

Linked Resources towards Enhancing the Conceptual Description of General Lexis Verbs Using Syntactic Information

Svetlozara Leseva

Institute for Bulgarian Language
Bulgarian Academy of Sciences
zarka@dcl.bas.bg

Ivelina Stoyanova

Institute for Bulgarian Language
Bulgarian Academy of Sciences
iva@dcl.bas.bg

Abstract

The paper describes the linking of three previously aligned resources (FrameNet, VerbNet and WordNet) both by expanding their coverage (by means of enhancing existing alignments) and by mapping elements of the semantic and syntactic description of the lexical items: FrameNet frame elements and VerbNet semantic roles, FrameNet valency patterns and VerbNet syntactic patterns. The study focuses on general lexis verbs as being more representative across languages. After describing the used resources and their interaction, we go on to outline the mapping procedures and the elements of the resulting resource. The discussion sums up the main challenges encountered in carrying out the described tasks.

Keywords: linked resources, FrameNet, VerbNet, WordNet, semantic description, syntactic patterns

The paper deals with linking complementary semantic and syntactic resources (FrameNet, VerbNet and WordNet) through aligning relevant elements of their semantic and syntactic description. We take as a point of departure previously made alignments between these resources where WordNet synsets or synset members have been assigned a FrameNet frame and/or a VerbNet class on the basis of equivalent or similar meaning. Through its membership to a class or a frame a lexical unit (in this case a verb belonging to a verb synset in WordNet) inherits the semantic and syntactic description associated with them. While the syntactic and semantic knowledge from FrameNet and VerbNet informs a rich linguistic description associated with each verb, there are a number of challenges to this approach: such a description includes a lot of information couched in different terms which on the one hand may be redundant and on the other does not provide linking between corresponding elements of meaning; such elements include: the semantic roles (SR)

describing the argument structure of each verb in VerbNet, the frame elements (FEs) in FrameNet, part of which (roughly speaking the core FEs) represent VerbNet SR counterparts; the selectional restrictions defined for semantic roles and the relevant semantic types of FrameNet FEs; the syntactic patterns that are associated with the contextual realisations of the verbs in the two resources.

In this work we describe a linked resource in which not only lexical units are aligned but also the elements of the semantic and the syntactic description associated with them. We use an aligned version of VerbNet and FrameNet and propose a methodology for linking semantic and syntactic knowledge in the resources so as to reduce redundant information and make the best use of both of them. We focus on general lexis verbs selected from WordNet using various criteria.

1 Related Work

Significant efforts have been invested in aligning and in some cases expanding the mapping between semantic and syntactic resources in the past two decades and this interest has been growing in recent years. A number of proposals have brought together the advantages of conceptual and lexical information encoded in resources such as WordNet, FrameNet, VerbNet and others. Such works include the mapping of WordNet, FrameNet and VerbNet by [Shi and Mihalcea \(2005\)](#), the elaboration of WordFrameNet¹ by [Laparra and Rigau \(2010\)](#) and MapNet² by [Tonelli and Pighin \(2009\)](#), the implementation of other FrameNet-to-WordNet mappings, such as the one by [Ferrández et al. \(2010\)](#). More enhanced linked resources include Semlink³ ([Palmer, 2009](#)), which unifies WordNet, FrameNet

¹<http://adimen.si.ehu.es/web/WordFrameNet>

²<https://hlt-nlp.fbk.eu/technologies/mapnet>

³<https://verbs.colorado.edu/semliink/>

and VerbNet with PropBank, and its follow-up Semlink+ that brings in a mapping to Ontonotes (Palmer et al., 2014).

More recently, the SynSemClass lexicon⁴ has marked a distinguishable effort towards combining the rich semantic description in the Vallex dictionary family with conceptual and syntactic information from external semantic resources in order to create a multilingual contextually-based verb lexicon. The aim of the lexicon is to provide a resource of classes of verbs that compares their semantic roles as well as their syntactic properties (Urešová et al., 2020a). In addition, each entry is linked to FrameNet, WordNet, VerbNet, OntoNotes and PropBank, as well as the Czech VALLEX.

VerbAtlas⁵, proposed by Di Fabio et al. (2019), is a hand-crafted lexical semantic resource which represents synsets as clusters with prototypical argument structures presented as frames, to a large extent inspired by VerbNet roles and semantic restrictions.

