Classification of Adjectives in BulNet: Notes on an Effort

Valentina Stefanova, Tsvetana Dimitrova

Institute for Bulgarian Language,
Bulgarian Academy of Sciences,
valentina@dcl.bas.bg, cvetana@dcl.bas.bg

Abstract. The paper presents an overview of an attempt at the semantic classification of adjectives in the Bulgarian Wordnet based on the information that is already available in WordNet, and other classifications proposed in the literature (classifications in the linguistic literature for Bulgarian and approaches implemented by other wordnets, more precisely, the Wordnet for German). The proposed approach to adjective classification is a work-in-progress and involves using information from other synsets that are linked to the adjective synset at hand via lexico-semantic relations as introduced by the Princeton WordNet, their semantic primes and the texts of definitions and usage examples.

Keywords: WordNet, semantic classification, adjectives

1 Introduction

The paper proposes a semantic classification of adjectives in WordNet and an approach to its application to the Bulgarian Wordnet drawing upon existing classifications proposed in the linguistic literature for Bulgarian and the classification of adjectives (in German) as found in the Wordnet for German (GermaNet, cf. [6]). The approach relies on the relational network of the Princeton WordNet as the Bulgarian Wordnet largely keeps the original structure of the Princeton WordNet while introducing language-specific concepts in the appropriate place of the lexical hierarchy as well as closed class words (pronouns, prepositions, conjunctions, particles, and interjections, cf. [10]).

WordNet concepts are nodes linked to each other via lexico-semantic relations (edges) that encode relationships between concepts such as hypernymy/hyponymy, meronymy (membership, partiality, etc.), antonymy, synonymy, similarity, derivativity, etc. [4].

Additionally, each synset is classified by a semantic primitive [13], [4]. Nouns are organized into 25 semantic classes (*noun.person, noun.animal, noun.plant, noun.event*, etc.), and verbs are classified under 15 primes (*verb.change, verb.contact, verb.motion*, etc.). Adjectives are classified into two larger classes: descriptive adjectives and relational adjectives; and additional class of adjectival participles [3]. More detailed classifications of adjectives, however, are already implemented in other wordnets such as the Wordnet for German (GermaNet) [6].

In Section 2, we present in brief our points of reference. Section 3 gives an overview of the implemented classification of adjectives that is still a work-in-progress. Section 4 presents the first steps of our approach to classifying adjectives.

2 The Reference Points in Brief

The organization of the information and the structure of the Bulgarian Wordnet follows, in principle, those of the Princeton Wordnet [9], [11]. Thus, descriptive and relational adjectives are linked to other synsets via different sets of relations and were classified under three semantic classes – *adj.all, adj.pert* and *adj.ppl* – to constitute separate non-intersecting structures. 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 and subsequently linked to a synset containing their source noun. Adjectival participles are marked as *adj.ppl* and are related via participle relation to synsets containing the verbs they are derived from. Adjectives are organized also via a set of relations encoding their properties of attribution, antonymy, similarity, derivation, fuzzynymy, thematic category, etc.; some relations are specific for one of the two classes (attribute, similarity, fuzzynymy – for *adj.all*; pertainym – for *adj.pert*; participle – *for adj.ppl*) [2].
The reference point for the development of the presented semantic classification of adjectives was the classification of adjectives implemented in the Wordnet for German which is among the most detailed ones. It is based on a classification by Hundsnurscher and Splett [7] which employs the modification property of the adjective – a (modifying) adjective is (semantically) linked with a certain (modified) noun to form a separate semantic entity. It is hierarchically organized into 13 semantic fields that are divided by several sub-features – resulting in 70 (sub)classes organized around a specific feature.
Some of these classes were also found in the classifications in the linguistic literature on Bulgarian following the observation that there is no clear-cut division between qualitative and relational adjectives [5], [15], [14] [1], [18], [12]. Although, traditionally, the adjectives in Bulgarian are divided into two larger classes – qualitative and relational (roughly, in parallel to adj.all and adj.pert in WordNet) – the adjective is usually analyzed as dependent lexical class whose semantic and syntactic properties are fully realized only in its relation to a noun [15]. Thus, even when an adjective expresses a property of being related to an object or an event (as with relational adjectives), it expresses a relational property of another object or event that manifests in a certain way, to a certain degree or in relation to a certain internal property of the modified object or event.

