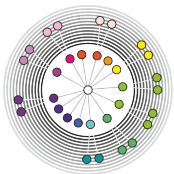




Semantic classes

Tsvetana Dimitrova
Valentina Stefanova

Institute of Bulgarian Language "Prof. Lubomir Andreychin"
Bulgarian Academy of Sciences



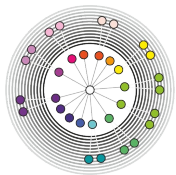
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Overview

- Corpus Pattern Analysis (CPA)
- Pattern Dictionary of English Verbs (PDEV)
- WordNet sentence frame
- WordNet noun hierarchy

The result: Mapping CPA with WordNet noun synsets

- Semantic classification of adjectives in the Bulgarian WordNet



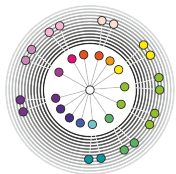
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Corpus Pattern Analysis (CPA)

The Corpus Pattern Analysis (CPA) semantic types are part of the Pattern Dictionary of English Verbs (PDEV) developed by Patrick Hanks and his observation about verb patterns and noun semantic types in sentences.

<https://pdev.org.uk/>

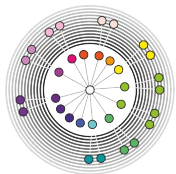
(Patrick Hanks. 2012. Mapping meaning onto use: a Pattern Dictionary of English Verbs. ACL 2008, Utah.)



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Corpus Pattern Analysis (CPA)

- a technique for mapping meaning of words onto their patterns of use as observed in real texts;
- a procedure in corpus linguistics which associates word meaning with word use through analysis of phraseological patterns and collocations.
- no attempt is made in CPA to identify the meaning of a verb or noun directly, as a word in isolation. Instead, meanings are associated with contexts.



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

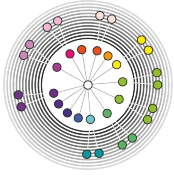
Pattern Dictionary of English Verbs (PDEV)

PDEV consists of verb patterns and semantic types of their nominal arguments organized within the CPA ontology.

Patterns can be described according to five types of arguments: Subject, Object, Complement, Adverbial, and Indirect Object.

Also can be described by using characteristics such as determiners, semantic types, contextual roles, and lexical sets.

- Determiners: “take place” and “take his place”;
- Semantic types: “building [[Machines]]” and “building [[Relationship]]”;
- Contextual roles: “[[Human = Film Director]] shoot” and “[[Human = Sports Player]] shoot”.



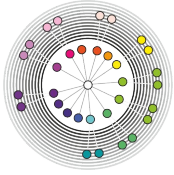
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

PDEV

- on the basis of corpus analysis (Sketch Engine and BNC)
- any verb pattern is construct of verbs and nouns use in particular sentence.
- 1. **[Human] | [Animal] consumes [Food] | [Beverage] | [Stuff]**
- 2. **[Human] | [Institution] | [Device] | [Activity] consumes [Artifact] | [Resource]**
- 3. **[Fire | Eventuality] consume [Physical_Object]**

Any verb pattern occupies a different use and verb meaning.

1. eat or drink
2. obtain
3. destroy



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

WordNet sentence frame

Each verb synset contains a list of generic sentence frames illustrating the types of simple sentences in which the verbs in the synset can be used.

Something ----s

Somebody ----s

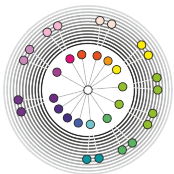
Somebody ----s something

Somebody ----s somebody

Somebody ----s something to somebody

Somebody ----s something from somebody

Organization is too general and some semantic classes can be missing. The sentence frames do not specify that the verbs like *achieve:1*, *accomplish:2*, *reach:9* 'to gain with effort' can be combined with nouns like *idea* (noun.cognition), *result* (noun.communication), *victory* (noun.event) but cannot co-occur with nouns such as *stone*, *table*, *sky*, etc.



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Example

There are many examples, such as in (1) where the sentence frame in (1a) signal that the verb can have both human and non-human subject argument. Further, (1c), which has a definition comparable to (1a), leaves only non-human subject arguments. In addition, the non-human subject arguments both in (1b) and (1c) may be specified as animate.

