

Computational lexicology, morphology and syntax

Diana Trandabăț

Course 2

Academic year 2022-2023

Lexicology as a part of linguistics

- **Lexicology** - part of linguistics, dealing with the vocabulary of a language and the properties of words as the main units of language.
- **Vocabulary** means the sum of all the words in the language.
- Good knowledge of the **description** of the vocabulary, rules of word-formation, origin and history of words helps to guess and remember the meaning of new-learned words, to master the standards of their usage.

Branches of Lexicology

- **General** - the general study of words, irrespective of the specific features of any particular language
- **Special** - the description of the vocabulary of a given language
- **Historical** - the study of the evolution of a vocabulary as well as of its elements. This branch discusses the origin of words, their change and development.
- **Descriptive** - deals with the description of the vocabulary of a given language at a given stage of its development.

Word structure

- A word is an independent unit of a language
- A morpheme is the minimum meaningful language unit. But unlike a word it is not independent. It occurs in speech just as a constituent part of words.

Morphemes

- A root is a basic element of words. The root in English is very often homonymous with the word. It is one of the most typical features of the English language. Roots are called productive if they are capable of producing new words.
- Affixes may be subdivided into prefixes, suffixes and infixes. A suffix is a morpheme following the root and forming a new word.
- A prefix is a morpheme standing before the root and modifying meaning.

Exercise 1

- Which of the lexicology branches will this course deal with, and why do you think so?
- Give the morphematic division (root, prefix, suffix) of these words:
 1. unpredictable,
 2. cooperation,
 3. multidimensional,
 4. leadership,
 5. impression.

Native and Borrowed Words

- A native word is a word which belongs to the original stock.
- Native words constitute about 80% of the 500 most frequent words in a language.
- They may be characterized by a high lexical and grammatical valency (ability to combine with other words), high frequency and developed polysemy.
- They are often monosyllabic, have great word building power and enter a number of set expressions.

Degree of Assimilation

1) ***Completely assimilated words***

E.g.: *wine, window, chair.*

1) ***Partially assimilated words***

a. Loan words not assimilated semantically

E.g.: *toreador.*

b. Loan words not assimilated grammatically

E.g.: *bacillus.*

c. Loan words not assimilated phonetically or graphically.

E.g.: *tabacco, e-mail.*

2) ***Unassimilated***

E.g. *chauffeur, haute couture.* The words from other languages for which there are corresponding English equivalents are so-called barbarisms. E.g.: *ciao.*

Wordformation

- Word-formation is a process of creating new words by means of existing elements and according to the patterns and rules of a given language.
 - 1) Derivation
 - 2) Compounding
 - 3) Conversion
 - 4) Quantitative changes

Derivation

- Derivation is a kind of word-formation when a new word is formed by adding a derivational morpheme (usually suffix or prefix) to the root.
 - 1) **Suffixation** is a kind of word-formation when a new word is formed by adding a suffix to the root.
 - 2) **Prefixation** is a kind of word-formation when a new word is formed by adding a prefix to the root.

1 Noun-forming suffixes

- age** (passage, marriage, mileage...); -**ance/-ence** (assistance, predominance, correspondence...); -**dom** (freedom, kingdom...)
- ee** (employee, referee...); -**eer/er** (engineer, profiteer, manager...); -**ess** (manageress, heiress...)
- ist** (economist...)
- hood** (adulthood, singlehood...)
- ing** (building, meaning...)
- ion/-sion/-tion/-ition/-ation** (production, conclusion, realisation...)
- ism** (consumerism, perfectionism...)
- ment** (agreement, investment...)
- ness** (effectiveness...)
- ship** (ownership...)
- ty/-ity** (productivity, prosperity...)
- ure/-ture** (procedure, expenditure...)

