Conceptual Description of Verbs and its Semantic and Syntactic Features

Ivelina Stoyanova and **Svetlozara Leseva**
Institute for Bulgarian Language Prof. Lyubomir Andreychun
Bulgarian Academy of Sciences
52 Shipchenski prohod blvd, bldg. 17, Sofia 1113, Bulgaria
{iva,zarka}@dcl.bas.bg

Relevant UniDive working groups: WG1, WG2

1 Introduction

The paper focuses on describing an effort at obtaining a rich semantic and syntactic description of verbs in WordNet[1] (Miller, 1995; Fellbaum, 1998) through mapping other lexical and conceptual resources to it – FrameNet[2] (Baker et al., 1998; Baker, 2008) and VerbNet (Kipper-Schuler, 2005; Kipper et al., 2008), in particular. This has been achieved through aligning elements of the semantic and syntactic description of the entities in these resources.

It has been acknowledged that combining WordNet with conceptual resources such as FrameNet produces a more complete semantic and syntactic representation of the lexical entries (Baker and Fellbaum, 2009; Schneider, 2012), thus expanding the possible applications of the resources for the purposes of syntactic and semantic parsing.

WordNet provides vast lexical coverage and representation of the lexis not only through the synsets' glosses and examples but also through the relationships established among synsets; FrameNet offers a semantic representation of the participants and circumstances used to define the situation described and their syntactic realisations; VerbNet adds a more general semantic description in line with the theory of semantic roles, along with syntactic patterns and alternations.

In this proposal, the bridging among the resources is implemented in the following way: verb synsets in WordNet are provided conceptual representation by assigning them a FrameNet frame; a suitable frame consists of a schematic description of the situation denoted by the synset by means of a set of frame elements spelling out the predicates' participants and props. This type of information is supplemented by assigning verb synsets a VerbNet class and the semantic roles associated with it.

The three resources have been aligned automatically using existing mappings, resulting in the assignment of FrameNet frames to 4,306 verb synsets. This number has been expanded by applying further mapping and validation procedures (Leseva and Stoyanova, 2020), leading to 13,104 automatic frame-to-synset alignments, of which more than 6,500 have been manually validated. VerbNet class-to-FrameNet frame alignments are taken as provided by previous initiatives and have not been separately validated (cf. Stoyanova and Leseva (2023) for a more detailed account of the mappings employed). However, in the exploration of syntactic patterns (see below) certain revisions and corrections have been undertaken.

After the semantic mapping, we go on to align the conceptual components of FrameNet and VerbNet, i.e. FrameNet's frame elements to VerbNet's semantic roles, by perfecting and supplementing previously implemented mappings between elements and roles cf. (Stoyanova and Leseva, 2023).

This step further enables the linking of the syntactic patterns associated with the units in FrameNet, VerbNet and WordNet, by unifying their representation and by matching the corresponding patterns at the level of syntactic groups. The alignment of the semantic components and their syntactic realisations is essential for better exploiting the abundance of information across resources and shedding light on cross-resource similarities, discrepancies and inconsistencies.

The syntactic patterns facilitate the extraction of examples illustrating the use of verb synset literals in corpora and their semantic characterisation through the association of the syntactic groups with the semantic components (frame elements or semantic roles) and can be employed in various tasks requiring semantic and syntactic description.

2 Components of the Conceptual Description

The alignment between WordNet, FrameNet and VerbNet results in a rich semantic and syntactic description of verbs in terms of:

(i) a set of semantic relations between verbs (lexical entries), including hypernymy and hyponymy,

[1] https://wordnet.princeton.edu/
[2] https://framenet.icsi.berkeley.edu/fndrupal/

synonymy, causativity, etc.; as well as derivational and morphosemantic relations between verb and noun synsets;

(ii) frames, frame elements and semantic restrictions associated with FrameNet lexical units and assigned to WordNet synsets, thus providing detailed valency patterns for the syntactic realisation of the frame elements for each verb (in the form of annotated sentences);

(iii) a set of frame-to-frame relations, which are translated into relations of inheritance, specialisation, etc. both between pairs of frames and between pairs of frame elements;

(iv) verb classes, predicate-argument structures (semantic role configurations), selectional restrictions and syntactic patterns realising the arguments of the verbs pertaining to the classes defined in the VerbNet lexicon which are also assigned to WordNet synsets and literals;

(v) aligned VerbNet classes and FrameNet frames providing correspondence between roles and frame elements applicable to lexical units.

Figure 1) exemplifies the successful mapping of the hierarchical structure of FrameNet and WordNet and their coarser-grained correspondence in VerbNet. In particular, the example illustrates a hypernym–hyponym pair of synsets, with the appropriate FrameNet frames assigned to them, which are themselves related by means of an inheritance relation (Cause_change_of_position_on_a_scale being an elaboration of the mother frame Cause_change). Both synsets are described by the *other_cos-45.4* class in VerbNet; respectively, for these particular synsets a correspondence between the pair of FrameNet frames and the *other_cos-45.4* VerbNet class is established.

3 Syntactic Description

We focus on mapping syntactic patterns from the resources which aims at providing a syntactic layer to the conceptual description of verbs in WordNet. We use SemCor (current version 3.0) (Miller et al., 1993; Landes et al., 1998) to extract usage examples for the syntactic patterns in which literals in the corresponding synsets appear in corpora. The extracted examples in English are analysed with a view to the differences in the syntactic patterns applicable to different literals.