(1)

a. {purr:1, make vibrant sounds:1}: 'indicate pleasure by purring; characteristic of cats'

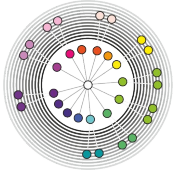
Something ----s, Somebody ----s

b. {moo:1, low:4}: 'make a low noise, characteristic of bovines'

Something ----s

c. {meow:1, mew:1}: 'cry like a cat; the cat meowed'

Something ----s



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

{agree:3} ‘achieve harmony of opinion, feeling, or purpose’

(WordNet) sentence frame: Somebody ----s; Somebody ----s PP (prepositional phrase)

(PDEV) verb pattern:

- [Human] | [Institution] agrees (with [Human] | with [Institution]) (about [Topic])

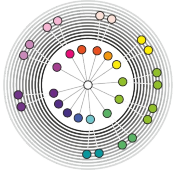
description: [[Human 1 | Institution 1]] indicates that {he | she | it} has the same opinion about [[Topic]] as [[Human 2 | Institution 2]]

- [Human] | [Institution] agrees (with [Proposition] | with [Activity])

description: [[Human | Institution]] indicates that {he, she, or it} has a favourable attitude towards [[Proposition | Activity]]

- [Human | Institution] agree [THAT]

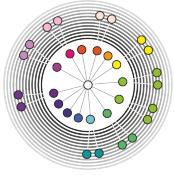
description: [[Human | Institution]] indicates that {he, she, or it} accepts {that-CLAUSE}



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

PDEV framework relies on semantic categories called semantic types, which refer to properties shared by a number of nouns that are found in verb pattern (argument) positions. Semantic types are formulated when they have been repeatedly observed in patterns and are organised into a relatively shallow ontology (up to 10 sublevels for some types).

