A 'Dip-dive' into Motion: Exploring Lexical Resources towards a Comprehensive Semantic and Syntactic Description

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

Abstract

In this paper I illustrate the semantic description of verbs provided in three semantic resources (FrameNet, VerbNet and VerbAtlas) in comparative terms with a view to identifying common and distinct components in their representation and obtaining a preliminary idea of the resources' interoperability. To this end, I provide a comparison of a small sample of motion verbs aligned with semantic frames and classes in the three resources. I also describe the semantic annotation of Bulgarian motion verbs using the framework defined in the Berkeley FrameNet project and its enrichment with information from the other two resources, which has been enabled by the mapping between: (i) their major semantic units – FrameNet frames, VerbNet classes and VerbAtlas frames, and (ii) their 'building blocks' – frame elements (FrameNet)and semantic roles (VerbNet, VerbAtlas).

Keywords: semantic annotation, semantic frames, verb classes, FrameNet, VerbNet, VerbAtlas, WordNet

1 Introduction

In this paper I embark on a comparison of the semantic description of verbs (using WordNet as a lexical inventory) provided in three semantic resources, FrameNet, VerbNet and VerbAtlas, to the end of identifying common and distinct components in their representation and obtaining a preliminary idea of the resources' interoperability. The study is based on a small sample of motion verbs aligned with semantic frames and classes in the three resources. The analysis is applicable both to theoretical studies of the semantic components of meaning and to the practical task of validating the automatic assignment of semantic information from the resources employed to WordNet synsets. In the second part of the paper I describe the semantic annotation of Bulgarian motion verbs using the framework defined in the Berkeley FrameNet project and its enrichment with information from the other two resources, which has been enabled by the mapping between: (i) their major semantic units – FrameNet frames, VerbNet classes and VerbAtlas frames, and (ii) their 'building blocks' – frame elements (FrameNet) and semantic roles (VerbNet, VerbAtlas). The two tasks have been implemented simultaneously, focusing on motion verbs. Part of the work has also been employed in the development of a FrameNet-like semantic resource for Bulgarian.

2 Language resources

Several large-coverage lexical-semantic and syntactic resources as well as a couple of annotated corpora have been employed in this study. In this Section I briefly present them and comment on how they are integrated with each other.

2.1 Lexical semantic resources

2.1.1 WordNet

The work described below makes use of the Princeton WordNet and its counterpart for Bulgarian, the Bulgarian WordNet.

The Princeton WordNet[1], PWN, (Miller, 1995; Fellbaum, 1998) is a large lexical database that represents comprehensively conceptual and lexical knowledge as a richly populated network whose nodes denote cognitive synonyms (synsets) linked to each other by means of a number of conceptual-semantic and lexical relations such as hypernymy, meronymy, antonymy, etc.

The Bulgarian Wordnet (Koeva, 2021)[2] has been developed by translation and adaptation of the PWN synsets and further enriched with language-specific synsets and features. The counterparts in

[1] https://wordnet.princeton.edu/

[2] The Bulgarian WordNet may be browsed at: https://dcl.bas.bg/bulnet/

the two wordnets are related to each other through unique interlingual identifiers, which also provide links to the respective synsets in all other wordnets that support them.

The two wordnets provide the verb inventories (for each language) that have been enriched with semantic and syntactic information through the mapping with the other lexical resources employed.

2.1.2 FrameNet

FrameNet (Baker et al., 1998; Baker, 2008) is a lexical resource which couches lexical and conceptual knowledge in a framework originating in the theory of Frame Semantics (Fillmore, 2003; Ruppenhofer et al., 2016). Frames are conceptual structures describing particular types of objects, situations, or events along with their components (frame elements, FEs) (Baker et al., 1998; Ruppenhofer et al., 2016). Each frame in FrameNet is represented by means of a definition that describes schematically the situation and the way in which at least the most essential FEs are involved in it. Each FE is also supplied with a definition that further clarifies its semantics and its interaction with other FEs.

Frame elements have different status reflecting their role in the description of a given frame, i.e. core, peripheral or extra-thematic (Ruppenhofer et al., 2016: 19–20). Core FEs instantiate conceptually obligatory components, while making the frame unique and different from other frames. Peripheral FEs make reference to notions such as Time, Place, Manner, Means, Degree, etc. Extra-thematic FEs characterise an event against a backdrop of another state of affairs. Frames are expressed, or 'evoked', by a set of lexical units (LUs), which are pairings of a word and a meaning whose conceptual semantics is represented by the frame.

FrameNet frames form a network by means of a number of frame-to-frame relations (cf. Ruppenhofer et al. (2016: 81–84) for a detailed account). Some of them, the strictest one being Inheritance, define relations involving the handing down of properties among hierarchically organised frames. In particular, Inheritance is defined as a relation between a more general (parent) frame and a more specific (child) frame where there should be a rather strict correspondence between the semantic characteristics of the parent and the child frame (Petruck, 2015). An illustration of this relation is represented by the sister frames `Self_motion`, `Fluidic_motion`, among others, which share the main configuration of frame elements defined for their parent frame `Motion`, but in addition provide a further specification of the THEME as an entity moving under its own power and will, i.e. a SELF_MOVER (in `Self_motion`), or as a FLUID (in `Fluidic_motion`).

Apart from their linguistic significance, Inheritance and other, primarily hierarchical relations play a role in some of the proposed FrameNet-to-WordNet mapping procedures.