One of the main concerns related to resource alignment has been the limited coverage. Hence, another line in the research on semantic resource linking has been the expansion of inter-resource coverage. Burchardt et al. (2005) have proposed a method for enriching FN frame-to-WN synset alignment based on exploring the structural features of the two resources. In particular, they study candidate frames evoked by literals (individual members of synsets) related to a target literal through certain semantic relations, such as synonymy, hypernymy, antonymy, and assign weights to them according to the adopted methodology.

Di Fabio et al. (2019) adopt a strategy of clustering WordNet synsets according to semantic similarity and associating them with frames that describe the predicate-argument structure and selectional restrictions of each cluster. While these frames are inspired by VerbNet, the clustering algorithm achieves much better coverage of WordNet synsets as compared with WordNet-VerbNet mappings relying on the lexical correspondence of the units in the two resources.

Another proposal for expansion of the mapping between FrameNet and WordNet proposed by Leseva et al. (2018b) and further refined in Leseva and Stoyanova (2019) makes use of the relational structure of the two resources. The method 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.

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 roset – 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.

In this paper we build upon previous efforts in aligning and expanding the coverage of semantic resources by mapping semantic and syntactic elements of the description of their basic units, in particular: FrameNet frame elements and VerbNet semantic roles (along with the selectional restrictions defined for them in the two resources) and the syntactic patterns associated with the verbs in the respective FrameNet frame and VerbNet class. Instead of using it directly, we employ the mapping provided in Semlink and Semlink+ (Palmer, 2009; Palmer et al., 2014) as a reference set to compare to our own mapping for a couple of reasons: some classes are only marginally corresponding to a given frame so the alignment needs to be considered more carefully; as FrameNet's and VerbNet's descriptions do not always correspond straightforwardly, the semantic roles and frame elements may have one-to-many or many-to-many mappings or actually not be counterparts of each other despite a seeming coincidence or similarity in the names or

⁴<https://ufal.mff.cuni.cz/synsemclass>

⁵<http://verbatlas.org/>

definitions. The availability of a mapping to compare our independent results will make the analysis of debatable cases more reliable.

In addition, we map the syntactic patterns capturing the expression of the semantic roles for the verb classes in VerbNet and the valency patterns for the verbs in a given frame in FrameNet. This procedure is aimed at providing an additional syntactic level of comparison between the two resources that may inform studies and applications both for English and for other languages. The syntactic correspondences are also applicable in semantic role – frame element mapping or mapping validation procedures, especially in cases where the semantic roles and the frame elements are not successfully mapped but have equivalent syntactic expression.

We consider two main research questions:

1. How can we integrate semantic and syntactic information to enhance the conceptual description of WordNet synsets and literals?
2. To what extent is syntactic information language independent and can it be transferred from English to less-resourced languages such as Bulgarian?

The contributions of the paper include:

- Mapping of VerbNet classes and their roles and FrameNet frames and frame elements – although initially relying on existing alignments, we extend them using additional mappings of FrameNet to WordNet which allows us to expand the dataset;
- Enhancing conceptual description of WordNet synsets and literals with syntactic patterns facilitates tasks such as syntactic and semantic parsing and semantic role labelling;
- Mapping of general (largely language-independent) and (language-)specific syntactic patterns for Bulgarian and English allows for cross-linguistic analyses, transferring valid patterns and adapting them for low-resourced language such as Bulgarian with limited resources on valency and syntactic realisation of conceptual frames.

2 Resources

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

2.1 WordNet

WordNet⁶ (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. We use both the Princeton WordNet and the Bulgarian WordNet, which are aligned at the synset level by means of unique synset identifiers.

2.2 FrameNet

FrameNet⁷ (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). 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.

2.3 VerbNet

VerbNet (Kipper-Schuler, 2005; Kipper et al., 2008) is a hierarchical network of English verbs which represents their syntactic and semantic patterns⁸. 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.