<https://pdev.org.uk/>

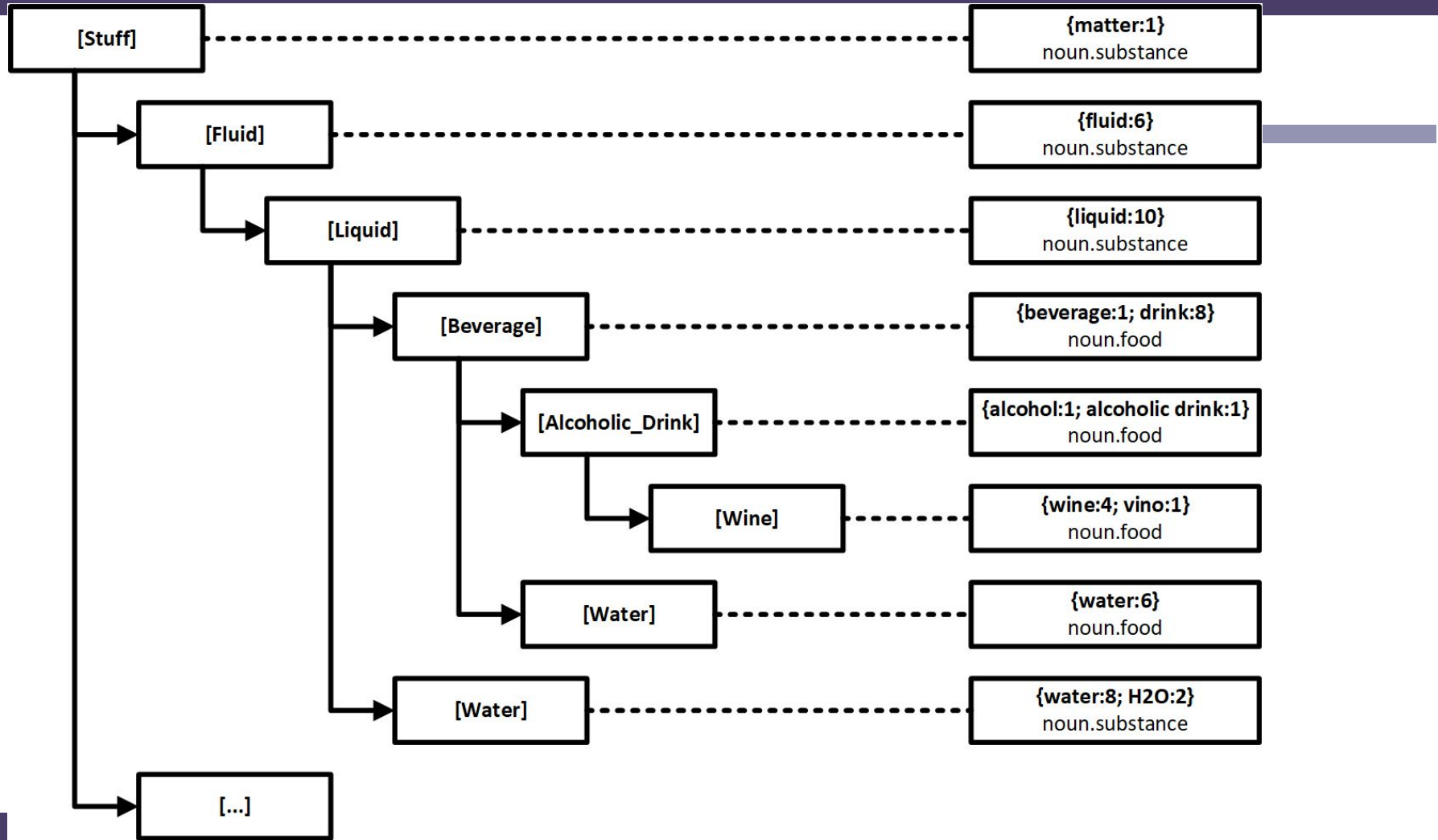


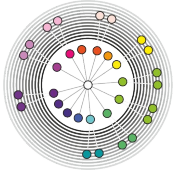
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

The CPA semantic types are corpus-driven types – they are formulated on the basis of real examples in corpora.

- The top-level type is [Anything] and it has six (sub)types: [Entity], [Eventuality], [Group], [Part], [Property], and [Not Connected].
- abstract types such as [**Anything**] are usually too broad
- some concepts are classified in different levels – [**Beverage**] is classified as both [**Physical Object**] [**Inanimate**] [**Artifact**] and as [**Physical Object**] [**Inanimate**] [**Stuff**] [**Fluid**] [**Liquid**].
- some concepts differs in more than one meaning: [**Water**] as [Liquid] and as [Beverage]

As in other ontologies, each semantic type inherits the formal property of the type above it in the hierarchy.



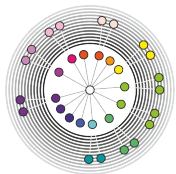


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

WordNet noun hierarchy

Nouns in WordNet are organised within the hypernym / hyponym hierarchy. The hierarchical semantic organisation is limited in depth, and distinguishing features are added to create lexical inheritance system where each word inherits the distinguishing features (attributes (modification), parts (meronymy), functions (predication)) from its hypernyms.

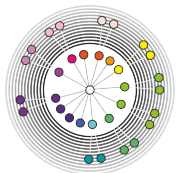
Noun synsets in WordNet are organised into 25 semantic classes (the so-called semantic primitives), namely nouns denoting humans (noun.person), animals (noun.animal), plants (noun.plant), acts or actions (noun.act), feelings and emotions (noun.feeling), spatial position (noun.location), foods and drinks (noun.food), etc.



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Example

{thing:1} – noun.Tops containing hyponyms labeled as noun.object;
{object:1; physical object:1} – noun.Tops, containing hyponyms that are noun.objects and noun.artifacts;
{causal agent:1; cause:1; causal agency:1} – noun.Tops, containing as hyponyms synsets labeled noun.person, noun.phenomenon, noun.state, noun.object, and noun.substance;
{matter:1} – noun.substance, containing hyponyms that are noun.substance and noun.object;
{process:1; physical process:1} – noun.process, with hyponyms marked as noun.process and noun.phenomenon;
{substance:7} – noun.substance (a sole synset).



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Mapping CPA with WordNet noun synsets

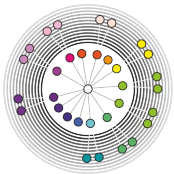
Our goal was then twofold: to identify the concept or the set of concepts to which a given CPA semantic type corresponds and to explore the structures of the two hierarchies: WordNet semantic primitives and CPA semantic types.

Mapping CPA with WordNet provide sets of concepts and their lexical representations linked to the CPA semantic types.