2.1.3 VerbNet

VerbNet (Kipper-Schuler, 2005; Kipper et al., 2008) is a hierarchical network of English verbs which represents their syntactic and semantic patterns[3]. 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 cross-linguistically.

2.1.4 VerbAtlas

VerbAtlas (Di Fabio et al., 2019) is a lexical-semantic resource representing the semantic description of the verb synsets in BabelNet. BabelNet is a very large, richly populated multilingual semantic network (covering more than 500 languages) which integrates lexicographic and encyclopaedic knowledge from WordNet and Wikipedia (Navigli and Ponzetto, 2010)) as well as from other sources. Each verb synset in VerbAtlas is assigned a frame corresponding to its prototypical predicate-argument structure. Obligatory components are described using semantic roles and the semantic properties governing their compatibility.

The alignment of these resources harnesses WordNet's rich lexical knowledge (glosses, examples, semantic primitives and semantic relations) and the detailed semantic and syntactic description implemented in FrameNet, VerbNet and VerbAtlas. Section 3 below describes the logic underlying the alignment efforts.

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

3 Inter-resource mapping

This work employs a WordNet-to-FrameNet mapping described in Leseva and Stoyanova (2020) which makes use of and builds upon several previously implemented ones, in particular: the direct mappings provided within FrameNet (Baker and Fellbaum, 2009), eXtendedWordFrameNet (Laparra and Rigau, 2010) and MapNet (Tonelli and Pighin, 2009), supplemented with additional indirect mapping through VerbNet (Palmer, 2009; Palmer et al., 2014). The combination of these previous mappings resulted in 4,306 unique synset-to-frame alignments, amounting to 30.5% of all the verb synsets (Leseva and Stoyanova, 2020).

Different methods for expanding the coverage between the two resources have been proposed. Two of the well-known approaches involve discovering a possibly suitable frame for a verb in WordNet not featured in FrameNet based on its semantic relation(s) to verbs having a correspondence in FrameNet (Burchardt et al., 2005) or applying graph-based algorithms to identify relevant senses of verbs evoking certain semantic frames (de Lacalle et al., 2016).

The procedure used in this paper is based on the exploration of the structural properties of WordNet and FrameNet (Leseva and Stoyanova, 2020). Verbs in a WordNet synset generally exhibit the same or very similar meaning, which implies that they are associated with the same semantic frame. Moreover, both resources are hierarchically organised based on the notion of inheritance from a more general to a more specific synset or frame. The alignment between the resources reflects the notion of inheritance – a more specific concept should be associated with the frame of its hypernym or with a more specific frame elaborating on (and possibly inheriting from) or otherwise related to the mother frame, although this is not borne out consistently in practice. The method involves: (1) manual assignment of frames to root verb synsets to ensure accurate mapping at the top hierarchical level (and thus alleviate error propagation); (2) automatic assignment of the hypernym's frame onto hyponyms which were not previously mapped; and (3) implementation of a set of enhancing procedures involving post-editing (Leseva and Stoyanova, 2020).

As a result, the coverage has been increased to 13,104 synset-to-frame alignments (94% of the verb synsets), of which more than 6,000 have been validated manually.

Finally, VerbAtlas is aligned with WordNet by design, as its verb frames have been defined specifically to provide appropriate semantic description in terms of predicate-argument structures and ensure complete coverage of the WordNet synsets.

3.1 Dataset selection

Two Bulgarian corpora have been employed as a source of examples for semantic role annotation: a semantically annotated corpus, BulSemCor (Koeva et al., 2006, 2011) designed according to the general methodology of the original SemCor (Miller et al., 1993; Landes et al., 1998) coupled with criteria for ensuring appropriate coverage of contemporary general lexis. The size of the corpus is close to 100,000 annotated units. As BulSemCor is manually annotated with WordNet senses, it provides disambiguated examples for the studied verbs.

As discussed above, most verb synsets have been assigned a FrameNet frame via one or another of the mapping efforts described in Section 3. While the assignment itself needs manual validation (which has also been underway), once it is verified, a SemCor-like corpus turns into a corpus available for semantic role annotation. This is exactly the approach adopted in this paper: the BulSemCor sentences containing verb synsets from the studied domain whose assigned FrameNet frames have been validated are extracted for annotation.

Where the number of examples is found to be insufficient, the dataset has been supplemented with sentences from the Bulgarian National Corpus, a corpus of 1.2 billion words of running Bulgarian text distributed in 240,000 text samples spanning the second half of the 20th century and the beginning of the 21st century (Koeva et al., 2012). As the corpus is not sense-disambiguated, the examples excerpted from it have been selected manually so as to correspond to the studied senses.

The resulting collection of example sentences has been annotated according to the FrameNet guidelines. The phrases realising syntactically the core frame elements have been explicitly marked in a similar fashion to the annotation in the Berkeley FrameNet corpus. The selection covers 893 annotated clauses or sentences distributed as follows across five frames: `Motion` – 149; `Self_motion` – 262; `Arriving` – 182; `Departing` – 178; `Traversing` – 122.

4 Preliminaries: Scratching the surface of the `Motion` domain

The examples in the following part of the paper feature motion verbs, in particular verbs evoking the FrameNet frame `Self_motion`, which is a direct descendant of the 'prototypical' frame for this semantic domain: `Motion`. This latter frame, as well as the ones that inherit its properties involve the translational motion of different entities. Its semantic description is presented below in more detail, as it serves as a point of departure in defining the more specific frames (e.g. `Self_motion`), while also showing the resemblances with the way motion is described in VerbNet and VerbAtlas.

