectives: drink → drinkable; אוכל → אוכל
- Denominalized adjectives: מזיאורה → מזיאורה
- Adjective nominalization: grammatical → grammaticality; שונים → שוליהות
- Negation: able → unable; ידוע → ידועי
In contrast to derivations and inflections, where affixes are attached to a stem, in compounding two or more lexemes’ stems are joint together, forming another lexeme.

Example:

- policeman; newspaper; יפה הוחם; עזרה דו

Both lexemes might undergo modification in the process.

In German, the concatenation is expressed in the orthography:

Example:

- Lebensversicherungsgesellschaftsangestellter
  - leben s versicherung s gesellschaft s angestellter
  - life insurance company employee
In what ways are morphemes expressed?

The simplest model of morphology is the situation where a morphologically complex word can be analyzed as a series of morphemes concatenated together.

An example: Turkish. Not only is Turkish morphology exclusively concatenative; in addition, all affixes are suffixes. Turkish words are of the form stem suffix*.

Example:

<table>
<thead>
<tr>
<th>çöp</th>
<th>lük</th>
<th>ler</th>
<th>imiz</th>
<th>de</th>
<th>ki</th>
<th>ler</th>
<th>den</th>
<th>mi</th>
<th>y</th>
<th>di</th>
</tr>
</thead>
<tbody>
<tr>
<td>trash</td>
<td>Aff</td>
<td>Pl</td>
<td>1p/Pl</td>
<td>Loc</td>
<td>Rel</td>
<td>Pl</td>
<td>Abl</td>
<td>Int</td>
<td>Aux</td>
<td>Past</td>
</tr>
</tbody>
</table>

“was it from those that were in our garbage cans?”
Linear concatenation is not the only way in which languages put morphemes together. Affixes may also attach as *infixes* inside words.

**Example: Bontoc (Philippines)**

<table>
<thead>
<tr>
<th>Original</th>
<th>Morpheme Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>fikas</td>
<td>f-um+ikas</td>
</tr>
<tr>
<td>strong</td>
<td>be strong</td>
</tr>
<tr>
<td>kilad</td>
<td>k-um+ilad</td>
</tr>
<tr>
<td>red</td>
<td>be red</td>
</tr>
<tr>
<td>fusul</td>
<td>f-um+usul</td>
</tr>
<tr>
<td>enemy</td>
<td>be an enemy</td>
</tr>
</tbody>
</table>
Realization of morphemes

- In the Bontoc case the infix must be placed after the first consonant of the word to which it attaches.
- In general, the placement of infixes is governed by prosodic principles.

**Example: Ulwa (Nicaragua)**

<table>
<thead>
<tr>
<th>suu+kí-lu</th>
<th>my dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>suu+ma-lu</td>
<td>your (Sg) dog</td>
</tr>
<tr>
<td>suu+ka-lu</td>
<td>his/her/its dog</td>
</tr>
<tr>
<td>suu+ni-lu</td>
<td>our (inclusive) dog</td>
</tr>
<tr>
<td>suu+kí+na-lu</td>
<td>our (exclusive) dog</td>
</tr>
<tr>
<td>suu+ma+na-lu</td>
<td>your (Pl) dog</td>
</tr>
<tr>
<td>suu+ka+na-lu</td>
<td>their dog</td>
</tr>
</tbody>
</table>
Some languages exhibit *circumfixes*, affixes which attach discontinuously around a stem.

**Example: German participles**

- säuseln: **ge+säusel+t**
- brüsten: **ge+brüst+et**
- täuschen: **ge+täusch+t**
In contrast to processes of attaching an affix to a stem, there exist also nonsegmental morphological processes.

A typical example is the Semitic root and pattern morphology.

Example: Hebrew *binyanim*

*a_a_, *ni_a_, *i_e_, *u_a_, *hi_i_, *hu_a_, *hit_a_e_.
Realization of morphemes

- Another nonsegmental process is *reduplication*.

**Example: Indonesian**

<table>
<thead>
<tr>
<th>Original</th>
<th>Reduplicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>orang</td>
<td>orang+orang</td>
</tr>
<tr>
<td>man</td>
<td>men</td>
</tr>
</tbody>
</table>

- Sometimes only part of the word is duplicated, as in Yidin (Australia) plural:

**Example:**

<table>
<thead>
<tr>
<th>Original</th>
<th>Reduplicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>mulari</td>
<td>mula+mulari</td>
</tr>
<tr>
<td>man</td>
<td>men</td>
</tr>
<tr>
<td>gindalba</td>
<td>gindal+gindalba</td>
</tr>
<tr>
<td>lizard</td>
<td>lizards</td>
</tr>
</tbody>
</table>
Morphotactics investigates the constraints imposed on the order in which morphemes are combined.

Various kinds of such constraints are known.

Example:

עִיֵּנָיוֹת → מֶבְטִיָּה מֶבְטִיָּה
but
עִיֵּנָיוֹת מֶבְטִיָּה; מֶבְטִיָּה עִיֵּנָיוֹת*
Morphology

Morphotactics

Types of constraints:

- Constraints on the type of the affix: םל is a prefix, ה is a suffix
- Syntactic constraints: [i] converts a noun to an adjective; [ut] converts an adjective to a noun
- 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.

Things are not that simple because of the often quite drastic effects of phonological rules. A great deal of the effort in constructing computational models of morphology is spent on developing techniques for dealing with phonological rules.

Since most computational analyses of morphology assume written input, phonological rules are often confused with orthographic ones.
Orthographic rules often do not correspond to phonological rules.

Example: Orthographic rules that don’t correspond to any phonological rule:

- city+s → cities (and not *citys)
- bake+ing → baking (and not *bakeing)

Example: A phonological rule (changing [aɻ] to [i]) is not reflected in the orthography:

- divine+ity → divinity

Example: A phonological rule (stress shift) is not reflected in the orthography:

- grammátical → grammaticáilty
## Phonology

- Examples of phonological rules

### Example: English: \([n]\) changes to \([m]\) before a labial consonant

im\_possible; im\_pose; im\_mortal

### Example: Finnish: vowel harmony

<table>
<thead>
<tr>
<th>NOM</th>
<th>PART</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>taivas</td>
<td>taivas+\textit{ta}</td>
<td>sky</td>
</tr>
<tr>
<td>puhelin</td>
<td>puheli+\textit{ta}</td>
<td>telephone</td>
</tr>
<tr>
<td>lakeus</td>
<td>lakeus+\textit{ta}</td>
<td>plain</td>
</tr>
<tr>
<td>syy</td>
<td>syy+\textit{tä}</td>
<td>reason</td>
</tr>
<tr>
<td>lyhyt</td>
<td>lyhyt+\textit{tä}</td>
<td>short</td>
</tr>
<tr>
<td>ystävällinen</td>
<td>ystävällinen+\textit{tä}</td>
<td>friendly</td>
</tr>
</tbody>
</table>