*general principle – the top synset or the synset that is on the highest level in tree structure in WN.

– [Permission] is mapped to permission:2, ‘approval to do something’ with semantic primitive noun.communication

– [Dispute] is mapped to disagreement:2, ‘the speech act of disagreeing or arguing or disputing’, semantic primitive noun.communication



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

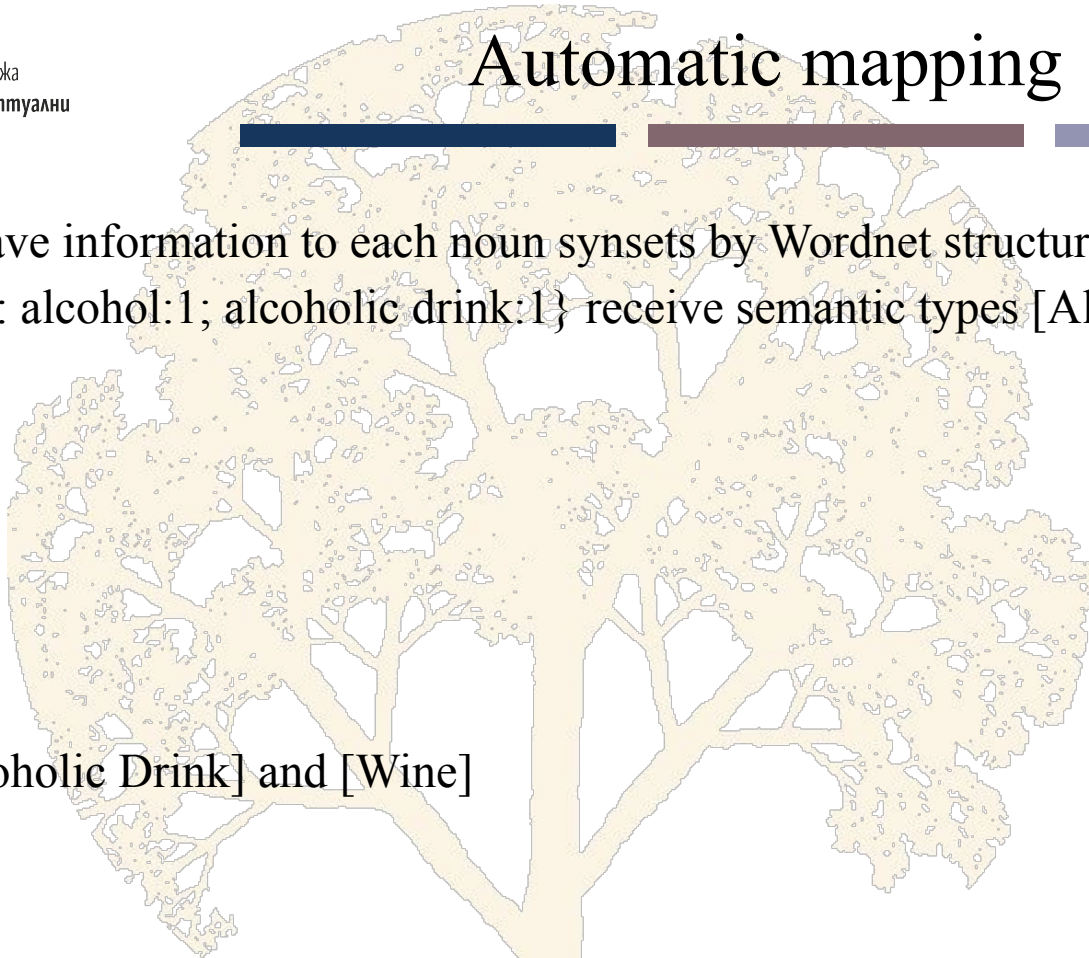
Automatic mapping

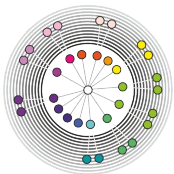
Automatically gave information to each noun synsets by Wordnet structure

Hyponyms of {n: alcohol:1; alcoholic drink:1} receive semantic types [Alcoholic Drink]:

- n: nipa:1
 - n: wine:4
 - n: pulque:1
 - n: liqueur:1
- etc.

n: wine:4 – [Alcoholic Drink] and [Wine]