2 Adjective-forming suffixes

- able/-ible** (manageable, permissible)
- al/-tal/-ial/-tial** (economical, statistical)
- ant/-ent** (redundant, dependent)
- ary** (monetary, inflationary)
- ate/-ete** (accurate, complete)
- ful** (dutiful, powerful)
- ish** (snobbish, reddish)
- ive** (effective, extensive)
- less** (effortless, powerless)
- like** (businesslike, lifelike)
- ly** (costly, orderly)
- ous/-ious** (ambiguous, nutritious)
- some** (troublesome, worrisome)
- y** (sexy, worthy)

3 Verb-forming suffixes

-en (brighten, moisten)

-ify/-fy (intensify, qualify)

-ize/*-ise (rationalize, advertise, stabilize)

***'ize'** is often used in **American English** (maximize) as an alternative spelling of 'ise' in British English (maximise).

4 Adverb forming suffixes

-ly (frequently, perfectly)

***-ward/-wards** (windward, backward, homewards)

-wise (vote-wise, percentage-wise) _

*Words formed with '**ward**' can usually be used as either **adverbs or adjectives** words formed with '**wards**' are mainly used as **adverbs** (e.g. westward, westwards).

Prefixation

- Prefixes modify the lexical meaning of the root;
- the simple word and its prefixed derivative usually belong to the same part of speech.
- The group of class-changing prefixes is rather small, e.g.: be- (belittle, befriend), de- (defrost, descale).

Negative prefixes

- ***Negative prefixes***

give negative, reverse or opposite meaning

a- (apolitical, asexual)

de- (destabilise, declassify)

dis- (disenfranchise, disinvest)

il- (before *l*: illegal)/ **im-**(before *p,b,m*: imperceptible)/ **in-**
(inadequate)/ **ir-**(before *r*: irresponsible)

non- (non-economic, non-profit)

un- (unacceptable, undemocratic)

Non-negative prefixes I

- 1) *Degree, measures or size:*
hyper- (hypercreative, hyperdevoted);
over- (overestimate, overcompensate); **semi-** (semiskilled, semi-annual); **super-** (super-dominant, supercharged);
ultra- (ultra-conservative, ultra-secret)
- 2) *Repetition or possibility:*
em-(before *p,b,m*)/**en-** (embark, enclose)
- 3) *Time, place, order relation:*
ex- (ex-employer, ex-tenant); **inter-** (inter-office, inter-government); **post-** (post-budget, post-election); **pre-** (pre-delivery, pre-budget)

Non-negative prefixes II

- 4) *Number and numeral relation:* **bi-** (bilateral, bilingual); **multi-** (multi-dimensional, multi-media); **uni-** (unilateral, unisex)
- 5) *Attitude, counteraction:* **anti-** (anti-EEC, antiestablishment); **auto-** (autodial, autonomy); **counter-** (countercharge, counteroffer); **pro-** (pro-business, pro-liberal)
- 6) *Pejoration:* **mis-** (miscalculation, mismanage); **pseudo-** (pseudo-creativity, pseudo-democratic)

Assignment 2

| <i>English</i> | <i>Suffix</i> | <i>Romanian</i> |
|----------------|---------------|-----------------|
| mileage | | |
| costly | | |
| manageable | | |
| salesmanship | | |
| expenditure | | |
| effortless | | |

Assignment 3

| <i>Word</i> | <i>Romanian</i> | <i>Noun</i> |
|-------------|-----------------|-------------|
| capable | | |
| efficient | | |
| proud | | |
| ready | | |
| wise | | |

Assignment 4

| Noun | Romanian | Verb | Adjective |
|---------------------|----------|------|-----------|
| <i>argument</i> | | | |
| <i>emptiness</i> | | | |
| <i>intensity</i> | | | |
| <i>satisfaction</i> | | | |
| <i>strength</i> | | | |

Compounds

- Endocentric compounds – the two constituent elements are clearly the *determinant* and *determinatum* (*ashtray, mousetrap, stepladder*)
- Exocentric compounds – the *determinatum* is not expressed (*hangover, killjoy, ladybird, forget-me-not*)
- Rhyme-motivated (*harum-scarum*)
- Pseudo-compounds (*mayday, hamburger*)
- Semiaffixes (*chairman, yes-man, kissproof*)