3 The Classification: Work-in-Progress

As a result of the observations on the literature, we embarked upon outlining semantic classes of adjectives covering: social and community affiliations (social-

related, Gesellschaft in GermaNet (GN)); place or location; local time (time, Zeit in GN); weather (natPhaenomen in GN); physical characteristics (body, Koerper); movement (motion, Bewegung in GN); knowledge (cognition, Geist in GN); attitude (relation, Relation in GN); feeling (Gefuehl in GN); behavior (Verhalten in GN). Three of the semantic classes in GN were too general and included adjectives that can be considered attributes: General (Allgemein), Pertainyms (Pertonym), Privative (Privativa). The classification of adjectives in WordNet needs also to comply with the classifications of nouns and verbs already implemented in WordNet through introducing the semantic primes to noun and verb synsets. When formulating semantic classes, we had to combine all this information.
We expanded the class of physical characteristics with adjectives associated with physical properties of animals (coating, fur, tail, color, etc.) and plants (flowering, roots, etc.). The substance class was additionally re-organized to cover ingredients; and material (for man-made objects). Additionally, we introduced classes for adjectives denoting a state (of a person or an object), a causing phenomenon or trigger of change of state, and adjectives that expresses quality characteristics of animate and inanimate objects.
The attempted semantic classification of the adjectives as applied to the Bulgarian Wordnet combines some of the classes outlined above (mostly adopted from the GermaNet classification) plus some information from the classification of verbs and nouns (following the Princeton WordNet classes – the semantic primes have been previously validated and some changes have been introduced into BulNet and Princeton WordNet – the effort is described in [8]).

(1) Adj.body: adjectives expressing physical characteristics of humans, animals, plants (it is possible for an adjective to refer to a property that may still be considered under another class but if it is referred to property that can be attributed to body part, it is still classified as adj.body)[1].:

Ex.: {*a: rus:1; rusokos:1; svetlokos:1*} / {*a: blond:1*} *being or having light colored skin and hair and usually blue or grey eyes*
antonym: a: brunet:1; brunette:1 ***[adj.body]***
eng_derivative: n: paleness:2; blondness:1; fairness:3 ***[noun.attribute]***
has_attribute: n: complexion:4; skin color:1; skin colour:1 ***[noun.attribute]***

(2) Adj.cognition: adjectives denoting cognitive processes and contents and ex-

[1] Examples in this section include: the synset in Bulgarian, in BulNet (transliterated); / the synset in English, in PWN 'the definition in English'. Other synsets linked via lexico-semantic relation to the adjective synsets, with semantic primes/classes given in bold, in square brackets. The semantic class of the adjectives – if given – can be found only in BulNet, but not in PWN – it can be checked at: http://dcl.bas.bg/bulnetedit/. There is a parallel view with Bulgarian vs. English, the parallel synsets are marked by a red arrow beside the synset (the white arrow turns into red) (for description of the viewer and its features, cf. [16]). Examples do not give the whole synset tree and do not include all literals in a synset are included.

pressing cognitive abilities of a person or entities resulting from the cognitive activities.
Ex.: {*a: pronitsatelen:3; shvatliv:3*} / {*a: perceptive:1*} *having the ability to perceive or understand; keen in discernment*
antonym: a: unperceptive:2; unperceiving:1 ***[adj.cognition]***
eng_derivative: v: perceive:1 ***[verb.cognition]***
eng_derivative: n: perceptiveness:1 ***[noun.attribute]***
eng_derivative: n: discernment:3; perceptiveness:2 ***[noun.cognition]***
eng_derivative: n: insight:4; perceptiveness:4; perceptivity:1 ***[noun.feeling]***

(3) Adj.feeling: adjectives for feelings and emotions of a person or related entities.
Ex: {*a: otchayan: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]***
similar_to: a: hopeless:2 ***[adj.feeling]***

(4) Adj.behavior: adjectives expressing behaviors, behavioral symptoms, etc.
Ex.: {*a: rezerviran:2*} / {*a: reserved:1*} *marked by self-restraint and reticence*
also_see: en - a: backward:5 ***[adj.behavior]***
also_see: en - a: undemonstrative:1 ***[adj.behavior]***

(5) Adj.perception: adjectives for seeing (color), hearing (voice, sound), and perception (taste, sense, etc.) and, rarely, estimation (liking/disliking, etc.)
Ex.: {*smradliv:2; zlovonen:2*} / {*a: fetid:1*} *offensively malodorous*
eng_derivative: n: harmfulness:1; noisomeness:1; noxiousness:1 ***[noun.attribute]***
eng_derivative: n: olfactory property:1; smell:7; odor:1; odour:1 ***[noun.attribute]***
eng_derivative: n: malodorousness:1; stinkiness:1; foulness:1 ***[noun.attribute]***
similar_to: a: malodorous:1; malodourous:1; unpleasant-smelling:1 ***[adj.perception]***