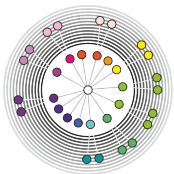




Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

The resulting data is available online, marked with the XML tag CPA in the
WordNet noun synsets.

http://dcl.bas.bg/PWN_CPA/



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

The structure

<SYNSET>

<ID>eng-30-09632518-n</ID>

<POS>n</POS>

<SYNONYM>

<LITERAL>worker

<SENSE>2</SENSE>

</LITERAL>

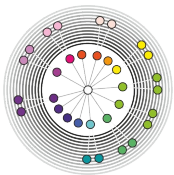
</SYNONYM>

<DEF>a person who works at a specific occupation</DEF>

...

<CPA>Human</CPA><CPA>Human_Role</CPA>

</SYNSET>



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

The structure

<SYNSET>

<ID>eng-30-02783324-n</ID>

<POS>n</POS>

<SYNONYM>

<LITERAL>ballroom

<SENSE>1</SENSE>

</LITERAL>

<LITERAL>dance hall

<SENSE>1</SENSE>

</LITERAL>

</SYNONYM>

<DEF>large room used mainly for dancing</DEF>

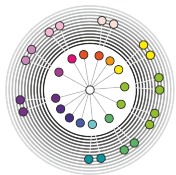
...

<CPA>Area</CPA>

<CPA>Room</CPA>

<CPA>Building</CPA>

</SYNSET>

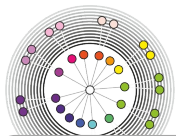


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Why we decided to map the two resources?

In the PDEV CPA ontology contains semantic types such as [Movie] and [Movie_Part], relevant for the semantic description of verbs like {film, shoot, take} 'make a film or photograph of something'. The PDEV pattern [Human] ----s [Movie] | [Movie_Part] represents the CPA semantic types of participants: [Human], [Movie] and [Movie_Part], that are relevant for distinguishing between different senses of these verbs.

For a comparison in the WordNet, nouns like {movie, film, picture, moving picture, moving-picture show} 'a form of entertainment that enacts a story by sound and a sequence of images giving the illusion of continuous movement' are classified within the more general semantic class: noun.communication



Обогатяване на
Семантич
Уърднет
Фреймо

The result can be prediction:

Wordnet

‘achieve harmony of
opinion, feeling, or
purpose’

verb.communication

Somebody ----s;
Somebody ----s PP

PDEV/CPA

{agree:3}

[Human] agrees (with [Proposition] | with [Activity])

defender

discussant

engineer

entertainer

expert

agrees

with

scenario

thesis

condition

provision

boundary
condition

with

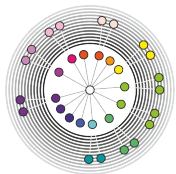
discussion

expression

examination

traffic

march

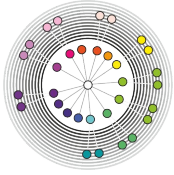


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

The adjectives in the Princeton WordNet

are divided into three semantic classes:

- descriptive adjectives (semantic prime **adj.all**) are organized into clusters based on similarity of meaning (synonymy) and binary opposition (antonymy).
- relational adjectives (**adj.pert**) are (derivationally) related with nouns and verbs, they are derived to.
- adjectival participles are marked as **adj.ppl** and are related via participle relation to synsets containing the verbs they are derived from.

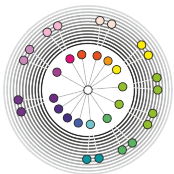


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Classification

A classification has been developed:

