Expanding the Conceptual Description of Verbs in WordNet with Semantic and Syntactic Information

Ivelina Stoyanova
Department of
Computational Linguistics
Institute for Bulgarian Language
Bulgarian Academy of Sciences
iva@dcl.bas.bg

Svetlozara Leseva
Department of
Computational Linguistics
Institute for Bulgarian Language
Bulgarian Academy of Sciences
zarka@dcl.bas.bg

Abstract

This paper describes an ongoing effort towards expanding the semantic and conceptual description of verbs in WordNet by combining information from two other resources, FrameNet and VerbNet, as well as enriching the verbs' description with syntactic patterns extracted from the three resources. The conceptual description of verb synsets is provided by assigning a FrameNet frame which provides the relevant set of frame elements denoting the predicate's participants and props. This information is supplemented by assigning a VerbNet class and the set of semantic roles associated with it. The information extracted from FrameNet and VerbNet and assigned to a synset is aligned (semi-automatically with subsequent manual corrections) at the following levels: (i) FrameNet frame: VerbNet class; (ii) FrameNet frame elements: VerbNet semantic roles; (iii) FrameNet semantic types and restrictions: VerbNet selectional restrictions. We then link 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 the better exploitation of the abundance of information across resources, including shedding light on cross-resource similarities, discrepancies and inconsistencies. The syntactic patterns can 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 components of semantic description (frame elements or semantic roles) and can be employed in various tasks requiring semantic and syntactic description. The resource is publicly available to the community. The components of the conceptual description are visualised showing the links to the original resources each component is drawn from.

1 Introduction

The paper focuses on describing an effort at obtaining a rich semantic and syntactic description of verbs in WordNet through mapping other lexical and conceptual resources to it (FrameNet and VerbNet, in particular). This has been achieved through aligning corresponding elements of the semantic and syntactic description of the entities in these resources. We rely on existing alignments between the resources – part of the verbs in WordNet have been assigned a FrameNet frame and/or a VerbNet class on the basis of equivalent or similar meaning. After the basic units of the resources have been aligned, we implement procedures for mapping their constituent parts: frame elements with semantic roles, syntactic groups with syntactic groups or syntactic positions. This type of alignment makes it possible to study the commonalities and differences in and possibly to perfect the representation of verbs across resources and languages, on the one hand, and to obtain a richer and more reliable description for the purposes of tasks in computational linguistics, on the other.

2 Related Work

In recent years, significant efforts have been invested in harnessing the strengths of lexical, conceptual and syntactically-oriented resources through mapping them on various levels. Such efforts include works on mapping WordNet, FrameNet and VerbNet (the earliest attempt probably being made by Shi and Mihalcea (2005)) or different combinations of these resources resulting in combined resources, such as WordFrameNet[1] by Laparra and Rigau (2010) and MapNet[2] by Tonelli and Pighin (2009), other other FrameNet-to-WordNet mappings, such as the one by Ferrández et al. (2010). The further enhancement of these

[1]http://adimen.si.ehu.es/web/WordFrameNet
[2]https://hlt-nlp.fbk.eu/technologies/mapnet

resources with others has resulted in the emergence of Semlink[3] (Palmer, 2009), which unifies WordNet, FrameNet and VerbNet with PropBank, and Semlink+ that brings in a mapping to Ontonotes (Palmer et al., 2014). Efforts such as the SynSemClass lexicon[4] centre not on any of the discussed resources, but on a different one (in this case the Vallex dictionary family), which is further enriched with conceptual and syntactic information from external semantic resources (Urešová et al., 2020a), including linking to FrameNet, WordNet, VerbNet, OntoNotes and PropBank, as well as the Czech VALLEX.

As the alignment between resources is limited by the overlap between the lexis covered by them, a major effort has been to expand the coverage of the mapping across resources by way of generalisation and transfer of existing descriptions from already described items (literals, synsets, lexical units, verbs in verb classes, etc.) to other units that share the same semantic and syntactic properties. VerbAtlas[5], proposed by Di Fabio et al. (2019) has adopted a representation of synsets as clusters with prototypical argument structures presented as frames (to a large extent inspired by VerbNet roles and semantic restrictions). The clustering leads to a significant expansion encompassing the entire verb inventory (13,767 synsets).

Another approach, adopted by (Leseva et al., 2018a) and further refined in (Leseva and Stoyanova, 2019, 2020), involves the mapping of FrameNet frames to WordNet synsets on the basis of the inheritance of conceptual features in hypernym trees, i.e., by assigning frames from hypernyms to hyponyms where possible and implementing a number of validation procedures based on the structural properties of the two resources, primarily the relations encoded in them. This has resulted in 13,104 automatic alignments, of which 6,000 have been validated and corrected manually in the framework of this project and previous initiatives.