⁶<https://wordnet.princeton.edu/>

⁷<https://framenet.icsi.berkeley.edu/fndrupal/>

⁸<https://verbs.colorado.edu/verbnet/>

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. Our focus on general lexis stems from the interest in studying the semantic and syntactic (a)symmetries in the vocabulary cross-linguistically. While we use wordnets for English and Bulgarian, any available wordnets for other languages (aligned at the synset level) can be used instead as at least the semantic components and for a number of languages – a part of the syntactic component may be used both for monolingual and comparative/contrastive research and applications.

FrameNet and VerbNet bring in rich semantic description in terms of aligned inventory of: (i) frames, frame elements and semantic restrictions associated with FN lexical units and detailed valency patterns representing the syntactic realisation of the frame elements for each verb (in the form of annotated sentences); (ii) 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. In implementing the task of 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 – 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.

As we use an expanded synset-to-frame mapping between WordNet and FrameNet (Leseva et al., 2018b; Leseva and Stoyanova, 2019), the number of verbs associated with a FN frame and all the information pertaining to it is larger than in the original mappings. An interesting research question to be tackled in the future is to what extent the indirectly aligned WordNet verbs (especially ones that do not correspond to a lexical unit in FrameNet or VerbNet) may be satisfactorily described semantically and syntactically by means of the information already available in the mapped

resources.

3 Dataset Compilation

The three resources have been mapped automatically using existing mappings or newly designed procedures in such a way that WordNet synsets are assigned corresponding verb classes from VerbNet and frames from FrameNet where possible. The previously implemented mappings have been supplemented and partially validated. In particular, the following have been employed: a mapping of the VerbNet 3.4 verb classes to WordNet synsets, as well as two types of mappings of the frames in FrameNet and the synsets in WordNet: indirectly via SemLink and directly through the system described by Laparra and Rigau (2010). In addition, in order to increase the inter-resource coverage between WN synsets and FN frames, we have used an expanded synset-to-FrameNet frame mapping described in detail in (Leseva and Stoyanova, 2020).

The focus of the study are general lexis verbs in WordNet. We are aiming at compiling a lexical resource of verbs of high frequency and wide usage supplied with conceptual description and syntactic frames. The main source of the description is the information from the FrameNet frame and VerbNet class aligned to the WordNet synset. The resource will serve as a model and can be further expanded to cover other verbs.

3.1 General Lexis Verbs and their Representation in WordNet

First, we identify verbs in WordNet that potentially belong to the general lexis using several criteria:

- verbs labelled as base concepts (BCS) in WordNet;
- verbs with high frequency in the Bulgarian National Corpus (considering the usage of all their senses);
- verbs identified in primary school textbooks in Bulgarian;
- verb senses included in Concepticon;
- verb senses marked with age of acquisition in primary school age;
- verb synsets that have been assigned FrameNet frames with high frequency (50+ verified occurrences assigned to WordNet

synsets), which in most cases have general meaning.

Base Concepts⁹ were introduced within the WordNet research framework (Vossen et al., 1998) as the building blocks for constructing wordnets for different languages. Base Concepts typically satisfy two main criteria: a high position in the semantic hierarchy and having many relations to other concepts. WordNet synsets lexicalising Base Concepts are therefore among the likeliest candidates for general lexis.

The Bulgarian National Corpus (Koeva et al., 2012) consists of 5.4 billion words (1.2 billion for Bulgarian) and represents the lexis of contemporary Bulgarian. For our purposes, we extracted verbs with high frequency (over 10 per million words) that are found across different domains, text types and genres (at least two different domains, one of them being either fiction or news articles).

We also cross-checked the identified verbs against a small corpus of primary school textbook texts (for children aged 7 to 11 years old) in 5 different subjects. Verbs of high frequency appearing in textbooks in at least two subjects are deemed to belong to the general lexis. For verbs occurring in more than one synset we have manually selected the more general and frequent senses (based on human expert evaluation).

Concepticon is an open-source online lexical database of linguistic concepts which links concept labels from 160 concept lists (compiled from various sources and for various purposes) to 2495 concept sets (structured by defining different relations between the concepts) (List et al., 2016)¹⁰. In essence, it is a concept meta-resource which is applicable across various languages and is also linked to lexical-semantic resources such as WordNet and BabelNet.