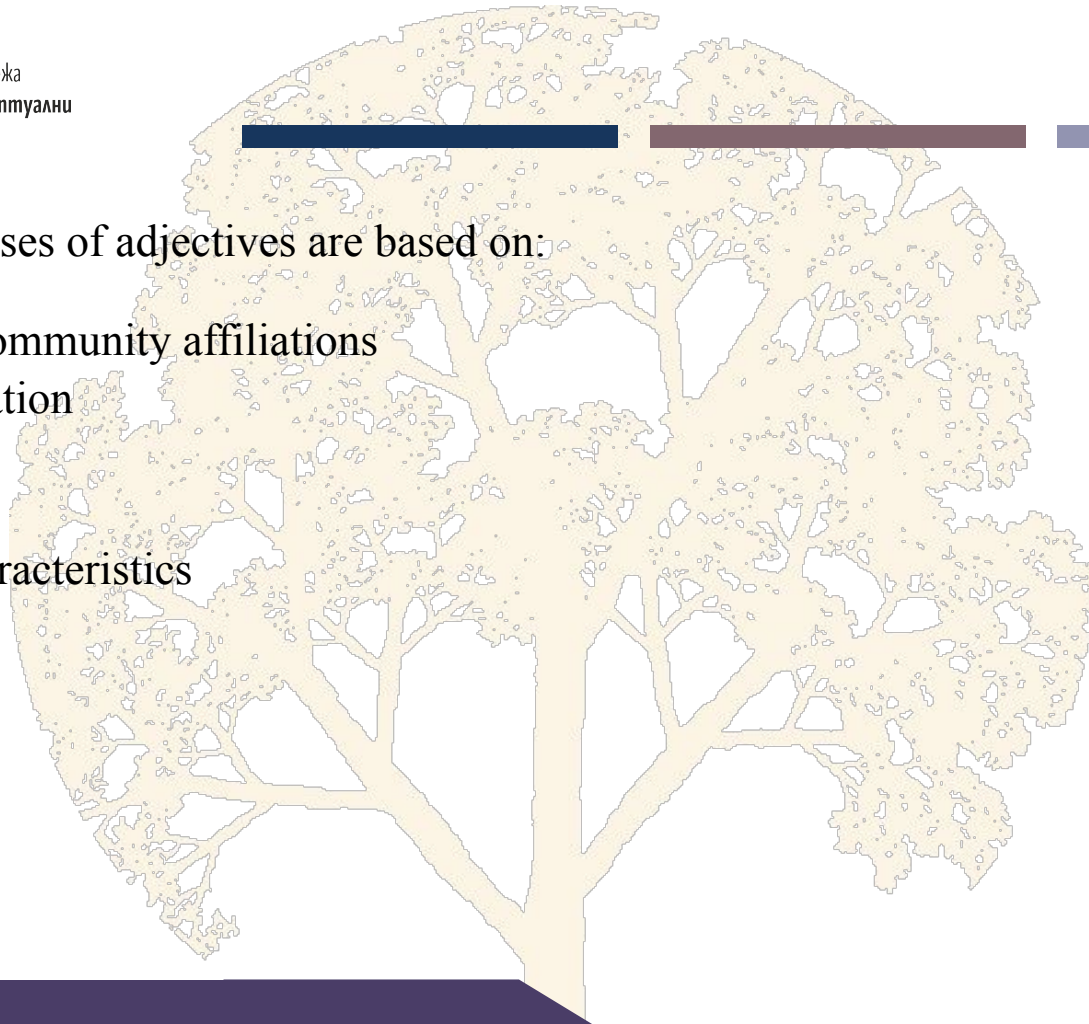
- after comparative analysis with another lexical-semantic network - the classification of adjectives in the WordNet for German, which is motivated by the fact that the modifying adjective is (semantically) related in a specific way with a given (modified) noun
- classifications proposed in linguistic literature
- on the base of information that is already available in WordNet based on relations between words from different parts of speech, their definitions and interrelations.
- on the base of the organization of the information and the structure of the Bulgarian WordNet follows those of the Princeton WordNet

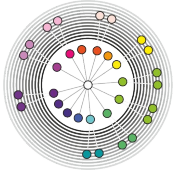


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

The semantic classes of adjectives are based on:

- social and community affiliations
 - place or location
 - time
 - weather
 - physical characteristics
 - movement
 - knowledge
 - attitude
 - feeling
 - behavior
- etc.





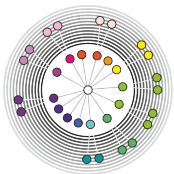
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Single groups

Adj.social: adjectives that express relations resulted from social norms and principles and concern entities or phenomena that are part of the social structure (incl. religion, ideology, marriage, etc.)

Ex.: {a: политически:1} / {a: political:1} involving or characteristic of politics or parties or politicians

similar to: a: governmental:1 [**adj.social**]

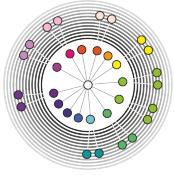


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Single groups

Adj.quantity: adjectives expressing quantity, size, degree, range, etc.