Another venue of research has been to map relevant information representing fragments of meaning associated with lexical units across resources, especially essential components of the semantic and the syntactic description such as semantic roles or their counterparts in the respective resources (e.g. frame elements, argument positions, valency slots). Alignments at the verb arguments' level have been carried out as part of the Semlink project and its more recent version Semlink 2.0. (Stowe et al.). The alignments described there include PropBank to VerbNet mappings (PropBank roleset – VerbNet senses, PB arguments – VerbNet semantic roles) as well as VerbNet to FrameNet mappings (VerbNet senses – FrameNet frames, VerbNet semantic roles – FrameNet frame elements). Another similar task, which makes use of the linking of various semantic resources (FrameNet, WordNet, VerbNet, OntoNotes and PropBank), has been implemented in the development of the SynSemClass Lexicon (Urešová et al., 2020a,b): the more general SynSemClass valency slots have been mapped to relevant FrameNet frame elements.

It has long been discussed that combining WordNet and other lexical and conceptual resources such as FrameNet produces a more complete semantic and syntactic representation of the meaning lexical entries (Baker and Fellbaum, 2009; Schneider, 2012) which expands the possible application of the resources for the purpose of syntactic and semantic parsing.

Our current effort builds on our previous work described in (Leseva et al., 2018b,a) and further refined in (Leseva and Stoyanova, 2019, 2020)[6], and proceeds onwards. Our interests lie in both: (i) expanding the alignment between the most lexically populated resource, WordNet, the rich conceptual apparatus and the more generalised argument-structure descriptions of FrameNet and VerbNet, respectively, and the syntactic descriptions available in the three resources; (ii) mapping the basic building blocks across resources, where possible, i.e. frame elements and semantic roles and respectively – their syntactic expressions.

In this paper we particularly focus on the latter: extending the description of WordNet verbs by mapping semantic and conceptual components of the description extracted from the three resources employed in the study, and supplementing it with syntactic patterns by combining and aligning the available syntactic information.

The proposed enhancements are directed to: (a) improving the existing mappings by aligning FrameNet frames and VerbNet verb classes assigned to the same synset; (b) enhancing the conceptual description of synsets with additional infor-

[3]https://verbs.colorado.edu/semlink/

[4]https://ufal.mff.cuni.cz/synsemclass

[5]http://verbatlas.org/

[6]The resource is distributed as a standoff file under CC by 4.0 license: https://dcl.bas.bg/semantic-relations-data/.

mation about the syntactic realisation of FrameNet frame elements and VerbNet semantic roles; and (c) suggesting further procedures for verification and improvements of conceptual descriptions of verb synsets in WordNet.

3 Lexical and Conceptual Resources

Below we describe in brief the used resources and how they are integrated with each other.

3.1 WordNet

WordNet[7] (Miller, 1995; Fellbaum, 1998) is a large lexical database that represents comprehensively conceptual and lexical knowledge in the form of a network whose nodes denote cognitive synonyms (synsets) linked by means of a number of conceptual-semantic and lexical relations such as hypernymy, meronymy, antonymy, etc. Of the three resources employed in this work, WordNet provides the greatest lexical coverage; the verbs represented in it are organised in 14,103 synsets (including verb synsets specific for Bulgarian). We use both the Princeton WordNet and the Bulgarian WordNet, which are aligned at the synset level by means of unique synset identifiers.

WordNet verb synsets are supplied with generalised sentence frames which specify the subcategorisation features of the verbs in the synset by indicating the kinds of sentences they can occur in (Fellbaum, 1990, 1999). The main purpose of these frames is to allow the identification of synsets sharing one or more syntactic frames, which facilitates the analysis of the syntactic realisation of semantically related verbs (e.g., verbs belonging to the same semantic class expressed by the semantic primitive, or synsets in the same hypernym tree).

There are 35 generic sentence frames illustrating the use of the literals in the synsets[8], e.g., (8) Somebody —-s something, (16) Somebody —-s something from somebody, (22) Somebody —-s PP, etc. As the syntactic frames describe the properties of individual verbs (literals), the generalised frames in WordNet can be applicable to all or only some of the literals in the synset.

Besides the rich lexical description (glosses, examples, semantic primitive) and the encoded relations, WordNet's main contribution to this work is the rich lexical coverage of verbs, including information about the membership of synsets to the so-called base concepts – a cross-lingual selection of synsets which we use as an approximation (together with other selection criteria) for establishing a set of general lexis verbs.