FrameNet definition: `Motion`

Some entity, the THEME starts out in one place SOURCE and ends up in some other place, the GOAL, having covered some space between the two (PATH). Alternatively, the AREA or DIRECTION in which the THEME moves or the DISTANCE covered may be mentioned.

Core frame element set: THEME, SOURCE, GOAL, PATH, AREA, DIRECTION, DISTANCE.

The THEME [4] is the entity that changes location either by moving on its own and/or under its own power or by being moved, carried, etc. by another entity or force. Its semantic specification includes **animate beings** and **physical objects**.

The SOURCE is a location or an entity occupying space that serves as the starting point or landmark where the moving entity is at before it starts changing location. Its semantic specification spans various **locations**, including but not limited to geological and other natural formations, geographical points, celestial bodies or **physical objects**, including man-made structures, such as buildings, constructions, facilities, etc.

The GOAL is the location or another entity serving as the point at which the moving entity ends up as a result of the motion. It has the same semantic specification as the SOURCE.

The PATH is any trajectory of motion confined between the SOURCE and the GOAL, (a part of) the ground over which the moving entity travels or a landmark by which it travels. Its specification encompasses **locations** or **physical objects** that may be construed as having extent along which the motion takes place, various **media** (water, air), etc.

The AREA identifies the setting in which the motion does not occur along a single linear trajectory. In consequence, it cannot co-occur with SOURCE, GOAL and PATH. The semantic specification of the AREA refers to **locations** or **physical objects** that may be construed as comprising some spatial expanse within which the motion takes place, various **media**, etc.

The DIRECTION indicates the general spatial orientation of the motion along a line (the PATH) from the deictic centre towards a (possibly implicit) reference point that is neither the GOAL of the posture change nor a landmark along the way of the moving part of the body. Its specification includes compass points (north, east, south, west), body relative directions (left, right, back, front, backward, forward, up, down), coordinates, etc.

The DISTANCE characterises the spatial extent of the motion, the distance travelled by the moving entity. Its specification includes **distance denotations** expressed either in a give system of measurement or in relative terms (farther, closer), etc.

`Self_motion` is an elaboration of the `Motion` frame and related to it by means of an Inheritance relation. It involves a similar configuration of core FEs with some further restrictions.

FrameNet definition: `Self_motion`

The SELF_MOVER, a living being, moves under its own direction along a PATH. Alternatively or in addition to PATH, an AREA, DIRECTION, SOURCE, or GOAL for the movement may be mentioned.

The most important distinction between this frame and `Motion` is the capability of the SELF_MOVER to change location by exercising their own will and power by the coordinated movement of their bodies, which is not a necessity with the `Motion` THEME. By metaphorical extension, SELF_MOVERS may be self-directed entities such as vehicles. Its semantic specification thus includes **animate beings** and **vehicles**.

The remaining core FEs in this frame are the ones defining the elements and aspects of the route of movement as defined for the parent frame `Motion` and have the same semantic specification as their counterparts there[5].

[4] For better generalisation, the definition of each frame element takes into account its description provided in the prototypical `Motion` frame and in other frames related to it.

[5] DISTANCE is not defined as a core FE, but I will treat it on a par with its equivalent in the `Motion` frame.

5 Motivation and prerequisites: What do we gain from resource alignment?

One of the goals of this paper is to illustrate the logic and motivation underlying the comparison undertaken towards the identification of common and distinct components of the semantic description of motion verbs in FrameNet, VerbNet and VerbAtlas. The purpose is to obtain a preliminary idea of the interoperability of the resources and the optimal ways to harness their strengths towards their further enrichment. To this end I compare a small sample of motion verbs and their semantic descriptions across resources. The inventory of verbs employed in the study are WordNet synsets to which FrameNet frames, VerbAtlas frames and possibly VerbNet classes have been assigned as described in the Sections 2.1 and 3.

The purpose of the study is twofold: (i) to illustrate the procedure of validating the automatic assignments of FrameNet frames to WordNet synsets and to suggest how the comparison of the semantic descriptions provided in different resources may give insights into the data; (ii) to describe the semantic annotation undertaken for Bulgarian using the validated FrameNet frame assignments and information from the other two semantic resources (VerbNet and VerbAtlas). The two tasks are implemented simultaneously, focusing on a particular class of verbs, in this case – motion predicates.

Let us illustrate the actual result of the alignment among the resources and what is gained by it. In most simple terms, the mapping between a WordNet synset and a FrameNet frame, on the one hand, or a WordNet synset and a VerbNet class, on the other, is based primarily on an existing correspondence between (i) one or more WordNet synset members (literals) and one or more FrameNet lexical units (LUs); or (ii) between one or more WordNet synset members and one or more verbs belonging to a VerbNet class. This lexical correspondence is the backbone of the inter-resource alignment, which is further expanded by more refined procedures aimed at obtaining greater and more perfect coverage. In the case of the mapping between WordNet and FrameNet, these include calculation of similarity between the definitions in different resources or between the definitions and other fragments of knowledge such as a frame's name (consider the correspondence between the gloss of the synset {change:2} 'undergo a change...' and the frame's name Undergo_change), the utilisation of the resources' structure, especially semantic inheritance, causative/inchoative relations, etc.