Kuperman et al. (2012) present a data resource of 30,000 English words labelled with age of acquisition (AoA) information. The initial list was compiled by selecting base words (lemmas) appearing with high frequency in an English corpus of movie and TV series subtitles and the AoA data was collected using web-based crowdsourcing. The data includes the mean AoA ratings (in years of age) and standard deviations (attesting to the reliability of judgement), as well as the number of respondents who gave ratings to the word. Addi-

tionally, we have lemmatised the AoA word list, extracted the verbs rated with AoA of up to primary school age (up to the age of 11) and matched them to WordNet synsets.

Finally, we have identified FrameNet frames and the corresponding VerbNet classes that have a high coverage in terms of WordNet synsets (synsets assigned the respective frames and/or classes) as established in the extended inter-resource mappings used in this study. The assumption is that the most populated frames and classes represent the general part of the lexicon.

Using the criteria above, we have identified a dataset of 4,927 verb synsets of which: (a) 2,362 belong to the category of base concepts; (b) 1,800 have a high frequency in the Bulgarian National Corpus (frequency of 200+ counted as accumulative frequency of all literals in the synsets across all of their possible senses); (c) 1,470 synsets whose literals appear in primary school textbooks (frequency of 20+ in the textbook collection counting all occurrences of the synset literals across all of their possible senses); (d) 322 are included in Concepticon; (e) 252 have age of acquisition in primary school years; (f) 1,844 verb synsets have been assigned a high frequency frame. 1,405 synsets (28.5% of the dataset) are confirmed by at least 3 of the features, 212 (4.3%) are confirmed by 4 or more, which shows that the features are complementary for the purpose of general vocabulary extraction.

3.2 Conceptual Description of General Lexis Verbs

Currently, we focus on a set of frames and their corresponding verb classes to build a uniform model for conceptual description that can be expanded both in size and in terms of description features later on.

The efforts to align different lexical-semantic resources aim at combining various information into an extensive complex representation of the lexical units (in our case verbs) and the description of the main participants in the corresponding conceptual frame.

We consider the WN synsets with their assigned FN frame and VN class. Each synset is characterised by a pair of a frame and a verb class. As shown in Example 1 (a-c), for different synsets a frame can be corresponding to a number of verb classes, e.g. the frame *Body_movement* can corre-

⁹<http://globalwordnet.org/resources/gwa-base-concepts/>

¹⁰<https://concepticon.clld.org/>

spond to the verb class *crane-40.3.2* (with explicit body part participating in the movement), the verb class *curtsey-40.3.3* where the body part is incorporated in the verb’s meaning, or the verb class *modes_of_being_with_motion-47.3* where the movement concerns the whole body. Although in general verb classes are more concrete than frames, there are also cases where a number of frames are linked to a single verb class, as in Example 1 (c-d), hence the frame-to-verb class correspondence is ‘many-to-many’.

Example 1. Alignment between FrameNet frames and VerbNet classes.

(a) **WordNet synset:** *eng-30-00145902-v purse* ‘contract one’s lips into a rounded shape’

FrameNet frame: *Body_movement*: Agent (Sentient); Body_part (Body_part)

VerbNet class: *crane-40.3.2*: Agent [+animate]; Patient [+body_part]; Topic; Recipient [+animate]

(b) **WordNet synset:** *eng-30-02040549-v curtsey*; *curtsey* ‘bend the knees in a gesture of respectful greeting’

FrameNet frame: *Body_movement*: Agent (Sentient); Body_part (Body_part)

VerbNet class: *curtsey-40.3.3*: Agent [+animate]; Topic; Recipient [+animate]

(c) **WordNet synset:** *eng-30-01865383-v bob* ‘move up and down repeatedly’

FrameNet frame: *Body_movement*: Agent (Sentient); Body_part (Body_part)

VerbNet class: *modes_of_being_with_motion-47.3*: Agent [+int_control]; Theme [+concrete]; Location [+location & -region]

(d) **WordNet synset:** *eng-30-01868258-v waver*, *wave* ‘sway to and fro’

FrameNet frame: *Self_motion*: Self_mover (Sentient); Area (Location) — Source (Source); Path (Path); Goal (Goal); Direction

VerbNet class: *modes_of_being_with_motion-47.3*: Agent [+int_control]; Theme [+concrete]; Location [+location & -region]

We have identified 96 pairs of FN frame and VN verb class assigned to 2,016 verb synsets. Out of the pairs only 12 have identically named frame and verb class, which suggests close correspondence (e.g., *Escaping* – *escape-51.1*, *Filling* – *fill-9.7*, *Destroying* – *destroy-44*, etc.). There are 20 frames mapped to more than one verb class, out of which 11 frames are mapped to 3 or more verb classes

each.