Ex.: {a: минимален:1} / {a: minimum:1} the least possible
antonym: a: maximal:1; maximum:1 **[adj.quantity]**



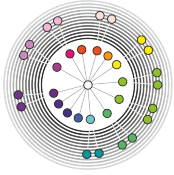
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Single groups

Adj.location: adjectives expressing spatial properties, succession in space etc.

Ex.: {a: тропически:1} / {a: tropical:3} relating to or situated in or characteristic of the tropics (the region on either side of the equator)

eng_derivative: n: Torrid Zone:1; tropical zone:1; tropics:1 [**noun.location**]



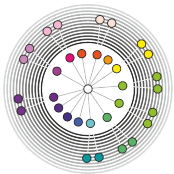
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Single groups

Adj.motion: adjectives related to manners of motion (vehicle, speed, etc.).

Ex.: {a: високоскоростен:1} / {a: high-speed:1; high-velocity:1} operating at high speed

similar_to: a: fast:5 [**adj.motion**]



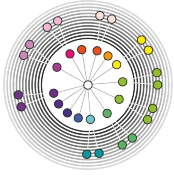
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Single groups

Adj.time: adjectives expressing age, historical period, succession in time, occurrence in a specific time period.

Ex.: {a: двумесечен:2} / {a: bimestrial:1} two months long; lasting two months

eng_derivative: n: bimester:1 **[noun.time]**



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

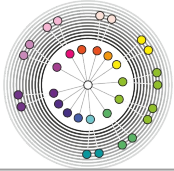
Single groups

Adj.weather: adjectives related to climate conditions (a limited number)

Ex.: {a: бурен: 6} / {a: stormy:1} two months long; lasting two months

eng_derivative: n: storm:1 [**noun.phenomenon**]

hypernym: n: atmospheric phenomenon:1 [**noun.phenomenon**]



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Couple

Adj.substance: adjectives expressing relation to substances

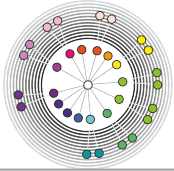
Ех.: { : алуминиев: 1 } / { a: aluminiferous: 1 } containing alum or aluminum

similar to: a: metallic:1; metal:1
[adj.substance]

Adj.material: adjectives expressing materials used for production of man-made objects.

Ех.: { a: дървен:1 } / { a: wooden:1 } made or consisting of (entirely or in part) or employing wood

a wooden box



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Close/ Similar

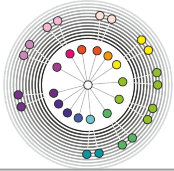
Adj.perception: adjectives for seeing (color), hearing (voice, sound), and perception (taste, sense, etc.) and estimation (liking/disliking, etc.)

Ex.: {a: златен:3; златист:1} / {a: gold:1; golden:3} having the deep slightly brownish color of gold
a gold carpet
eng derivative: n: amber:2; gold:3
[noun.attribute]

Adj.body: adjectives expressing physical characteristics of humans, animals, plants

Ex.: рус:1; рускоч:1; / blond:1 being or having light colored skin and hair and usually blue or grey eyes

antonym: a: brunet:1; Brunette:1 [adj.body]



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Close/ Similar

Adj.quality: adjectives expressing a property of an entity that is considered more or less an inherent attribute of this entity.

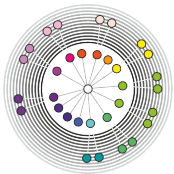
Ex.: {a: качествен:1} / {a: qualitative:1}
involving distinctions based on qualities

antonym: a: quantitative:1

Adj.state: adjectives expressing states of a person or an entity which are more or less stable for a period of time but can be subjected to change (physical, cognitive, etc.)

Ex.: {a: туберкулозен:2} / {a: tubercular:1}
constituting or afflicted with or caused by tuberculosis or the tubercle bacillus

eng derivative: n: tuberculosis:1; TB:3;
T.B.:1 [noun.state]



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

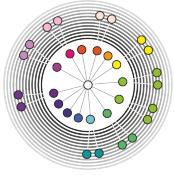
Related to person

Adj.cognition: adjectives denoting cognitive processes and contents and expressing cognitive abilities of a person or entities resulting from the cognitive activities.