(6) Adj.time: adjectives expressing age, historical period, succession in time, longevity, occurrence in a specific time period.
Ex.: {*a: dvumesechen:2*} / {*a: bimestrial:1*} *two months long; lasting two months*
eng_derivative: n: bimester:1 ***[noun.time]***
similar_to: a: long:5 **[adj.time]**

(7) Adj.location: adjectives expressing spatial properties, placement, succession in space, etc.
Ex.: {*a: tropicheski: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]***
eng_derivative: n: tropic:3 **[noun.location]**
similar_to: a: equatorial:1 **[adj.location]**

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

Ex.: *a: visokoskorosten:1, barzohoden:1 / a: high-speed:1; high-velocity:1 operating at high speed*
similar_to: a: fast:5 ***[adj.motion]***

(9) Adj.social: adjectives that express relations resulting from social norms and principles or concern entities or phenomena that are part of the social structure (incl. religion, ideology, marriage, etc.)
Ex.: {*a: politicheski:1*} / {*a: political:1*} *involving or characteristic of politics or parties or politicians*
antonym: a: nonpolitical:1 ***[adj.social]***
similar_to: a: governmental:1 ***[adj.social]***

(10) Adj.substance: adjectives expressing relation to substances.
Ex.: {*a:kristalen:3, kristalinen:1*} / {*a: crystalline: 1*} *consisting of or containing or of the nature of crystals*
similar_to: a: crystallized:1; crystalised:1 ***[adj.quality]***

(11) Adj.material: adjectives expressing materials used for production of man-made objects.
Ex.: {*a: darven:1*} / {*a: wooden:1*} *made or consisting of (entirely or in part) or employing wood*
similar_to: a: woody:3 ***[adj.substance]***

(12) Adj.weather: adjectives related to climate conditions (a limited number).
Ex.: {*a: tropichen:1*} / {*a: tropical:1*} *of weather or climate; hot and humid as in the tropics*
eng_derivative: n: tropic:3 ***[noun.location]***
similar_to: a: hot:9 ***[adj.perception]***

(13) Adj.quantity: adjectives expressing quantity, size, degree, range, etc.
Ex.: {*a: minimalen:1*} / {*a: minimum:1*} *the least possible*
also_see: a: meager:1; meagre:1; meagerly:1; stingy:1; scrimpy:1
antonym: a: maximal:1; maximum:1 ***[adj.quantity]***
*similar_to: a: negligible:1****[adj.quantity]***

(14) 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: tuberkulozen: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]***
eng_derivative: n: tubercle:3 ***[noun.state]***
similar_to: a: ill:8; sick:7 ***[adj.state]***

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

Ex.: {*a: stimulirasht:3*} / {*a: stimulative:1*} *capable of arousing or accelerating physiological or psychological activity or response by a chemical agent*
eng_derivative: v: stimulate:1; arouse:3; brace:1; energize:1; energise:1; perk up:1 ***[verb.body]***
eng_derivative: v: stimulate:2; excite:1 ***[verb.change]***
eng_derivative: v: stimulate:7; excite:8; stir:9 ***[verb.perception]***

(16) Adj.quality: adjectives expressing a property of an entity that is considered more or less an inherent attribute of this entity.
Ex.: {*a: kachestven:1*} / {*a: qualitative:1*} *involving distinctions based on qualities*
antonym: a: quantitative:1

(17) Adj.relation: adjectives denoting an explicit relation to an entity such as possession, purpose, function, composition, similarity, etc.
Ex.: {*a: mazhki:1*} / {*a: male:3*} *for or pertaining to or composed of men or boys*
eng_derivative: n: maleness:1; masculinity:2 ***[noun.attribute]***
eng_derivative: n: male:5; male person:1 ***[noun.person]***
similar_to: a: male:2

The adjective expresses an attribute of an entity but in WordNet the attribute can refer to different entities – and this can be tracked down the synset tree as well as in the definiton (and examples). For example, the adjectives expressing physical state can express also attributes of a human, an animal, a plant, etc. Many adjectives express attributes pertaining to entities of different nature. For example, {perceptive:1} having the ability to perceive or understand; keen in discernment can be used both for human abilities and for actions attributed to human abilities; the synset is linked to nouns classified as *noun.attribute, noun.cognition, noun.feeling*, as well as to *verb.cognition*. Another example is the adjective {consumptive:2} – which was classified here as *adj.state* – with the definiton 'afflicted with or associated with pulmonary tuberculosis', and examples 'a consumptive patient' (a state of a person – referring to possible classification of *adj.body*) and 'a consumptive cough' (a hint at *adj.cause* – 'a cough caused by tuberculosis). There are more than one way to resolve this – we may either split a synset, or classify it into more than one semantic class (or formulate subclasses). These approaches need further consideration. The next section presents the first stage of our effort.