3.3 Alignment between Semantic Roles and Frame Elements

The challenges to the mapping of frame elements and semantic roles stem from several sources: (i) differences in the conceptualisation of the situations between frames and verb classes; (ii) differences in the status of the frame elements and semantic roles (not all core elements necessarily have a semantic role counterpart and vice versa); (iii) differences in the syntactic description across the resources, etc.

Example 2. FrameNet frame *Escaping* aligned to VerbNet class *escape-51.1*.

WordNet synset: *eng-30-02074677-v escape*; *get away*; *break loose*

FrameNet frame: *Escaping*

Core FN FEs: *Escapee* (Semantic Type: Animate_being); *Undesirable_location* (Semantic Type: Source)

VerbNet class: *escape-51.1*

VN roles: *Theme*; *Initial_location*; *Destination*; *Trajectory*

FN element and status	Semantic type	VN role	VN restriction
Escapee	Animate_being	Theme	[concrete +]
Goal	Goal	Destination	[concrete +]
Means	State_of_affairs		
Manner	Manner		
Undesirable_location	Source	Initial_location	[concrete +]
Speed	Speed		
Vehicle			
Time	Time		
Purpose	State_of_affairs		
Place	Locative_relation		
Depictive			
Path		Trajectory	[concrete +]
Degree	Degree		
Distance			
Explanation			

Consider Example 2, which represents the mapping of the FrameNet frame *Escaping* to the VerbNet class *escape-51.1*. Judging from their names, one expects the alignment to be very straightforward. However, *Escaping* has two core frame elements – *Escapee* and *Undesirable situation*, while *escape-51.1* is associated with four semantic roles: *Theme*, *Destination*, *Initial location* and *Trajectory*. Table 1 shows the mapping of the frame elements and the semantic roles: *Escapee* maps to *Theme*

and *Undesirable location* maps to *Initial Location*. In addition, the semantic role *Destination* corresponds to the peripheral frame element *Goal*, and *Path* aligns with *Trajectory*.

The table of Example 2 shows the alignment between the frame elements of the frame *Escaping* to the roles of the VerbNet class *escape-51.1*.

The judgment of which frame element corresponds to which semantic role is made by employing semantic information from the two resources, including comparison of definitions and similarity in the naming of the elements and roles (where possible) and inferred knowledge abstracted away from the structure of FrameNet where the frame elements are too specific. The latter case 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 the Inheritance relation between frames (Leseva et al., 2018a). For instance, the fact that *Text_creation* inherits its properties from several frames (forming a chain of inheritance from a more specific to a more general frame) – *Text_creation* > *Intentionally_create* > *Creating* > *Transitive_action* – allows us to identify a corresponding inheritance relation between relevant frame elements involved in these frames: *Author* > *Creator* > *Creator* > *Agent* and *Text* > *Created_entity* > *Created_entity* > *Patient*, that is the frame elements expressed as the subject and the direct object position in the frames under discussion. Having obtained this correspondence to more general frame elements, we try to map them to relevant roles in the semantic role set of the VerbNet verb class aligned with the respective frame. Thus, *Author* will be mapped to *Agent* in VerbNet.

Similarly, in Example 2 above *Escaping* inherits its properties from *Departing*, enabling us to infer the frame element Inheritance relations *Escapee* > *Theme* and *Undesirable_location* > *Source*. The comparison between the most abstract frame elements and the semantic roles in the respective VerbNet class, leaves us with the straightforward alignment: *Theme* – *Theme* and the more or less transparent one: *Source* – *Initial_location*.

3.4 Alignment between Syntactic Patterns

After aligning FN frames to VN verb classes (assigned to synsets or groups of synsets), and FN FEs to VN roles, we move towards mapping syntactic patterns from the resources aiming at providing a

new, syntactic layer to the conceptual description of general lexis verbs. The criteria for equivalence between two syntactic patterns obtained from the two resources 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).