3.2 FrameNet

FrameNet[9] (Baker et al., 1998; Baker, 2008) is a lexical semantic resource which couches lexical and conceptual knowledge in the terms of frame semantics. Frames are conceptual structures describing types of objects, situations, or events along with their components (frame elements) (Baker et al., 1998; Ruppenhofer et al., 2016). Depending on their status, frame elements (FEs) may be core, peripheral or extra-thematic (Ruppenhofer et al., 2016). In terms of the conceptual description, we deal primarily with core FEs, which instantiate conceptually necessary components of a frame, and which in their particular configuration make a frame unique and different from other frames.

FrameNet frames represent conceptual rather than lexical knowledge and thus are to a large extent language independent. FrameNet frames apply at synset (sense) level and in most cases cover all literals. Each frame is associated with a set of syntactic patterns showing the realisation of different configurations of the FEs in sentences. Here, we consider the configurations of core FEs which describe the obligatory participants in the situation. Example 1 shows the FrameNet frame *Cause_motion* and its description.

Example 1. FrameNet frame *Cause_motion* and its description.
Frame definition: An Agent causes a Theme to move from a Source, along a Path, to a Goal.
Frame elements: Agent (Sentient); Cause; Theme; Source; Goal; Path; Initial_state; Area; Result.
Syntactic patterns (total of 116 patterns):
NP (Agent) V NP (Theme);
NP (Agent) V INI (Goal) NP (Theme);
NP (Agent) V PP[off] (Source) NP (Theme);
NP (Agent) V PP[into] (Goal) PP[across] (Path) NP (Theme);
NP (Theme) V PP[around] (Area) PP[by] (Cause);
NP (Theme) V PP[by] (Agent) NP (Path); etc.
Examples: She$_{\text{Agent}}$ THREW {her shoes}$_{\text{Theme}}$ {into the dryer}$_{\text{Goal}}$.
Croquet$_{\text{Theme}}$ was PUSHED out$_{\text{Source}}$ by tennis$_{\text{Cause}}$.

[7] https://wordnet.princeton.edu/
[8] https://wordnet.princeton.edu/documentation/wninput5wn
[9] https://framenet.icsi.berkeley.edu/fndrupal/

The storm$_{\text{Cause}}$ TOSSED the sailor$_{\text{Theme}}$ from the boat$_{\text{Source}}$.

3.3 VerbNet

VerbNet (Kipper-Schuler, 2005; Kipper et al., 2008) is a hierarchical network of English verbs which represents their syntactic and semantic patterns[10]. It is organised into 274 classes extending Levin's classification (Levin, 1993) through refining and adding subclasses so as to provide better syntactic and semantic coherence among members of a class. VerbNet explicitly projects semantic relations onto syntactic structures and encodes information about thematic roles, arguments' selectional restrictions and syntactic frames. While the syntactic dimension of the resource is more specific to English, the semantic roles and the selectional restrictions employed provide well-motivated semantic generalisations.

Each VerbNet class is associated with a number of syntactic patterns which have a generalised form and express the configurations in which the thematic roles appear in sentences. Unlike FrameNet patterns, the VerbNet patterns do not account for syntactic transformations such as passivisations, etc. Example 2 shows the VerbNet class *run-51.3.2* with its corresponding description.

Example 2. VerbNet class *run-51.3.2* and its description.

Roles: Theme [+animate | +machine]; Trajectory [+concrete]; Initial_Location [+concrete]; Destination [+concrete].
Syntactic patterns (total of 6 patterns): NP V
NP V PP.location
NP V PP.location
There V PP NP
There V NP PP
PP.location V NP
Syntax: Theme VERB
Examples: The horse$_{\text{Theme}}$ RAN.
The horse$_{\text{Theme}}$ RAN to the barn$_{\text{Destination}}$.
The horse$_{\text{Theme}}$ JUMPED {over the fence}$_{\text{Trajectory}}$.
{Out of the box}$_{\text{Initial\_location}}$ JUMPED {a little white rabbit}$_{\text{Theme}}$.

[10] https://verbs.colorado.edu/verbnet/

4 Alignment between Resources

4.1 Mapping VerbNet classes and FrameNet frames to WordNet synsets

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, 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 hierarchical and non-hierarchical relations, which are translated into relations of inheritance, specialisation, etc. both between pairs of frames and between pairs of frame elements; these relations are also reflected in the alignment between WordNet synsets and FrameNet frames;

(iv) verb classes, predicate-argument structures (in the form of 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 semantic roles and frame elements applicable to lexical units.

By aligning the lexical items in FrameNet and VerbNet we focus particularly on mapping core frame elements as they are most likely to represent a verb's arguments and hence – constitute counterparts of the semantic roles. Differences between frames' core FEs sets and corresponding predicate argument structures reveal valuable language- and resource-specific features of the semantic and syntactic description.