Figure 1 shows a straightforward mapping of the synset {walk:1} 'use one's feet to advance; advance by steps' to the FrameNet LU *walk.v*, which evokes the frame `Self_motion`. Each LU is supplied with a dictionary definition, in this case one borrowed from the Concise Oxford Dictionary, and the comparison to the synset gloss confirms the correspondence between the units in the two resources. The accuracy of the alignment with the VerbNet class run-51.3.2 is borne out by the correspondences between pairs of VerbNet semantic roles and FrameNet frame elements[6], and/or by VerbNet lemmas sharing class membership with *walk*, which have as their WordNet counterparts either hyponyms – e.g. {limp:1}, {shuffle:1}, {amble:1}, or sisters – e.g. {swim:1}, {run:34} – of {walk:1} or are otherwise close to this synset in the WordNet structure (and hence, have strong similarity).

Unlike FrameNet and VerbNet where the alignment with WordNet is a secondary feature implemented on the basis of the already developed resources – semantic frames and lexical units (FrameNet) and semantic classes and lists of verb lemmas (VerbNet) – VerbAtlas has been developed specifically to describe the verb synsets in BabelNet (of which WordNet is a part), both in terms of their lexical semantics and predicate argument structure. The 13,767 verb synsets in WordNet are covered by 466 frames. WordNet's {walk:1} is assigned the VerbAtlas frame 0137f GO-FORWARD. The definition provided for this frame describes the relationship among a set of semantic roles, part of which may be straightforwardly aligned to frame elements in the frame `Self_motion` and to semantic roles in the VerbNet class run-51.3.2[7].

Given the very different premises of how the frames or classes are assigned, there is not and cannot be a one-to-one correspondence among the resources. In a test set of 206 manually validated synsets mapped to the FrameNet frame `Self_motion`, 88 are aligned with the VerbNet class run-51.3.2 and 135 are assigned one of 4 VerbAtlas frames as shown in Table 1, while the

[6]The colour coding on Figure 1 shows the matching frame elements and semantic roles that have been associated to each other empirically.

[7]Part of the correspondences, e.g. the one between FrameNet's PATH / VerbNet's TRAJECTORY and the VerbAtlas LOCATION are not straightforward and have been established by annotating example sentences on the VerbAtlas web platform: https://verbatlas.org/.

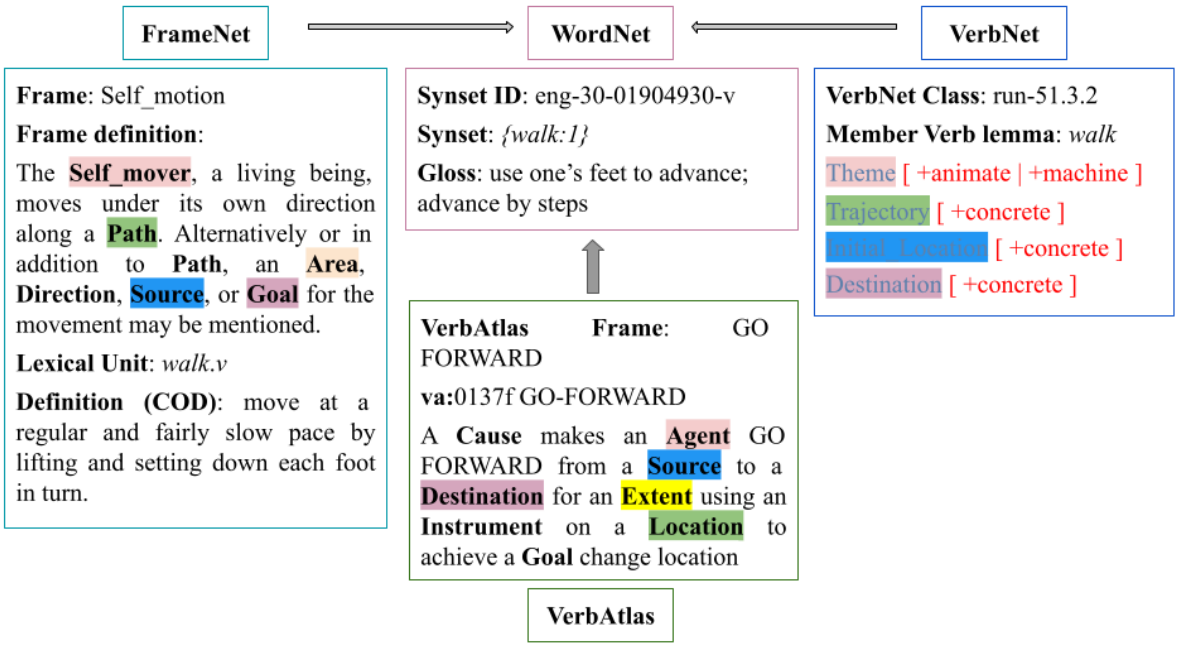

Figure 1: An illustration of the alignment between a WordNet synset and its counterparts in FrameNet, VerbNet and VerbAtlas, with the empirically confirmed frame element-to-semantic role correspondences.

remaining synsets are assigned one of 20 other VerbAtlas frames in fewer numbers[8].

The examples below illustrate the semantic and syntactic annotation across the different resources. For each verb in the examples I have selected a sentence from the manually annotated FrameNet corpus and have run it (possibly simplified) through the VerbAtlas SRL platform[9] and the VerbNet SLR parser[10]. The resulting automatic annotation, with corrections where deemed necessary, is given below along with the original FrameNet sentence.