Example. Corpus data for the FN frame – VN class pair <Becoming_aware : see-30.1> and the

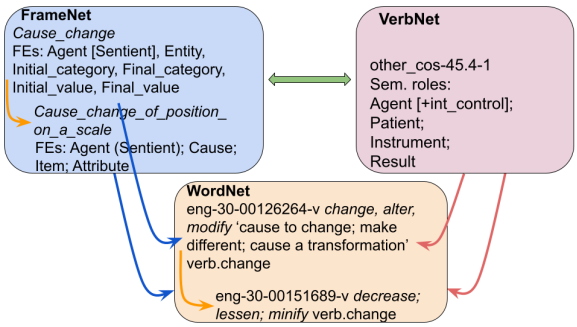

Figure 1: Frames inheritance (Cause_change → Cause_change_of_position_on__scale) reflected in synset hypernym / hyponym relations (*change* → *decrease*)

synset eng-30-00598954-v verb.cognition *learn; hear; get word; get wind; pick up; find out; get a line; discover; see* 'get to know or become aware of, usually accidentally' aligned with this FN – VN pair.

Most frequent aligned patterns:

VN: NP (Experiencer) V NP (Stimulus)
FN: NP (Cognizer) V NP (Phenomenon)

VN: NP (Experiencer) V PP.stimulus[about,of] (Stimulus)
FN: NP (Cognizer) V PP (Phenomenon)

VN: NP (Experiencer) V S[that,wh*,∅] (Stimulus)
FN: NP (Cognizer) V S[that,wh*,∅] (Phenomenon)

Corpus examples:

*We **learned** this year that our older son, Daniel, is autistic.*
*Have you ever **heard** of thuggee?*
*We had merely been **discovered** by the pool sharks.*
*We want to **find_out** who knew about it.*
*Williams is **learning** the difficulties of diplomacy rapidly.*
*I was anxious to **hear** about those dazzling days on the Great_White Way.*
*What obsessions had she **picked_up** during these long nights of talk?*

While the aggregate of all the patterns of the literals in a synset – as well as the aggregate of the patterns of all the synsets associated with a given frame – are a valuable point of departure for any analysis (especially in order to make up for the cases where few examples of distinct patterns are found for the individual verbs), we consider with the greatest priority the patterns associated with a particular verb, as syntactic differences may be observed across literals denoting the same sense.

4 Towards a Cross-Language Description

The combination of semantic and syntactic information is seen as a possible way of transferring knowledge across languages (especially under-resourced ones) by relying on the universality of the semantic description.

In our work on describing the conceptual and syntactic properties of Bulgarian verbs, we have found the applicability of the conceptual description encoded in the FrameNet frames (the same holds to a lesser degree for VerbNet classes, the main reason for this being the lack of definitions of the classes) to be to a great extent language-independent and transferrable cross-linguistically, even if in some cases corrections may be necessary. Given the fact, that the alignment between equivalent senses in the wordnets developed for different languages is ensured by means of shared identification numbers with the original Princeton WordNet, the alignment with FrameNet and VerbNet is also mappable across languages via WordNet[3].

Some of the syntactic patterns can be directly transferred to other languages, while others need adaptation (e.g., considering prepositions or other lexical information), or might be language specific (e.g., constructions such as 'THERE (Aux) is / are ...'). Even so, English syntactic patterns taken both from FrameNet and from VerbNet have been found to be a valuable point of departure in the analysis of Bulgarian syntactic data: they help establishing what is valid or invalid in Bulgarian by comparing the syntactic properties of the Bulgarian verbs to those of their English counterparts and the example sentences in the resources.

In addition, annotated corpora also provide material for studying the syntactic properties of verbs, including syntactic patterns realising the predicate-argument structures, word order specifics, alternations and other variations. Our future work envisages the employment of the UD framework[4] in syntactic annotation.

The syntactic descriptions are applicable to semantic role labelling, word sense disambiguation and other NLP tasks; the accurate identification of situation participants may benefit areas such as information extraction, text recognition and generation, question answering, machine translation.

[3]For a list of existing wordnets in the world, see http://globalwordnet.org/resources/wordnets-in-the-world/.

[4]https://universaldependencies.org/

References

Collin F. Baker. 2008. FrameNet, present and future. In *The First International Conference on Global Interoperability for Language Resources*, Hong Kong. City University, City University.

Collin F. Baker and Christiane Fellbaum. 2009. WordNet and FrameNet as complementary resources for annotation. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 125–129, Suntec, Singapore. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *COLING-ACL '98: Proceedings of the Conference. Montreal, Canada*, pages 86–90.

Christiane Fellbaum, editor. 1998. *WordNet: an electronic lexical database*. MIT Press, Cambridge, MA.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. Language resources and evaluation. *Communications. ACM*, 42(1):21–40.

Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon. PhD Thesis*. Computer and Information Science Dept., University of Pennsylvania. Philadelphia, PA.

Shari Landes, Claudia Leacock, and Rundee Tengi. 1998. Building semantic concordances. In Christiane Fellbaum, editor, *WordNet: An electronic lexical database*, chapter 8.

Svetlozara Leseva and Ivelina Stoyanova. 2020. Beyond lexical and semantic resources: linking WordNet with FrameNet and enhancing synsets with conceptual frames. In *Towards a Semantic Network Enriched with a Variety of Semantic Relations*. Prof. Marin Drinov Academic Publishing House of the Bulgarian Academy of Sciences.

George A. Miller. 1995. WordNet: A lexical database for English. *Commun. ACM*, 38(11):39–41.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. 1993. A semantic concordance. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.

Gerold Schneider. 2012. Using semantic resources to improve a syntactic dependency parser. In *LREC 2012 Conference Workshop "Semantic Relations II", Istanbul, Turkey, 22 May 2012*, pages 67–76.

Ivelina Stoyanova and Svetlozara Leseva. 2023. Expanding the conceptual description of verbs in wordnet with semantic and syntactic information. In *Proceedings of the 12th Global Wordnet Conference, Donostia-San Sebastian 23 - 27 January 2023*.