The three resources have been aligned automatically using existing mappings (see Section 2) on top of which further mapping procedures have been implemented. In particular, the following resources have been employed: a mapping of the VerbNet 3.4 verb classes to WordNet 3.0 synsets, as well as two types of mappings of the frames in FrameNet and the synsets in WordNet 3.0[11]: indirectly via SemLink and directly through the system described by Laparra and Rigau (2010). These mappings have resulted in the assignment of FrameNet frames to 4,306 verb synsets.

The number of synsets that are assigned a FrameNet frame have been supplemented using the expanded synset-to-FrameNet frame mapping described in (Leseva et al., 2018a) and further refined in (Leseva and Stoyanova, 2019, 2020) which involves the mapping of FrameNet frames to WordNet synsets on the basis of the inheritance of conceptual features in hypernym trees, i.e., by assigning frames from hypernyms to hyponyms where possible and implementing a number of validation procedures based on the structural properties of the two resources, primarily the relations encoded in them. This has resulted in 13,104 automatic alignments, of which over 6,000 have been validated and corrected manually in the framework of this project and previous initiatives. VerbNet class-to-FrameNet frame alignments have not been separately validated.

Example 3 represents the different blocks of information obtained from the three resources through the mapping. 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.

Example 3. Alignment between FrameNet frames and VerbNet classes (Figure 1).

(a) **WordNet synset:** *eng-30-00126264-v change; alter; modify* verb.change 'cause to change; make different; cause a transformation'
FrameNet frame: *Cause_change*: Agent (Sentient); Entity (Entity); Initial_category; Final_category; Initial_value; Final_value; Attribute [unexpressed]; Cause [unexpressed]

[11]Additional mappings between WordNet versions were also involved.

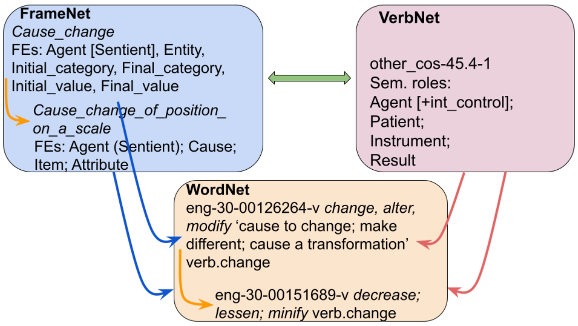

Figure 1: Frames inheritance (Cause_change → Cause_change_of_position_on__scale) reflected in synset hypernym / hyponym relations (*change* → *decrease*)

VerbNet class: *other_cos-45.4*: Agent [+int_control]; Patient; Instrument; Result

(b) **WordNet synset:** *eng-30-00151689-v decrease; lessen; minify* 'make smaller'
FrameNet frame:*Cause_change_of_position _on_a_scale*: Agent (Sentient); Cause; Item; Attribute
VerbNet class: *other_cos-45.4*: Agent [+int_control]; Patient; Instrument; Result.

4.2 Mapping FrameNet frame elements to VerbNet semantic roles

The mapping between FrameNet frame elements and VerbNet semantic roles is based on extracting semantic information from the two resources: (i) establishing correspondence between frame elements and semantic roles, where possible, and (ii) inferring knowledge from the structure of FrameNet, many frame elements being more specific than VerbNet semantic roles. The former case (i) involves heuristic procedures such as establishing identity, similarity or correspondences in the naming of elements and roles, and possibly comparing their definitions. Example 4 shows a FrameNet frame–VerbNet class alignment where the frame *Breathing* has been automatically aligned to the VerbNet class *breathe-40.1.2*. The frame elements and semantic roles Agent and Source have been aligned on the basis of their identical names. In addition, the frame element Goal has been aligned to the role Destination based on established general (frame/class non-specific) correspondences in the naming conventions adopted in the two resources. The latter case (ii) involves knowledge about the relations between more general and more concrete frame elements, which is obtained from a shallow hierarchy of frame elements based on inheritance

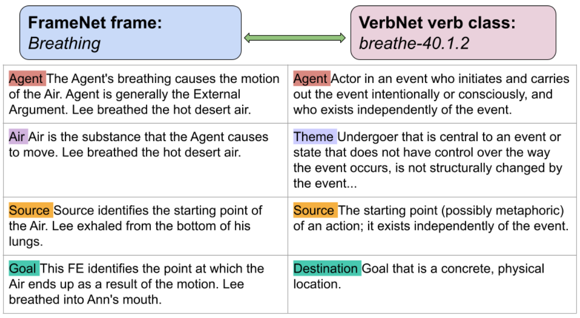

Figure 2: Frame inheritance reflected in the hypernym/hyponym relation between synsets.

between frames (Leseva et al., 2018b). *Breathing* inherits its properties from a series of frames that form a chain of inheritance from a more specific to a more general frame – *Breathing* > *Fluidic_motion* > *Motion*. The frame-to-frame relations help identify corresponding inheritance relations between the relevant frame elements in these frames: *Air* > *Fluid* > *Theme*. The FE-to-FE relations are obtained semi-automatically based on their syntactic expression and/or similarity of definitions. After establishing inheritance chains, we try to map the more general FEs to relevant roles in the semantic role set of the VerbNet verb class aligned with the respective frame. As a result, the Breathing *Air* is mapped to the *Theme* in the VerbNet class *breathe-40.1.2*.