Example 1: AMBLE

FrameNet frame: **`Self_motion`**

[*He*]$_{\text{SELF\_MOVER}}$ *AMBLED* [*down the long winding corridor*]$_{\text{PATH}}$.

VerbAtlas frame: **MOVE-ONESELF**

[*He*]$_{\text{THEME}}$ *AMBLED* [*down the long winding corridor*]$_{\text{DESTINATION}}$.

VerbNet class: **run-51.3.2**

[*He*]$_{\text{THEME}}$ *AMBLED* [*down the long winding corridor*]$_{\text{TRAJECTORY}}$.

Example 2: LUMBER

FrameNet frame: **`Self_motion`**

[*He*]$_{\text{SELF\_MOVER}}$ *LUMBERED* [*down the steep narrow staircase*]$_{\text{PATH}}$.

VerbAtlas frame: **TRAVEL**

[*He*]$_{\text{THEME}}$ *LUMBERED* [*down the steep narrow staircase*]$_{\text{DESTINATION}}$.

VerbNet class: **run-51.3.2**

[*He*]$_{\text{THEME}}$ *LUMBERED* [*down the steep narrow staircase*]$_{\text{TRAJECTORY}}$.

Testing such sets of sentences gives a glimpse into the similarities and differences in the annotation and how these differences may be remedied.

Below I also examine some differences in the definition of FrameNet frames and VerbAtlas frames. As shown in Table 1, {lumber:1} is assigned the frame TRAVEL in the latter resource, while FrameNet and VerbNet assign it to the same frame/class as {walk:1}. Looking at the definitions of the two VerbAtlas frames, one notes several differences. One major distinction is that TRAVEL's roleset does not define a SOURCE role, but it does define a GOAL, and possibly a PATH (the role LOCATION may be construed in this way based on the automatic annotation)[11]. The FrameNet `Travel` frame specifies a similar set of elements of the route of motion as its parent frame `Self_motion`: SOURCE, PATH, GOAL, AREA, DIRECTION. Example 3 below presents an annotated FrameNet

[8] Not all synsets are assigned a VerbNet class.

[9] https://verbatlas.org/

[10] https://verbnetparser.com/

[11] For the sake of clarity I will refrain from analysing the remaining roles defined in the GO-FORWARD frame but not in the TRAVEL frame.

Synset	FrameNet	VerbNet	VerbAtlas	Synsets	VerbAtlas frames
{lumber:1}	Self_motion	run-51.3.2	TRAVEL	45	A **cause** makes an **agent** TRAVEL on a **location** to a **destination**
{walk:1}	Self_motion	run-51.3.2	GO-FORWARD	39	A cause makes an **agent** GO FORWARD from a **source** to a **destination** for an **extent** using an **instrument** on a location to achieve a goal Change location
{amble:1}	Self_motion	run-51.3.2	MOVE-ONESELF	29	An **agent** makes a **theme** MOVE ONESELF on a **location** for an **extent** from a **source** to a **destination** to a **patient** having a **result** (+ **attribute**)
{run:34}	Self_motion	run-51.3.2	RUN	22	An **agent** make a **theme** RUN with a **co-theme** on a **location** or from a **source** to a **destination** for an **extent** for a **purpose** using an **instrument**

Table 1: The most frequent correspondences between the FrameNet frame Self_motion, the VerbNet frame run-51.3.2 and the corresponding VerbAtlas frames.

sentence for the verb *journey.v*, which is assigned the FrameNet frame Travel, the VerbAtlas frame TRAVEL and the VerbNet class run-51.3.2. In this case, a SOURCE (*from Heathrow*) and a GOAL (*to Edinburgh*) are expressed, but both are annotated as DESTINATION by the VerbAtlas SRL system, possibly because the role SOURCE is not part of the description of the VerbAtlas frame. Likewise, *from Heathrow* is parsed by the VerbNet parser not as INITIAL_LOCATION, the role which has empirically been confirmed to correspond to the FrameNet SOURCE (as illustrated in Figure 1), but as DIRECTION. After inspecting the list of syntactic frames of run-51.3.2, one can speculate that this can be attributed to the fact that there is not an explicit syntactic frame associated with this VerbNet class or its subclasses that allows the INITIAL_LOCATION and the DESTINATION to be expressed simultaneously (there is, however a syntactic frame describing the simultaneous expression of the INITIAL_LOCATION, the TRAJECTORY and the DESTINATION). In addition, the role DIRECTION is assigned not from VerbNet but from the PropBank roleset (marked by an underline in the examples). The same goes for the MANNER role, roughly corresponding to the FrameNet frame element MODE_OF_TRANSPORTATION, which does not have a (clear-cut) counterpart in the corresponding VerbAtlas frame and VerbNet class.

Example 3: JOURNEY

FrameNet frame: **Travel**

[*He*]$_{\text{TRAVELER}}$ *JOURNEYED* [*from Heathrow*]$_{\text{SOURCE}}$ [*to Edinburgh*]$_{\text{GOAL}}$ [*by overnight coach*]$_{\text{MODE\_OF\_TRANSPORTATION}}$.

VerbAtlas frame: **TRAVEL**

[*He*]$_{\text{AGENT}}$ *JOURNEYED* [*from Heathrow*]$_{\text{DESTINATION}}$ [*to Edinburgh*]$_{\text{DESTINATION}}$ [*by overnight coach*]$_{\text{ATTRIBUTE}}$.