Conversion - zero derivation

- The process of converting words from one part of speech to another **without adding any derivative element** is called **conversion** or *zero derivation*.
- *"In English almost every word can be verbed..."*

Conversion - classification

- **Verbs**

(*to nurse, to hand, to e-mail, to finger, to hammer, to empty, to up, to blind*)

- **Nouns**

(*a go, a hunt, a lift, a find, pros and cons, whys, ups and downs, a black, breakdown, make-up, comeback, take-off*)

- **Occasional formations (nonce-words)**

Occasional words are usually emotionally coloured words coined for a unique occasion.

E.g. Don't *darling* me!, Don't *yes-mum* me!

- **Marginal Cases of Conversion**

Cases of formations by **shift of stress** are neither regular, nor productive.

E.g. verb > noun (*abstract, import, refill, transfer*)
verb > adjective (*frequent, moderate, perfect*)

Quantitative Changes

- *Clipping*
- *Blending*
- *Graphical Abbreviations*
- *Back-formation*

Clipping

- **Clipping (shortening)**

The shortening of words consists of the reduction of a word to one of its parts, as a result of which the new form is used as an independent lexical unit.

This type of word-formation is in English highly productive.

a. Final clipping – the beginning of the prototype is retained.

ad, advert < advertisement,
memo < memorandum, *lab* < laboratory,
gym < gymnasium, *vac* < vacuum cleaner.

b. Initial clipping – the final part is retained.

chute < parachute, *phone* < telephone,
copter < helicopter, *plane* < aeroplane.

c. The middle is retained. E.g. *Liz* < Elizabeth ,

flu < influenza, *tec* < detective.

d. The middle is left.

maths < mathematics, *ag'st* < against.

Blending

- Blending is a word-formation process of forming a new lexeme from **parts of two or more other words**.

E.g. *smog* < smoke + fog, *brunch* < breakfast + lunch, *tranceiver* < transmitter + receiver, *bit* < binary digit, *chunnel* < channel + tunnel...

Graphical abbreviations

- New lexical units formed from the *initial letters* of the words and pronounced as one word – *acronyms*. E.g. *UNESCO, AIDS, NATO, laser, radar*.
- New lexical units formed from the *initial letters* of the words with *alphabetic reading*. E.g. *BC, SOS, TV, VIP, VAT*.
- Initial abbreviation in which the first element is a letter and the second a complete word. E.g. *A-bomb, E-mail*.
- Latin abbreviations. E.g. *AD, BC, i.e., e.g.*
- Shortenings formed by a part of a word and the remaining part expressed by a capital letter or a figure. E.g. *2-nite, 4 you, par-T*.

Dictionaries

- Types of dictionaries:
 - Monolingual vs. Bilingual vs. Multilingual
 - Electronic vs. Printed
 - General Language vs. Specialized Language (Glossaries)
 - Alphabetic vs. Semantic
- Dictionaries for different formal aspects of lexical units:
 - Dictionaries of synonyms/antonyms
 - Dictionaries of frequency counts

eDTLR

- Electronic Thesaurus Dictionary of the Romanian Language
- What it contains:
 - spelling
 - pronunciation
 - inflected and derivative forms
 - etymology
 - part of speech
 - definitions
 - illustrative uses of alternative senses
 - synonyms and antonyms
 - special usage notes

WordNet

- What is missing in traditional dictionaries
 - It does not say, for example, that trees have roots, or that they consist of cells having cellulose walls, or even that they are living organisms
 - “Sense” of the super ordinate term aka **hypernym** (living plant or industrial plant)
 - **Coordinate terms** (bushes, shrubs, ...)
 - **Hyponyms** - types of trees (pine, tropical, deciduous..)
 - Information assumed to be known to everyone (trees have barks and leaves, they grow from seeds, they make their own food by photosynthesis- probably information for encyclopedia!)