4 Testing Ground

At the first stage, we have classified the adjectives manually taking into account additional information available from wordnet such as the semantic class of noun and verb synsets linked via lexico-semantic relations to the adjective synsets, the semantic class of the adjectives in the synset tree, and the text in the definition.

A set of 2500 adjective synsets – only the ones marked as *adj.all* – have been manually classified (as of April, 2017) experimentally.
We have used information provided in the definitions to extract synsets of adjectives – f.ex. the query [sem_class('adj.all')&definition('in time')] (cf. [17] for the modal language for wordnet) returns [antecedent:1, anterior:1, subsequent:1, precedent:1, previous:1 (...)].
The adjective classes comply with the semantic classes of nouns and verbs – either the ones directly linked to the adjective or the ones that are indirectly linked via other adjectives in the synset tree. Adjectives are linked to noun and verb synsets via lexico-semantic relations: ⟨ has_attribute ⟩; ⟨ eng_derivative ⟩; ⟨ category_domain ⟩; ⟨ usage_domain ⟩ – linking synsets in a topical class). Table 1 presents an overview of the semantic classes of the nouns to which some of the adjectives are directly linked (but not indirectly, i.e., via other adjective synsets directly linked to noun synsets)[2]. Further, an adjective synset is linked to other

Adj.class	All	Attribute noun.class	Derivative noun.class	Category noun.class
body	327	16: attribute: 15	51: state: 18, body: 13, attribute: 9, animal: 8 (...)	40: cognition:38(...)
behavior	297	46: attribute: 44, state: 2	91: attribute: 55, act: 15, person:13, feeling: 9(...)	4
perception	265	22: attribute: 22	54: attribute: 32, food: 10, cognition: 5(...)	4
quantity	256	19: attribute:18	20: quantity, attribute: 6(...)	5
cognition	214	9: attribute:4, cognition:5	49: cognition: 17, state: 14, person: 11(...)	39: cognition:34(...)
time	202	21: attribute:17(...)	17: time:8. attribute:3 (...)	7: cognition: 7(...)
feeling	175	11: feeling:5(...)	45: feeling: 31, attribute: 13, state: 5(...)	0
cause	168	9: attribute:3, state:3	42: attribute:12, state: 11, artifact:8, substance:6, act:5(...)	3
social	161	15: attribute:4(...)	22: person:11, attribute:4 (...)	4
location	151	14: attribute:8, location: 6	12: location:6, artifact, object: 2 (...)	4

Table 1. Distribution of adjective classes and noun classes

adjective synsets in a synset tree via relations for antonymy (⟨ antonym ⟩), and similarity (via relation ⟨ similar_to ⟩ – for semantic similarity between focal and

[2] Table 1 gives numbers for adjective classes with over 150 synsets being classified.

satellite synsets with close referential meaning; and relation ⟨ also_see ⟩ to link semantically related synsets. Information about semantic classes of adjectives can be obtained also by analyzing these relations. It can be expected that semantically related adjectives are classified in the same semantic class (or similar) but there are a number of exceptions as exemplified in Table 2 for the combinations in synset trees of *adj.behavior, adj.feeling, adj.perception* and *adj.cognition.* The proposed classification resolves some issues concerning the structure of the

Adj.class	Adj.class	similar_to	also_see
adj.cognition	adj.behavior	12	3
adj.cognition	adj.cognition	87	34
adj.behavior	adj.feeling	15	12
adj.behavior	adj.behavior	104	61
adj.perception	adj.perception	226	28
adj.feeling	adj.feeling	72	20

Table 2. Combination of adjective classes in adjective synset trees

wordnet and the synset references while leaving others aside for further consideration. First, 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, animals, etc. Further, there are adjectives which may express differentially attributed properties – an issue that is mirrored in the combined synsets in the synset tree – see with *adj.behavior* and *adj.cognition*, *adj.behavior* and *adj.feeling* in Table 2. The keywords in the definitions are a telling sign (keywords for experiences or feelings or relations to *verb.emotion* or *noun.feeling* signals for *adj.feeling*; showing, reacting, behaving, etc. – *adj.behavior*; words for causing or relation to *verb.cause* – *adj.cause*; animate and inanimate objects related to *noun.state* or *verb.state* fall into *adj.state*; etc.).