VerbNet class: **run-51.3.2-1**

[*He*]$_{\text{THEME}}$ *JOURNEYED* [*from Heathrow*]$_{\underline{\text{DIRECTION}}}$ [*to Edinburgh*]$_{\text{DESTINATION}}$ [*by overnight coach*]$_{\underline{\text{MANNER}}}$.

Similarly, in Example 4, the peripheral frame element COTHEME specified in the Self_motion frame does not have a counterpart in the respective rolesets in the other two resources. As a result, it is not labelled or, in the case of the VerbNet parser, is assigned the PropBank role COMITATIVE.

Example 4: WALK

FrameNet frame: **Self_motion**

I hope you won't mind if [*I*]$_{\text{SELF\_MOVER}}$ *WALK* [*home*]$_{\text{GOAL}}$ [*with you*]$_{\text{COTHEME}}$.

VerbAtlas frame: **GO FORWARD**

I hope you won't mind if [*I*]$_{\text{AGENT}}$ *WALK* [*home*]$_{\text{DESTINATION}}$ [*with you*].

VerbNet class: **run-51.3.2.2-1**

I hope you won't mind if [*I*]$_{\text{THEME}}$ [*WALK home*]$_{\text{DESTINATION}}$ [*with you*]$_{\underline{\text{COMITATIVE}}}$.

The comparison of the semantic descriptions provided by the three lexical resources points to the conclusion that at least some of the differences in the construals of verb senses may need to be reconsidered in a cross-resource perspective. The correspondences shown in Table 1 and the analysis of the examples demonstrate that some fine-grained distinctions made in a resource may lead to inaccurate predictions. For instance, the role COTHEME described in the VerbAtlas RUN frame (represented by verbs such as *run*) is just as applicable in the GO FORWARD frame (e.g. *walk*, Example 4), although it is not defined there, and the same is true for the RUN role PURPOSE. The Self_motion frame, which is evoked by both *run* and *walk* in the relevant senses, accounts for the respective fragments of the semantic description via the peripheral elements by the same names (Examples 4, 5).

Thanks to the alignment between VerbNet and PropBank, the VerbNet parser suggests an accurate label for the PURPOSE in Example 5, just as it does for the COTHEME in Example 4. In this case, the VerbNet-to-PropBank mapping provides a notable example of a richer and more robust semantic description obtained through combining the strengths of different resources.

Example 5: WALK

FrameNet frame: **Self_motion**

[*He*]$_{\text{SELF\_MOVER}}$ *WALKED* [*round*]$_{\text{PATH}}$ [*to stare out of the window*]$_{\text{PURPOSE}}$.

VerbNet class: **run-51.3.2.2-1**

[*He*]$_{\text{THEME}}$ *WALKED* [*round*]$_{\text{DESTINATION}}$ [*to stare out of the window*]$_{\underline{\text{PURPOSE}}}$.

Cross-resource comparison also matters, as it may point to possible distinctions that have gone unnoticed in one or more resources. Consider the synsets {dance:1} ('move in a graceful or rhythmical way') and {dance:2} ('move in a pattern; usually to musical accompaniment; do or perform a dance'). The former, like {walk:1} is assigned the VerbAtlas frame GO-FORWARD as it describes a manner of motion along a route, e.g. *She danced into the room*. The latter is described by the frame DANCE defined as 'An agent DANCES with a co-agent...', e.g. *He was dancing tango with his girlfriend on the shining dance floor*. The only lexical unit *dance.v* in FrameNet evokes the Self_motion frame. As attested by the Concise Oxford Dictionary definition provided for it, however, it encompasses two distinct dictionary senses: 'move rhythmically to music, typically following a set sequence of steps' and '2a: move in a quick and light or lively way', which may be considered as corresponding to the distinction made in VerbAtlas. Through this comparison, one is alerted to the fact that a reconsideration of the FrameNet description of *dance* may be necessary so as to account for the different senses.

In such a way, the detailed exploration of resemblances and differences in the assignments provides insights into the points of divergence in the construal of verb semantics and presents a challenge and an intriguing topic for research in both theoretical and practical terms.

6 A corpus of annotated examples

The corpus of annotated Bulgarian examples (Section 3.1) consists of sentences illustrating the use of a number of representative verbs from 5 highly frequent motion frames: Motion, Self_motion, Traversing, Arriving and Departing.

FrameNet provides a number of valence patterns for each lexical unit evoking a frame. Valence patterns represent co-occurring combinations of frame elements attested in the FrameNet corpus, i.e. the actual realisations of a lexical unit in context. Following the Berkeley FrameNet approach, each example in the Bulgarian corpus is supplied with annotation that marks both the frame elements realised in a sentence and their syntactic category and function. In such a way, information is obtained about: (i) the combinatorial properties of the annotated lexical unit, i.e. the particular combinations of both core and non-core frame elements that actually occur in the examples; (ii) the syntax-semantic interface, i.e. the regularities in the syntactic projection of the frame elements into syntactic positions; (iii) the syntactic groups realising the frame elements (along with their morpho-syntactic properties) and the prepositions, complementisers and other lexical items that introduce them.

Example 6 illustrates the manual annotation of a Bulgarian sentence, with explicit marking of the boundaries of the constituents and the type of phrase realising each core frame element and selected non-core ones. The English original of the sentence is provided for reference.