Example 3. Aligned syntactic patterns for the FrameNet frame *Escaping* and the VerbNet class *escape-51.1*.

VN	NP(Theme)	V	
FN	NP.Ext(Escapee)	V	
VN	NP(Theme)	V	NP(Destination)
FN	NONE		
VN	NONE		
FN	NP.Ext(Escapee)	V	DNI.(Undesirable_location)
VN	NP(Theme)	V	NP(Initial_Location)
FN	NP.Ext(Escapee)	V	NP.Obj(Undesirable_location)
VN	NP(Theme)	V	NP(Trajectory)
FN	NONE		
VN	NP(Theme)	V	PP.destination(Destination)
FN	NP.Ext(Escapee)	V	PP[into].Dep(Goal)
VN	NP(Theme)	V	PP.initial_location (Initial_Location)
FN	NP.Ext(Escapee)	V	PP[from].Dep (Undesirable_location)
VN	NP(Theme)	V	PP.initial_location (Initial_Location)
FN	NP.Ext(Escapee)	V	PP[from].Dep (Undesirable_location)
VN	NP(Theme)	V	PP.initial_location (Initial_Location)
FN	NONE		PP.destination(Destination)
VN	NP(Theme)	V	PP.trajectory(Trajectory)
FN	NP.Ext(Escapee)	V	PP[from].Dep(Path)

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 (Examples 3 and 4). Where no correspondence is discovered, the table cell is marked as NONE.

Example 3 shows the alignment of the syntactic patterns between the frame *Escaping* and the class *escape-51.1* following the mapping between the FEs and VN semantic roles (Theme – Escapee, Initial_location – Undesirable_location, Destination – Goal and Path – Trajectory). Misalignment occurs in the cases of additional semantic roles that are not considered core FEs (e.g., Trajectory).

Example 4. Aligned syntactic patterns for the FrameNet frame *Killing* and the VerbNet class *murder-42.1* (e.g., *kill, slay, annihilate, assassinate*, etc.).

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

Example 5. Aligned syntactic patterns for the FrameNet frame *Killing* and the VerbNet class *suffocate-40.7* (e.g., *asphyxiate, choke, suffocate*, etc.).

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

Examples 4 and 5 show different degree 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 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 under 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 5 (e.g., *suffocate to/into unconsciousness*).

The asymmetries in the syntactic patterns covered by matched FN frames and VN verb classes for particular WN synsets are indicative of the need for more detailed syntactic analysis and the study of both the alignment between the FEs and the semantic roles and their syntactic realisation.

4 Discussion of Results

The task of aligning FrameNet and VerbNet poses a number of challenges.

(1) Aligning frame elements and semantic roles at a different level of granularity.

This task is approached by employing (i) the semantic alignment of the fine-grained FrameNet frame elements to the more generalised VerbNet semantic roles using straightforward correspondences and the frame element hierarchy discussed in 3.3; (ii) the syntactic mapping – correspondences in terms of syntactic categories, prepositions, subordinating conjunctions, types of clauses, etc. – between frame elements and VerbNet roles with similar semantics and place in the conceptual description of particular verbs.

(2) Aligning the syntactic patterns for frames and verb classes with a different number of components or ones that allow alternative syntactic realisation. For instance, the syntactic description of the frame *Statement* includes the pattern:

NP.Ext(Medium) V Sfin

The sign announced that the bar was closed.

while no syntactic patterns with a finite clause are found in the description of the corresponding VerbNet class *talk-37.5*.

The semantic and syntactic information coming from different resources can serve for the validation of the linguistic generalisations captured in each of them. Thus, discrepancies across resources may be a sign of missing information in one of them and can be used for the enhancement of the poorer description.

Alternatively, the lack of correspondence may also be a red flag of the lack of semantic correspondence between seemingly identical or similar senses and hence should be studied with caution.

(3) Taking care of alterations such as passives (which are defined in FrameNet as separate syntactic patterns but are not represented in VerbNet).

Our approach would be to use the more comprehensive and explicit description in order to validate

the various alternations both within a language and cross-linguistically.

(4) Adapting syntactic patterns across languages and capturing significant parallels and differences in the syntactic projection cross-linguistically.