Example 4. FrameNet frame *Breathing* aligned to VerbNet class *breathe-40.1.2* along with the alignment between frame elements and semantic roles (Figure 2).

WordNet synset: *eng-30-00001740-v breathe; take a breath; respire; suspire* verb.body 'draw air into, and expel out of, the lungs'
FrameNet frame: *Breathing*: Agent (Sentient); Air; Source; Goal
VerbNet class: *breathe-40.1.2*: Agent [+int_control]; Theme; Source; Destination

While often there is no full frame-to-verb class equivalence, the greater the correspondence between the frame elements and semantic roles in terms of their number and semantics, the better the match is.

5 Corpus Resources

The semantically annotated corpus SemCor (current version 3.0) (Miller et al., 1993; Landes et al., 1998) is compiled by the Princeton WordNet team and covers texts excerpted from the Brown Corpus. SemCor is supplied with POS and grammatical tagging and all open-class words (both single words and multiword expressions, as well as named entities) are semantically annotated by assigning each word a unique WordNet sense (synset ID).

BulSemCor (Koeva et al., 2010, 2011) has been generally modelled on the SemCor methodology and structure. While only open-class words are annotated with WordNet senses in SemCor, all lexical units in BulSemCor have been annotated; for that purpose the Bulgarian wordnet has been expanded with closed-class words (Koeva et al., 2010).

We use SemCor and BulSemCor 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. Examples from BulSemCor serve the purpose to provide material for the investigation of the possible syntactic knowledge transfer from English to Bulgarian.

6 Compilation of Syntactic and Semantic Description of Verbs in WordNet

After aligning FrameNet frames to VerbNet classes (assigned to synsets or groups of synsets), and FrameNet frame elements to VerbNet roles, we move towards mapping syntactic patterns from the resources to the end of providing a new, syntactic layer to the conceptual description of the verbs in WordNet. In order to make the alignment between the patterns obtained from VerbNet and FrameNet as precise as possible, we perform this procedure at the literal level and then transferred onto the the FrameNet frame and VerbNet class pair, i.e. for each literal in a synset which is mapped to a lexical unit in FrameNet and an entry in VerbNet, the corresponding patterns from the two resources are aligned according to several criteria. These include:

- correspondence in the number of elements or roles expressed in a syntactic pattern;
- correspondence between the frame element and the semantic role mapped to it as part of the previous task;
- correspondence in the syntactic restrictions (PP heads, clause types or subordinating elements) defined for the mapped frame elements and semantic roles;

- correspondence between the syntactic expression of each mapped frame element and semantic role – both in terms of the type of syntactic phrase by means of which they are expressed (NP, PP, etc.), and the syntactic position in which they are projected (e.g. subject, object).

The syntactic pattern alignment procedure is implemented as a set of mapping rules. As a result of their application, we obtain a list of the equivalent syntactic models for a given FrameNet frame and VerbNet class pair (Examples 5, 6 and 7). Where no correspondence is discovered, the table cell is marked as NONE.

Example 5. Aligned syntactic patterns for the WordNet synset *eng-30-00001740-v breathe; take a breath; respire; suspire*, FrameNet frame *Breathing* and the VerbNet class *breathe-40.1.2*. (FN syntactic patterns with frequency of 3+ are labelled by a *.)

WN	Somebody	–s	
VN	NP(Agent)	V	
FN	*NP.Ext(Agent)	V	
WN	Somebody	–s	
VN	NP(Agent)	V	PP.destination[on,onto] (Destination)
FN	*NP.Ext(Agent)	V	INC(Air) PP[in,on](Goal)
WN	Somebody	–s	something
VN	NP(Agent)	V	NP(Theme)
FN	*NP.Ext(Agent)	V	NP(Air)
WN	Somebody	–s	something
VN	NP(Agent)	V	NP(Theme) PP.destination[on,onto] (Destination)
FN	*NP.Ext(Agent)	V	NP(Air) PP[in,on](Goal)
WN	Somebody	–s	
VN	NONE		
FN	NP.Ext(Agent)	V	INC(Air) PP[down] (Path)
	NP.Ext(Agent)	V	INC(Air) PP[in] (Place)
	*NP.Ext(Agent)	V	INC(Air) PP[at] (External_cause)
	*NP.Ext(Agent)	V	INC(Air) AVP (Manner)
	NP.Ext(Agent)	V	INC(Air) PP[by,without] (Means)
	NP.Ext(Agent)	V	INC(Air) PP[as] (Depictive)
	NP.Ext(Agent)	V	INC(Air) PP[from,out] (Source)
	NP.Ext(Agent)	V	INC(Air) VPto (Purpose)
WN	Somebody	–s	something
VN	NONE		
FN	NP.Ext(Agent)	V	NP(Air) PP[down] (Path)
	NP.Ext(Agent)	V	NP(Air) AVP (Manner)
	NP.Ext(Agent)	V	NP(Air) PP[in](Goal) PP[from,out] (Source)
	NP.Ext(Agent)	V	NP(Air) PP[in](Goal) PP[through] (Instrument)