Ex.: {a: схватлив:3} / {a: perceptive:1} having the ability to perceive or understand; keen in discernment

eng derivative: v: perceive:1 [**verb.cognition**]

eng derivative: n: insight:4; perceptiveness:4; perceptivity:1 [noun.feeling]



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Adj.feeling: adjectives for feelings and emotions of a person or related entities.

Ex: {a: отчаян:7} / {a: despondent:1} without or almost without hope

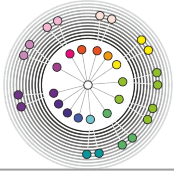
eng derivative: v: despond:1 [**verb.emotion**]

eng derivative: n: despondency:1; despondence:1; heartsickness:1 [**noun.feeling**]

Adj.behavior: adjectives expressing behaviors, behavioral symptoms, etc.

Ex.: {a: резервиран:2} / {a: reserved:1} marked by self-restraint and reticence

also see: en - a: undemonstrative:1 [**adj.behavior**]



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Abstract

Adj.cause: adjectives expressing abilities relating to change of state.

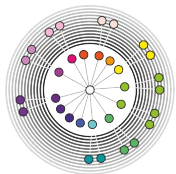
Ex.: {a: стимулиращ:3} / {a: stimulative:1}
capable of arousing or accelerating physiological or psychological activity or response by a chemical agent

eng derivative: v: stimulate:2; excite:1
[verb.change]

Adj.relation: adjectives denoting an explicit relation to an entity such as possession, purpose, function, composition, similarity, etc.

Ex.: {a: мъжки:1} / {a: male:3} for or pertaining to or composed of men or boys

eng derivative: n: maleness:1; masculinity:2
[noun.attribute]

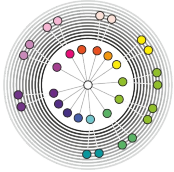


Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Attribution and verification

- classified a set of adjectives manually by applying only one semantic class to an adjective synset
- important is the semantic class of noun and verb
- important is also a lexico-semantic relations to other adjective synsets (the semantic class of other adjectives in the synset structure)
- the text in the definition also suggests a semantic class (to extract* synsets of adjectives)

*Hydra system – search in the database of the Bulgarian wordnet using regular expressions (Rizov, 2014)



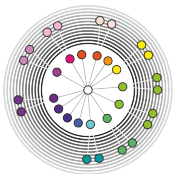
Attribution and verification

Keywords in the definition –

- the query [**lang('bg')&definition('#'...'')&pos('a')**] searches for all synsets in Bulgarian (lang('bg')) of the adjective part of speech (pos('a')) whose definition contains a string of symbols #'...'. .

For example: 'to feel', 'feeling', 'emotion' – adj.feeling / 'knowledge', 'cognition', 'experience' – adj.cognition / 'characterize', 'distinguish', 'quality' – adj.quality.

- on the other hand, the definition of a synset may cover more than one class of referents, as with adj.body where the property can be attributed to bodies of humans and/or animals



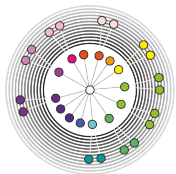
Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

Attribution and verification

Structure of the synset –

- the query `[lang('bg')&pos('a')&<eng_derivative>$s]`
- the query `[lang('bg')&sem_class ('adj.behavior')&<antonym>$s]`

It can be expected that semantically related adjectives are classified in the same semantic class (or similar) but there are a number of exceptions like the combinations in synset trees of adj.behavior, adj.feeling, adj.perception and adj.cognition.



Обогатяване на
Семантичната мрежа
Уърднет с концептуални
фреймове

- Dimitrova, Stefanova 2018: T. Dimitrova, V. Stefanova. Semantic classification of adjectives in the Bulgarian Wordnet: Toward a multiclass approach. – In: Etudes Cognitives, 18, 2018, 1-17.
- Koeva et al. 2018: S. Koeva, T. Dimitrova, V. Stefanova, D. Hristov. Mapping WordNet concepts with CPA ontology. – In: Proceedings of the 9th Global WordNet Conference (GWC'2018). Global WordNet Association, Singapore, 2018, 70-77.



**Enriching the Semantic Network WordNet with Conceptual Frames,
supported by the Bulgarian Science Fund, Contract KP-06-H50/1 of 2020.**