Both semantic and syntactic patterns may be adopted and possibly adapted cross-linguistically. Using an already available predefined set of patterns and refining or modifying them where needed, allows for a uniform representation of the data across languages and may be used to obtain a more complete description in the cases where corpora are not large enough to yield examples for all possible syntactic frames. Even so expert validation is indispensable.

With respect to the two research questions we have obtained the following results.

How can we integrate semantic and syntactic information to enhance the conceptual description of WordNet synsets and literals?

Such an integration may be implemented by employing semantic correspondences and syntactic patterns which apply to all (or most of) the synonyms in a given synset. In addition, more specific syntactic frames are needed in many cases to fine-tune these patterns and to cater for the syntactic realisation of individual literals, e.g. specific prepositions introducing prepositional phrases for different synonyms.

For example, the Bulgarian correspondence of the synset eng-30-00811375-v *avoid* includes, among others, the verbs *izbyagvam* and *stranya*. The former is associated with patterns corresponding to the ones defined in FrameNet and VerbNet for English:

NP.Ext(Agent) V NP.Obj(Undesirable_situation)

EN: *Her friends now avoided her.*

BG: *Priyatelite i sega ya izbyagvaha.*

The latter, *stranya*, however, requires its *Undesirable_situation* element to be realised as a PP headed by the preposition *ot* (from), which is not the case in English:

NP.Ext(Agent) V PP[from](Undesirable_situation).

EN: *Her friends now avoided her.*

BG: *Priyatelite i sega stranyaha ot neya.*

This necessitates the definition of language-specific syntactic frames on the basis of evidence from the language under study.

To what extent is syntactic information lan-

guage independent and can it be transferred from English to less-resourced languages such as Bulgarian?

Although by no means identical, semantic descriptions are largely applicable across languages as far as senses are defined in a similar manner and should be largely uniform within a given synset. Syntactic frames are much more divergent cross-linguistically, yet there are major trends and similarities that may be transferred with caution across languages and resources.

With respect to general lexis verbs, our expectations are that they are realised by means of more common and well-established syntactic patterns with less specific features. Many of them are similar between languages.

However, an extensive analysis of syntactic structures should be carried out in order to determine the degree to which syntactic patterns defined for English can be adapted automatically to serve Bulgarian. To this end there are various corpora that can be used to extract occurrences of certain verbs, to study their context, combinations with prepositions, etc.

5 Conclusions and Future Work

The research presented in this paper aims at providing a reliable alignment between: (a) FrameNet frame elements and VerbNet semantic roles on the basis of mapped FN frame – VN verb class pairs assigned to a number of WordNet synsets; (b) FrameNet lexical units' syntactic patterns and VerbNet syntactic frames. These combined allows for expanding the conceptual description of verbs with information about their syntactic realisation. Further, the data offer extensive opportunities to investigate to what extent the conceptual and the syntactic information can be transferred between languages, especially languages from one language family. These observations can play a crucial role in expanding semantic and syntactic description of Bulgarian verbs and thus, boost the development of new NLP applications.

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).

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, 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.
- Aljoscha Burchardt, Katrin Erk, and Anette Frank. 2005. A WordNet detour to FrameNet. In *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*, volume 8 of *Computer Studies in Language and Speech*. Lang, Frankfurt, Germany.
- 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, editor. 1998. *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.
- Svetla Koeva, Ivelina Stoyanova, Svetlozara Leseva, Rositsa Dekova, Tsvetana Dimitrova, and Ekaterina Tarpomanova. 2012. *The Bulgarian National Corpus: theory and practice in corpus design*. *Journal of Language Modelling*, 0(1):65–110.
- Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44:978–990.
- 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.
- Johann-Mattis List, Michael Cysouw, and Robert Forkel. 2016. Concepticon: A Resource for the Linking of Concept Lists. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pages 2393–2400. European Language Resources Association (ELRA).
- George A. Miller. 1995. WordNet: A lexical database for English. *Commun. ACM*, 38(11):39–41.
- 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.
- 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. Springer, Berlin, Heidelberg.
- 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.

Piek J. T. M. Vossen, Laura Bloksma, and Rodriquez Horacio. 1998. The EuroWordNet base concepts and top ontology. Workingpaper, Vrije Universiteit.