What is WordNet?

- WordNet is a lexical database for the English language.
- WordNet 3.0 has [1]:
 - 117,097 nouns (average noun has 1.23 senses)
 - 11,488 verbs (average verb has 2.16 sense)
 - 22,141 adjectives
 - 4,601 adverbs
- Created and maintained at the Cognitive Science Laboratory of Princeton University
- Accessible online @
<http://wordnetweb.princeton.edu/perl/webwn>
(Also Downloadable)
- Interfaces available in C, .Net , Java, Perl, Php, Python, Sql etc..(JWNL, WordNet.Net, RTiA wordNet, pywordnet ..)

WordNet Structure

- Words are organized as synsets in WordNet
- There are four disjoint kinds of synsets, containing either
 - Nouns
 - Verbs
 - Adjectives
 - Adverbs

What is a synset?

- Basic unit of WordNet
- A group of synonymous words which refer to a common semantic concept
- Words may belong to more than one synset – first sense is the most frequent sense
- Words also include collocations (“eye contact”, “mix up”)

Synset example

- “car” as in
 - {car, auto, automobile, machine, motorcar}
 - {car, railcar, railway car, railroad car}.
- “Chocolate” as in-

| Synset | Gloss |
|---|--|
| {cocoa, chocolate hot chocolate} | (made from baking chocolate or cocoa powder and milk and sugar) |
| {chocolate} | (made from roasted ground cacao beans) |
| {chocolate, coffee, deep brown umber, burnt umber} | (a medium to dark brown color) |

How are synsets related?

- A list of pointers associated with each synset to express the relationship between synsets
- WordNet defines 17 relations
 - 10 between synsets
 - 5 between wordsense
 - "gloss" (between a synset and a sentence, i.e a textual definition for each synset)
 - "frame" (between a synset and a verb construction pattern)

WordNet relations

| Relation | Also called | Definition | Example |
|-----------------|---------------|---|---|
| Hypernym | Superordinate | From concepts to superordinates | <i>breakfast</i> ¹ → <i>meal</i> ¹ |
| Hyponym | Subordinate | From concepts to subtypes | <i>meal</i> ¹ → <i>lunch</i> ¹ |
| Member Meronymy | Has-Member | From groups to their members | <i>faculty</i> ² → <i>professor</i> ¹ |
| Has-Instance | | From concepts to instances of the concept | <i>composer</i> ¹ → <i>Bach</i> ¹ |
| Instance | | From instances to their concepts | <i>Austen</i> ¹ → <i>author</i> ¹ |
| Member Holonym | Member-Of | From members to their groups | <i>copilot</i> ¹ → <i>crew</i> ¹ |
| Part Meronymy | Has-Part | From wholes to parts | <i>table</i> ² → <i>leg</i> ³ |
| Part Holonym | Part-Of | From parts to wholes | <i>course</i> ⁷ → <i>meal</i> ¹ |
| Antonym | | Opposites | <i>leader</i> ¹ → <i>follower</i> ¹ |

Figure 19.2 Noun relations in WordNet.

| Relation | Definition | Example |
|----------|---|---|
| Hypernym | From events to superordinate events | <i>fly</i> ⁹ → <i>travel</i> ⁵ |
| Troponym | From a verb (event) to a specific manner elaboration of that verb | <i>walk</i> ¹ → <i>stroll</i> ¹ |
| Entails | From verbs (events) to the verbs (events) they entail | <i>snore</i> ¹ → <i>sleep</i> ¹ |
| Antonym | Opposites | <i>increase</i> ¹ ↔ <i>decrease</i> ¹ |

Figure 19.3 Verb relations in WordNet.

Beyond WordNet

- eXtended WordNet
- SentiWordNet
 - Each term in WordNet database is assigned a score of 0 to 1 in SentiWordNet which indicates its polarity
- WordNet for languages other than English
- FrameNet
- SentiFrameNet

Until next week...

“The beginning of wisdom is
the definition of terms.”

— Socrates