Example 6
FrameNet frame: `Self_motion`
A [*po vodnite alei*]$_{\text{PATH:PP}}$ *PLUVAHA* [*kucheshkata*]$_{\text{MANNER:AdvP}}$ [*oshte igove*]$_{\text{SELF\_MOVER:NP.Ext}}$.
And [*other Igs*]$_{\text{SELF\_MOVER:NP.Ext}}$ *were DOG-PADDLING* [*along the water lanes*]$_{\text{PATH:PP}}$.

An important feature of the FrameNet methodology and by extension of the annotation adopted in the Bulgarian corpus, is that syntactically non-overt but semantically obligatory frame elements are also marked (as null instantiations, NIs) cf. Ruppenhofer et al. (2016: 28–30), as illustrated below.

Example 7
FrameNet frame: `Self_motion`
[ ]$_{\text{SELF\_MOVER:DNI:NP.Ext}}$ *VARVYAHME* [*v neshto kato ledena zala*]$_{\text{PATH:PP}}$.
[*We*]$_{\text{SELF\_MOVER:DNI:NP.Ext}}$ *were WALKING* [*through a kind of ice hall*]$_{\text{PATH:PP}}$.

Table 2 (see the Appendix) illustrates the distribution of the most frequent valence patterns and their syntactic expression across the five studied frames as attested in the corpus of examples. For generalising purposes, only the configurations involving core frame elements have been presented.

Table 3 illustrates the typical valence patterns discovered for the verb *varvya* (one of the Bulgarian counterparts of the synset {walk:1}). Each of them is represented in a separate row along with the syntactic realisations attested in the dataset[12].

The corpus is being expanded both in terms of the verb coverage and the number of examples. As certain valence patterns are much more frequent than others, in order to obtain an accurate picture of the combinatorial properties of verbs and the syntactic realisation of frame elements other selection procedures are being tested.

A further extension inspired by the alignment with VerbNet and VerbAtlas that has been initiated involves the automatic assignment of the corresponding VerbNet and VerbAtlas semantic roles based on the identification of the counterparts of the FrameNet frame elements in the respective rolesets, as illustrated by the example in Table 4. Based on previous efforts (Leseva and Stoyanova, 2022), this mapping is also assisted by the automatic annotation of test sentences using the VerbAtlas SRL system and the VerbNet parser in order to confirm or reject possible frame element-to-semantic role alignments. This procedure will translate the interoperability of the resources into interoperability of the annotation. The Bulgarian corpus is currently made available for online search[13].

7 Conclusion

The work described in this paper represents a cursory dip-dive into verb semantics, aiming to delineate the specifics of: (i) a cross-lingual 'deep dive' study implemented through the adaptation of semantic resources developed for English to Bulgarian; (ii) a cross-resource analysis implemented through the exploration of the conceptualisation of verb semantics in resources constructed according to different theoretical frameworks. This undertaking has been enabled through two types of alignment: (i) a mapping between verbs (in WordNet) and their semantic descriptions represented as FrameNet semantic frames, VerbAtlas predicate-argument structure frames and VerbNet verb classes and predicate-argument structures; (ii) an alignment between equivalent facets of the semantic description, e.g. between corresponding frame elements and semantic roles or between a semantic frame and a predicate-argument structure. The end goal is the devising of a synergistic approach to the semantic description of verbs and the harnessing of the wealth of linguistic information to semantic annotation, semantic role labelling and similar tasks.

Acknowledgments

This paper is carried out as part of the scientific programme under the project *An Ontology of Activity Predicates – Linguistic Modelling with a Focus on the Bulgarian Language* funded by the Bulgarian National Science Fund (Grant Agreement No. KP–06–N80/9 of 8.12.2023).

[12]To avoid redundancy, nominal (NP.Ext) and pro-drop subjects are represented as aggregates in both tables. Pro-drop subjects are treated as an instance of definite null instantiation (DNI), i.e. as syntactically non-overt elements whose referent is retrievable from the broader context.

[13]The data are deployed at: https://dcl.bas.bg/corpus-data-semantic-frames-2024/.

References

Colin F. Baker and Christiane Fellbaum. 2009. WordNet and FrameNet as Complementary Resources for Annotation. In *Proceedings of the Third Linguistic Annotation Workshop (ACL-IJCNLP '09), Association for Computational Linguistics, Stroudsburg, PA, USA*, pages 125–129.

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 Resourcen*, 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.

Charles Fillmore. 2003. Valency and semantic roles: the concept of deep structure case. In Vilmos Ágel, Ludwig M. Eichinger, Hans Werner Eroms, Peter Hellwig, Hans Jürgen Heringer, and Henning Lobin, editors, *Dependenz und Valenz: Ein internationales Handbuch der zeitgenössischen Forschung*, pages 457 — 475. Walter de Gruyter.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. Language Resources and Evaluation. *Commun. 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. 2021. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences*, 8(1):47–70.

Svetla Koeva, Svetlozara Leseva, Borislav Rizov, Ekaterina Tarpomanova, Tsvetana Dimitrova, Hristina Kukova, and Maria Todorova. 2011. Design and development of the Bulgarian sense-annotated corpus. In *Information and communications technologies: present and future in corpus analysis: Proceedings of the III International Congress of Corpus Linguistics*, pages 143 – 150.

Svetla Koeva, Svetlozara Leseva, and Maria Todorova. 2006. Bulgarian sense tagged corpus. In *Proceedings of LREC 2006*, pages 79 – 86.

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.