Example 5 shows the alignment of the syntactic patterns between the frame *Breathing* and the class *breathe-40.1.2* following the mapping between the frame elements and semantic roles (Agent – Agent, Air – Theme, Source – Source, Goal–Destination). Misalignment occurs in the cases of additional semantic roles that are not considered core FEs (e.g., Path, Manner, etc.) which have no correspondence to VerbNet roles and participants in WordNet basic sentence frames.

Example 6. Aligned syntactic patterns for the FrameNet frame *Killing* and the VerbNet class *murder-42.1* for the synset eng-30-01323958-v *kill* 'cause to die; put to death, usually intentionally or knowingly'.

WN	Somebody	–s	somebody	
VN	NP(Agent)	V	NP(Patient)	
FN	NP.Ext(Killer)	V	NP.Obj(Victim)	
WN	Somebody	–s	somebody	
VN	NP(Agent)	V	NP(Patient)	{with} PP.instrument (Instrument)
FN	NP.Ext(Killer)	V	NP.Obj(Victim)	PP[with].Dep (Instrument)
WN	Something	–s	somebody	
VN	NP.instrument (Instrument)	V	NP(Patient)	
FN	NP.Ext (Instrument)	V	NP.Obj(Victim)	
WN	Something	–s	somebody	
VN	NONE			
FN	NP.Ext(Cause)	V	NP.Obj(Victim)	

Example 7. Aligned syntactic patterns for the FrameNet frame *Killing* and the VerbNet class *suffocate-40.7* (e.g., *asphyxiate, choke, suffocate*, etc.).

WN	Somebody	–s	somebody	
VN	NP(Agent)	V	NP(Patient)	
FN	NP.Ext(Killer)	V	NP.Obj(Victim)	
WN	Somebody	–s	somebody	
VN	NP(Agent)	V	NP(Patient)	{with} PP.instrument (Instrument)
FN	NP.Ext(Killer)	V	NP.Obj(Victim)	PP[with].Dep (Instrument)
WN	Something	–s	somebody	
VN	NONE			
FN	NP.Ext (Instrument)	V	NP.Obj(Victim)	
WN	Something	–s	somebody	
VN	NONE			
FN	NP.Ext(Cause)	V	NP.Obj(Victim)	
WN	Somebody	–s	somebody	
VN	NP(Agent)	V	NP(Patient)	{to, into} PP.result(Result)
FN	NONE			

Examples 6 and 7 show different degrees of misalignment between the syntactic patterns of the corresponding frames and verb classes. The frame *Killing* allows for the Instrument to appear as an external argument NP which matches a syntactic pattern within the verb class *murder-42.1* but not the verb class *suffocate-40.7*. Further, while the verbs evoking the frame *Killing* incorporate the result (the death of the Patient / Victim), the verb class *suffocate-40.7* also allows for a different Result as shown in the last row of the table in Example 7 (e.g., *suffocate to/into unconsciousness*).

Further, in order to increase the number of mapped frames we generalise some unmapped FrameNet frames by excluding optional or unexpressed arguments, thus reducing the pattern to a more basic form.

The asymmetries in the syntactic patterns covered by matched FrameNet frames and VerbNet classes for particular WordNet synsets are indicative of the need for more detailed syntactic analysis and the study of both the alignment between frame elements and semantic roles and their syntactic realisation.

Example 8 shows sentences featuring the literals from a given synset which are extracted from SemCor.

Example 8. Corpus data for the FN frame – VN class pair <Becoming_aware : see-30.1> on 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'

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*,$\emptyset$] (Stimulus)
FN: NP (Cognizer) V S[that,wh*,$\emptyset$] (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?*

As illustrated by the examples: (a) some literals appear more frequently in the data while others do not appear at all (e.g., *get wind*) and for the latter we cannot draw any conclusions; (b) some literals have a restricted number of patterns applicable to them (e.g., multiword expressions such as **get word** cannot have a Phenomenon as a direct object) or accept particular lexical entries (e.g., prepositions *hear of* but **pick_up of*).

