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

<div dir="rtl">

הם דיברו כל הלילה

</div>

Observations:

- דיברו is third person, plural, past form of the verb דיבר
- this form is obtained by concatenating the suffix ו 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.

Example

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

gǒu	bú	ài	chī	qīngcài
dog	not	like	eat	vegetable

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

çöp	lük	ler	imiz	de	ki	ler	den	mi	y	di
trash	Aff	Pl	1p/Pl	Loc	Rel	Pl	Abl	Int	Aux	Past

"was it from those that were in our garbage cans?"

המשבפחי ו?

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"

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

amó

Example

am	ó
love	1p/Sg/Pres/Indicative/Active
"I love"	

Inflections and derivations

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

Example: [dibber] ⇒ [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*.

Inflections and derivations

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: החלטיות ← החלטי ← החלטה ← החליט

Inflections and derivations: distinctive criteria

- Inflection does not change the part-of-speech, derivation might.
  החלטי–החלטה; החלטות–החלטה

- 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*.

החלטה	–	החלטי		החלטה	–	החלטות
הסכמה	–	הסכמי*		הסכמה	–	הסכמות

Verbal morphology

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"

	Cicerōnem	laudātur
	Cicero	praise/3/Sg/Pres/Ind/**Pass**

"Cicero is praised"

Verbal morphology

Example: causativization

Example

אכל → האכיל; יצא → הוציא

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

Example

vir	Cicerōnem	laudābō
man	Cicero	praise/3/Sg/**Future**/Ind

"the man will praise Cicero"

vir	Cicerōnem	laudāvit
man	Cicero	praise/3/Sg/**Perf**/Ind

"the man has praised Cicero"

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

Example

The	princess	kisses	the	frog
*The	princess	kiss	the	frog

הם	דברו	כל	הלילה
*הם	דברה	כל	הלילה

In some languages (e.g., Georgian and Chicheŵa) 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 semantic function)
- gender (feminine, masculine, neuter)

Latin has five cases: nominative, genitive, dative, accusative, ablative.

Finnish has fourteen different cases!

Example: the inflection paradigm of the noun *magnus* (big) in Latin.

The inflection paradigm of Latin *magnus*

Example

		masculine	feminine	neuter
sing.	nom	magn+**us**	magn+**a**	magn+**um**
	gen	magn+**ī**	magn+**ae**	magn+**ī**
	dat	magn+**ō**	magn+**ae**	magn+**ō**
	acc	magn+**um**	magn+**am**	magn+**um**
	abl	magn+**ō**	magn+**ā**	magn+**ō**
plur.	nom	magn+**ī**	magn+**ae**	magn+**a**
	gen	magn+**ōrum**	magn+**ārum**	magn+**ōrum**
	dat	magn+**īs**	magn+**īs**	magn+**īs**
	acc	magn+**ōs**	magn+**ās**	magn+**a**
	abl	magn+**īs**	magn+**īs**	magn+**īs**

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

Example

gwyn	gwynn+**ed**	gwynn+**ach**	gwynn+**af**
white	as white	whiter	whitest
teg	tec+**ed**	tec+**ach**	tec+**af**
fair	as fair	fairer	fairest

Derivational morphology

In general, derivational morphology is not as productive as inflectional morphology.

Nominalization: destroy → destruction; שמר → שמירה; פיתח → פיתוח; הסכים → הסכם

Deverbal adjectives: drink → drinkable; אכל → אכיל

Denominalized adjectives: מציאות → מציאותי

Adjective nominalization: grammatical → grammaticality; שליח → שליחות

Negation: able → unable; חוטי → אלחוטי

Compounding

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; עורך דין; יפת עיי ים

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

What are morphemes?

In order to know what morphemes are, it is useful to check in what ways they are 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

çöp	lük	ler	imiz	de	ki	ler	den	mi	y	di
trash	Aff	Pl	1p/Pl	Loc	Rel	Pl	Abl	Int	Aux	Past

"was it from those that were in our garbage cans?"

What are morphemes?

Linear concatenation is not the only way in which languages put morphemes together. Affixes may also attach as *infixes* inside words.
Example: Bontoc (Philippines)

Example

fikas	→	f-**um**+ikas
strong		be strong
kilad	→	k-**um**+ilad
red		be red
fusul	→	f-**um**+usul
enemy		be an enemy

What are 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)

Example

suu+**ki**-lu	my dog
suu+**ma**-lu	your (Sg) dog
suu+**ka**-lu	his/her/its dog
suu+**ni**-lu	our (inclusive) dog
suu+**ki+na**-lu	our (exclusive) dog
suu+**ma+na**-lu	your (Pl) dog
suu+**ka+na**-lu	their dog

What are morphemes?

Some languages exhibit *circumfixes*, affixes which attach discontinuously around a stem.
Example: German participles

Example

säuseln	**ge**+säusel+**t**
brüsten	**ge**+brüst+**et**
täuschen	**ge**+täusch+**t**

What are morphemes?

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*

Example

_a_a_, ni__a_, _i__el, _u__a_, hi__i_, hu__a_, hit_a__e_.

What are morphemes?

Another nonsegmental process is *reduplication*.
Example: Indonesian

Example

orang	→	orang+orang
man		men

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

Example

mulari	→	mula+mulari
man		men
gindalba	→	gindal+gindalba
lizard		lizards

So, what are morphemes?

In its most general definition, a morpheme is an ordered pair
$\langle \text{CAT}, \text{PHON} \rangle$, where CAT is the morphological category expressed
by the morpheme (for example, its syntactic and semantic
features), and PHON represents its phonological form, including the
ways in which it is attached to its stem.

Example

$\langle (Adj \rightarrow N, \text{"state of"}), ([ut], suffix) \rangle$ שליח ← שליחות

$\langle (root \rightarrow V, causative), (\_i\_\_e\_) \rangle$ ג.ד.ל ← גידל

What are words, then?

A morpheme is a pairing of syntactic/semantic information with phonological information. In the same way, it is useful to assume that words have dual structures: phonological and morphological. The two structures are not always isomorphic.

It is a fairly traditional observation in morphology that there are really two kinds of words from a structural point of view: phonological words and syntactic words. These two notions specify overlapping but not identical sets of entities. furthermore, the orthographic word might not correspond to any of these.

The answer depends on the application:

Sometimes it is sufficient to know that דברו is an inflected form of דיבר; sometimes morphological information is needed, either as a list of features (דברו is third person, plural, past form of the verb דיבר) or as a structure tree; sometimes it is better to produce a list of phonemes without determining word boundaries. For some applications, the root ד.ב.ר might be needed.

Morphotactics

Morphotactics investigates the constraints imposed on the order in which morphemes are combined.

Various kinds of such constraints are known.

Example:

על טבע → טבעי → טבעיות → טבעיות על but

טבעיות על*; *טבעותי על

Morphotactics

Types of constraints:

- Constraints on the type of the affix: על is a prefix, ות 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.
An orthographic rule that does not correspond to any phonological rule:

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

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

divine+ity → divinity

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

grammátical → grammaticálity

Phonology

Examples of phonological rules
English: [n] changes to [m] before a labial consonant:
impossible; **im**pose; **im**mortal
Finnish: vowel harmony

Example

NOM	PART	gloss
taivas	taivas+**ta**	sky
puhelin	puheli+**ta**	telephone
lakeus	lakeus+**ta**	plain
syy	syy+**tä**	reason
lyhyt	lyhyt+**tä**	short
ystävällinen	ystävällinen+**tä**	friendly