5 Conclusion

The paper presented the first step of an attempt to classify adjectives in WordNet drawing upon information that is already available in the structure of the resource (both in the lexico-semantic relational structure as inherited from the Princeton WordNet, and in content of the definitions and examples). We are still at the stage of formulating the semantic classes while classifying the candidate synsets manually. The goal of this manual effort is to extract patterns to help us further with extraction and possible automatic classification.

6 Acknowledgements

The work reported in this paper is carried out within the project Semantic Classification of the Adjectives in the Bulgarian WordNet of the Institute for

Bulgarian Language which is supported under the Program for Career Development of Young Scientists (2016) at the Bulgarian Academy of Sciences. We would also like to thank the three anonymous reviewers and the participants in the workshop *Challenges for Wordnets* co-located with *LDK 2017: The first conference on Language, Data and Knowledge.*

References

1. Barbolova, Z.: Modern Bulgarian Language. Morphology. Sofia (1997) [in Bulgarian]
2. Dimitrova, T., Stefanova, V.: Adjectives in Wordnet: Semantic Issues. In: Proceedings of the 12th International Conference Linguistic Resources and Tools for Processing the Romanian Language, 131–141. (2016)
3. Fellbaum, C., Gross, D., Miller, K.: Adjectives in WordNet. In: Introduction to WordNet: an On-line Lexical Database. Five Papers on WordNet. (1993)
4. Fellbaum, C., Osherson, A., Clark, P.: Putting Semantics into WordNets Morphosemantic Links. In: Proceedings of the Third Language and Technology Conference, vol. 5603, 350–358. Poznan (2009)
5. Grammar of the Contemporary Bulgarian Standard Language. Vol. 2: Morphology. Sofia: Bulgarian Academy Publishing House, 148–156. (1983) [in Bulgarian]
6. Hamp, B., Feldweg, H.: GermaNet – a Lexical-Semantic Net for German, https://aclweb.org/anthology/W/W97/W97-0802.pdf (1997)
7. Hundsnurscher, F., Splett, J.: Semantik der Adjektive im Deutschen: Analyse der semantischen Relationen. Wiesbaden: Westdeutsches Verlag (1982)
8. Koeva, S., Leseva, S., Stoyanova, I., Dimitrova, T., Todorova, M.: Automatic Prediction of Morphosemantic Relations. In: Proceedings of the Eighth Global Wordnet Conference, 27–30. Bucharest (2016)
9. Koeva, S. Derivational and Morphosemantic Relations in Bulgarian Wordnet. Intelligent Information Systems, 359368. (2008)
10. Koeva, S., Tinchev, T., Mihov, S.: Bulgarian Wordnet – Structure and Validation. Romanian Journal of Information Science and Technology, 7(1-2), 61-78.
11. Koeva, S.: WordNet and BulNet. In: Language Resources and Technologies for Bulgarian Language. Sofia, 154–173. (2014) [in Bulgarian]
12. Kucarov, I.: Theoretical Grammar of Bulgarian Language. Sofia (2007) [in Bulgarian]
13. Miller G., Beckwith R., Fellbaum C., Gross D., Miller K.: Introduction to WordNet: an On-line Lexical Database. Five Papers on WordNet. (1993)
14. Radeva, B.: Relational Adjectives in Contemporary Bulgarian Language. Ph.D. thesis. Sofia (2011) [in Bulgarian]
15. Radeva, V.: Word Formation in Bulgarian Literary Language. Sofia (1991) [in Bulgarian]
16. Rizov, B., Dimitrova, T., Barbu Mititelu, V.: Hydra for Web: A Multilingual Wordnet Viewer. In: Proceedings of the 11th International Conference "Linguistic Resources and Tools for Processing the Romanian Language", 19–30. (2016)
17. Rizov, B.: Hydra: A Modal Logic Tool for Wordnet Development, Validation and Exploration. In: Proceedings of the 6th Language Resources and Evaluation Conference. (2008)
18. Spasova, A.: On Relational Adjectives in Contemporary Bulgarian Language. Bulgarian Language 4–5, 381–390 (1959) [in Bulgarian]