7 Results

The processing of the data included the following key procedures:

(1) Identifying FrameNet-frame-to-WordNet-synsets alignments and selecting only manually validated ones so as to ensure the quality of the dataset.

(2) Identifying VerbNet-class-to-WordNet-synsets alignments. Out of these, as a matter of validation, we select only those that have been aligned to FrameNet frames.

(3) The resulting dataset covers 1,121 WordNet synsets and a total of 5,264 verb literals. Each synset is assigned a pair <FN frame : VN class>. There are a total of 329 such pairs involving 195 FrameNet frames and 165 VerbNet classes. As already illustrated (e.g., Example 3), there are VerbNet classes that correspond to more than one pair of alignments, as well as FrameNet frames that correspond to more than one class (e.g., Examples 6 and 7).

The VerbNet classes represented in the dataset include 32 unique semantic roles which are matched to a total of 217 FrameNet frame elements.

The synsets in the dataset cover 29 (out of the 35) generalised WordNet sentence frames. These are aligned to 451 VerbNet syntactic patterns and 13,884 FrameNet syntactic patterns. The greater number of FrameNet syntactic realisations is due to: (a) the large number of peripheral and extra-thematic frame elements[12] and the variety of configurations they enter in the different realisations; and (b) the representations of alternations and variations (e.g., passives, incorporation of FEs, various prepositions in PPs, etc.). The FrameNet patterns

[12] Although we focus on the core FEs, the syntactic patterns include some peripheral and extra-thematic elements with high frequency.

have been filtered based on frequency (of examples exhibiting the pattern included in the FrameNet dataset), which has resulted in 811 FrameNet syntactic patterns with frequency of 3 or more.

The dataset is supplemented with a set of 16,059 corpus examples illustrating the annotated synsets (on average, 14 examples per synset). Additionally, we have also included the usage examples provided in all of the resources – WordNet examples (which are often not full sentences but phrases) and FrameNet and VerbNet illustrative examples.

The newly developed resource containing pairs of a FrameNet frame and a VerbNet class with their corresponding syntactic patterns for realisation of FEs and semantic roles is distributed under a CC by 4.0 license[13].

7.1 Towards Literal-Specific Description

Our efforts are aimed at expanding the description of WordNet synsets towards a complex conceptual and syntactic representation. While the conceptual description applies to a large extent to the whole synsets, the considered syntactic patterns are relevant to individual literals in the synset. The corpus examples provide material to confirm the syntactic patterns valid for certain literals. However, for some literals there are insufficient number of examples or no examples at all. These will require the use of a general corpus with no semantic annotation where ambiguity also needs to be taken into account. However, the syntactic models applying to some of the literals in the synset can serve to extract detailed semantic description of the semantic roles and frame elements co-occuring with the particular use of the verb and its subcategorisation frame, and this knowledge can inform algorithms for synonym detection in a general corpus and identifying verbs belonging to the same synset and analysing their syntactic realisation.

7.2 Towards a Cross-Language Description

Further, efforts can be invested into the cross-language transfer of knowledge in order to develop conceptual and syntactic description of synsets for other languages, especially under-resourced languages such as Bulgarian. For this purpose, once again, we consider the applicability of the conceptual description contained in FrameNet frames and VerbNet classes as largely language-independent, which can be transferred and / or adapted. The syntactic patterns need further examination and filtering in order to match the Bulgarian data. We have extracted a dataset of 6,249 sentences from the BulSemCor corpus containing instances of the synsets under analysis. Some of the syntactic patterns can be directly transferred to Bulgarian, while others need adaptation (e.g., considering prepositions or other lexical information), or are not relevant (e.g., constructions such as 'THERE (Aux) is / are . . . ' which are not found in Bulgarian).

In the future our efforts will be focused on validating the syntactic description for Bulgarian and expanding the dataset of examples in order to provide more linguistic material for reliable decisions on the syntactic realisation of verbs and their subcategorisation frames.

8 Conclusions

In this paper we present a dataset of WordNet synsets supplied with extensive semantic, conceptual and syntactic information obtained by combining (i) WordNet's description and semantic relations with (ii) the conceptual information from the relevant FrameNet frame (including the frame elements and the specific semantic restrictions) and VerbNet class assigned to the synsets and (iii) the syntactic patterns compiled from all the three resources and aligned both in terms of the syntactic realisation and the frame element or semantic role of each component.