Maddalen Lopez de Lacalle, Egoitz Laparra, Itziar Aldabe, and German Rigau. 2016. A Multilingual Predicate Matrix. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2662–2668, Portorož, Slovenia. European Language Resources Association (ELRA).

Shari Landes, Claudia Leacock, and R. Tengi. 1998. Building Semantic Concordances. In *WordNet: An Electronic Lexical Database*.

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

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 and Ivelina Stoyanova. 2022. Linked Resources towards Enhancing the Conceptual Description of General Lexis Verbs Using Syntactic Information. In *Proceedings of the Fifth International Conference Computational Linguistics in Bulgaria (CLIB 2022)*, pages 214 – 223. Institute for Bulgarian Language.

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*.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a Very Large Multilingual Semantic Network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics.

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.

Miriam R. Petruck. 2015. The Components of FrameNet. Tutorial at NAACL 2015. Http://naacl.org/naacl-hlt-2015/tutorial-framenet-data/FNComponentsMRLP.pdf.

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.

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*.

Appendix A

Valence pattern	Frame	Occurrences
[NP.EXT]$_{\text{THEME}}$ [PP]$_{\text{PATH}}$	Motion Self_motion Traversing	55 86 23
[NP.EXT]$_{\text{SELF\_MOVER}}$ [ ]$_{\text{PATH\_DNI}}$	Self_motion	45
[NP.EXT]$_{\text{THEME}}$ [PP]$_{\text{GOAL}}$	Motion Self_motion Arriving Departing	25 35 75 35
[NP.EXT]$_{\text{THEME}}$ [PP]$_{\text{AREA}}$	Motion Self_motion	15 37
[NP.EXT]$_{\text{THEME}}$ [ ]$_{\text{SOURCE\_DNI}}$	Departing	53
[NP.EXT]$_{\text{THEME}}$ [NP.OBJ]$_{\text{SOURCE}}$	Departing	38
[NP.EXT]$_{\text{THEME}}$ [PP]$_{\text{GOAL}}$ [ ]$_{\text{SOURCE\_DNI}}$	Departing	35
[NP.EXT]$_{\text{THEME}}$ [PP]$_{\text{SOURCE}}$	Departing	15
[NP.EXT]$_{\text{THEME}}$ [NP.OBJ]$_{\text{PATH}}$	Traversing	79
[NP.EXT]$_{\text{THEME}}$ [ ]$_{\text{GOAL\_DNI}}$	Arriving	36

Table 2: Distribution of the most frequent valence patterns across the studied motion frames.

SELF_MOVER	PATH	GOAL	SOURCE	DIRECTION	MANNER	TIME	COTHEME	No.
NP.EXT		PP						3
NP.EXT		ADVP						3
NP.EXT	PP							11
NP.EXT	PP				ADVP			3
NP.EXT	PP + PP							1
NP.EXT	PP						PP	1
NP.EXT	PP			PP				1
NP.EXT				ADVP				1
NP.EXT	PP			ADVP				1
NP.EXT	PP				PP			1
NP.EXT	PP				ADVP			1
NP.EXT					S	NP		1
NP.EXT						NP		1

Table 3: Distribution of the valence patterns and their syntactic realisations for the verb *varvya* in the synset {varvya:9; hodya:6} ({walk:1}). The data include 40 annotated sentences.

.

Annotation	**An annotated Bulgarian sentence and its English translation annotated in the styles of FrameNet, VerbNet and VerbAtlas**
Corpus of BG examples	[*Togava*]$_{\text{TIME:ADVP}}$ [*grazhdanite*]$_{\text{SELF\_MOVER:NP.EXT}}$ *BYAGAHA* [*s chanti i kufarcheta*]$_{\text{DEPICTIVE:PP}}$ [*po stramnite i tamni ulichki*]$_{\text{PATH:PP}}$ [*kam dvata tunela pod grada*]$_{\text{GOAL:PP}}$.
FN frame: `Self_motion`	[*Then*]$_{\text{TIME:ADVP}}$ [*the inhabitants*]$_{\text{SELF\_MOVER:NP.EXT}}$ *RAN* [*purses and briefcases in hand*]$_{\text{DEPICTIVE:SABS}}$ [*down the steep dark alleys*]$_{\text{PATH:PP}}$ [*to the two tunnels under the city*]$_{\text{GOAL:PP}}$.
VerbNet class: run-51.3.2	[*Then*]$_{\text{TIME:ADVP}}$ [*the inhabitants*]$_{\text{THEME:NP.SUBJ}}$ *RAN* [*purses and briefcases in hand*]$_{\underline{\text{ADVERBIAL}}}$ [*down the steep dark alleys*]$_{\text{TRAJECTORY:PP}}$ [*to the two tunnels under the city*]$_{\underline{\text{DIRECTION:PP}}}$.
VerbAtlas frame: RUN	[*Then*]$_{\text{TIME:ADVP}}$ [*the inhabitants*]$_{\text{THEME:NP.SUBJ}}$ *RAN* [*purses and briefcases in hand*]$_{\text{ATTRIBUTE}}$ [*down the steep dark alleys*]$_{\text{DESTINATION:PP}}$ [*to the two tunnels under the city*]$_{\text{DESTINATION:PP}}$.

Table 4: An example of a manually annotated sentence according to the FrameNet methodology, its English translation (for reference) and the automatic assignment of the roles from the relevant VerbNet class (if available) and VerbAtlas frame, using the VerbNet parser and the VerbAtlas SRL platform. The underlined semantic roles are assigned from PropBank.