The combination of semantic and syntactic information is seen as a possible way to transfer knowledge across languages (e.g., from English to Bulgarian) by relying on the universality of semantic description. Various annotated corpora will be further used in studying the syntactic properties of verbs to the end of: enhancing their applicability to NLP tasks such as semantic role labelling, word sense disambiguation, etc. Another promising venue of research is related to facilitating the more precise identification of the participants in the situations described by verbs, thus enabling better information extraction, text recognition and generation, question answering, machine translation.

9 Acknowledgments

This paper is carried out as part of the scientific programme under the project *Enriching the Semantic Network Wordnet with Conceptual Frames* funded by the Bulgarian National Science Fund (Grant Agreement No. KP-06-N50/1 of 2020).

[13] https://dcl.bas.bg/enriching-wordnet-results/

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.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. Verbatlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, November 3 – 7, 2019*, page 627 – 637. Association for Computational Linguistics.

Christiane Fellbaum. 1990. English Verbs as a Semantic Net. *International Journal of Lexicography*, 3(4):278–301.

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

Christiane Fellbaum, editor. 1999. *WordNet: an Electronic Lexical Database*. MIT Press, Cambridge, MA.

Óscar Ferrández, Michael Ellsworth, Rafael Muñoz, and Collin F. Baker. 2010. Aligning FrameNet and WordNet based on semantic neighborhoods. In *Proceedings of the 7th Conference on Language Resources and Evaluation (LREC 2010), May 17-23, Valletta, Malta*, pages 310 – 314.

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.

S. Koeva, S. Leseva, B. Rizov, E. Tarpomanova, T. Dimitrova, H. Kukova, and M. Todorova. 2011. Design and development of the Bulgarian Sense-Annotated Corpus. In *Proceedings of the Third International Corpus Linguistics Conference (CILC), 7-9 April 2011, Valencia, Spain*, pages 143–150. Universitat Politecnica de Valencia.

Svetla Koeva, Svetlozara Leseva, Ekaterina Tarpomanova, Borislav Rizov, Tsvetana Dimitrova, and Hristina Kukova. 2010. Bulgarian Sense-Annotated Corpus – results and achievement. In *Proceedings of the 7th International Conference of Formal Approaches to South Slavic and Balkan Languages (FASSBL-7)*, pages 41–49.

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

Egoitz Laparra and German Rigau. 2010. eXtended WordFrameNet. In *Proceedings of LREC 2010*, pages 1214–1219.

Svetlozara Leseva and Ivelina Stoyanova. 2019. Enhancing conceptual description through resource linking and exploration of semantic relations. In *Proceedings of 10th Global WordNet Conference, 23 – 27 July 2019, Wroclaw, Poland*, pages 229–238.

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.

Svetlozara Leseva, Ivelina Stoyanova, Hristina Kukova, and Maria Todorova. 2018a. Integration of subcategorisation information in WordNet's relational structure. *Balgarski ezik*, 65(2):11–40.

Svetlozara Leseva, Ivelina Stoyanova, and Maria Todorova. 2018b. Classifying verbs in WordNet by harnessing semantic resources. In *Proceedings of CLIB 2018, Sofia, Bulgaria*.

Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago and London: The University of Chicago Press.

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*.

Martha Palmer. 2009. Semlink: linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*. 9–15.

Martha Palmer, Claire Bonial, and Diana McCarthy. 2014. SemLink+: FrameNet, VerbNet and event ontologies. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929–2014), Baltimore, Maryland USA, June 27, 2014*, pages 13–17. Association for Computational Linguistics.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher. R. Johnson, Collin. F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: extended theory and practice*. International Computer Science Institute, Berkeley, California.

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.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: combining FrameNet, VerbNet and WordNet for robust semantic parsing. In A. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing. CICLing 2005. Lecture Notes in Computer Science*, volume 3406. Springe, Berlin, Heidelbergr.

Kevin Stowe, Jenette Preciado, Kathryn Conger, Susan Brown, Ghazaleh Kazeminejad, James Gung, and Martha Palmer. Semlink 2.0: chasing lexical resources. In *Proceedings of the 14th International Conference on Computational Semantics, pages 222–227 June 17–18, 2021*, pages 222–227. Association for Computational Linguistics.

Sara Tonelli and Daniele Pighin. 2009. New features for Framenet – Wordnet mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL'09), Boulder, USA*.

Zdenka Urešová, Eva Fucíková, Eva Hajičová, and Jan Hajič. 2020a. SynSemClass linked lexicon: Mapping synonymy between languages. In *Proceedings of the Globalex Workshop on Linked Lexicography, Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020*, pages 10 – 19.

Zdenka Urešová, Eva Fučíková, Eva Hajičová, and Jan Hajič. 2020b. Syntactic-semantic classes of context-sensitive synonyms based on a bilingual corpus. In *Human Language Technology. Challenges for Computer Science and Linguistics*, pages 242–255. Springer International Publishing.