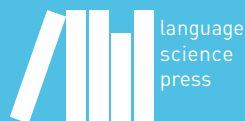


Universality of semantic frames and language specific Bulgarian data

Edited by

Svetla Koeva

Frame-Based Approaches to Semantics 1



Frame-Based Approaches to Semantics

Editors: Voula Giouli (Institute for Language and Speech Processing, ATHENA Research and Innovation Centre / National and Kapodistrian University of Athens, Greece), Ana Ostroški Anić (Institute for the Croatian Language, Croatia), Alexander Ziem (Heinrich-Heine-Universität Düsseldorf, Germany)

In this series:

1. Koeva, Svetla (eds.). Universality of semantic frames and language specific Bulgarian data

Universality of semantic frames and language specific Bulgarian data

Edited by

Svetla Koeva

Svetla Koeva (ed.). 2024. *Universality of semantic frames and language specific Bulgarian data* (Frame-Based Approaches to Semantics 1). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/474>

© 2024, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: **no digital ISBN**

no print ISBNs!

no DOI

Source code available from www.github.com/langsci/474

Errata: paperhive.org/documents/remote?type=langsci&id=474

Cover and concept of design: Ulrike Harbort

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: \LaTeX

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

<http://langsci-press.org>

Storage and cataloguing done by FU Berlin

Freie Universität



Berlin

Contents

Preface	iii
1 Universality of semantic frames vs. specificity of conceptual frames Svetla Koeva	1
2 Language-independent and language-specific properties of semantic description: A case study on verbs of communication Svetlozara Leseva & Ivelina Stoyanova	65
3 The complex conceptual structure of verbs of change Ivelina Stoyanova	129
4 The conceptualisation of the route: Non-directed and directed motion verbs in Bulgarian and English Svetlozara Leseva	171
5 Frame semantics and verbs of contact Maria Todorova	233
6 Frame semantics and verbs of emotion Hristina Kukova	255
7 Basic verb vocabulary: An empirical approach to argument structure and word associations Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova	293
Index	326

Preface

The book *Universality of semantic frames and language specific Bulgarian data* is dedicated to the principles of data organisation in the Bulgarian FrameNet, which has been under development for more than 20 years and has gone through various phases. Originally it was developed as an independent resource, but for about fifteen years it has been correlated with the Berkeley FrameNet, observing the following basic principles: The information in FrameNet that is relevant for the description of Bulgarian is considered language-independent (e.g. definition of frames and relations between them, definitions of frames and elements and relations between them, etc.) and automatically transferred into a structure called a superframe. For each superframe, there can be one or more Bulgarian frames in which the language-independent information is restructured, if necessary, so that it corresponds exactly to the description for the Bulgarian language. The Bulgarian verbs of communication, change, movement, contact and emotion are described in more detail, their subclasses are delineated and the similarities and differences in the semantic and syntactic description for Bulgarian and English are compared and discussed.

Chapter 1 *Universality of semantic frames vs. specificity of conceptual frames* introduces the Bulgarian FrameNet, which is based on the FrameNet and at the same time offers the possibility to encode language-specific semantic structures, either by replicating or reconstructing existing semantic frames or by introducing new frames. An abstract representation, called *superframe*, is developed to replicate language-independent information (at least for English and Bulgarian) from semantic frames. In Bulgarian FrameNet, the *conceptual frames* inherit either all or part of the language-independent information from the semantic frames via the superframes and may contain additional language-specific data to represent scenarios evoked by the Bulgarian lexical units. Each conceptual frame is extended by a set of nouns that represent the lexical realisations of the frame elements corresponding to the target lexical units.

The study presents the FrameNet (Fillmore et al. 2003, Fillmore & Baker 2009), the creation of FrameNets for other languages, the motivation for the introduction of conceptual frames and superframes that combine the semantic and conceptual frames in a *multilingual network*, and the structure of the Bulgarian Frame-

Preface

Net, which includes Lexical, Grammatical, Frame and Syntactic sections (valence patterns).

The overall aim is to present our approach to the identification and transfer of *language-universal* knowledge from the FrameNet semantic frames, which is universal in the sense that it applies to both English and Bulgarian, and the definition and integration of *language-specific components* of the conceptual frames for Bulgarian (as compared to English).

Chapter 2 *Language-independent and language-specific aspects of the semantic description of verbs: a case study of verbs of communication* focuses on the complex semantic description of verbs compiled from two main resources – the Princeton WordNet (Fellbaum 1999) and its Bulgarian counterpart BulNet (Koeva 2021) on the one hand and the FrameNet on the other. The verb synsets in WordNet are assigned FrameNet frames, which represent the frame elements that denote the participants and props of the predicate. This chapter discusses the notion of universality in terms of semantic and conceptual features and relations, which enables the cross-linguistic transfer of language-independent descriptions (that apply to at least two languages). In addition, this chapter presents a case study on verbs of communication, analysing the transferability of FrameNet valency patterns from English to Bulgarian and the language-specific features to be taken into account. Special attention is paid to the differences resulting from alternative constructions of identical or similar situations. The analysis of these differences can shed light on the factors underlying the different conceptualisations and help to identify tendencies and regularities in the representation of the conceptual/argument structure.

Chapter 3 *Complex conceptual structure of verbs of change* discusses the class of verbs of change in WordNet (Fellbaum 1999) and the conceptual frames from FrameNet (Fillmore et al. 1998) that describe them, together with the corresponding frame elements. The analysis begins with an overview of the treatment of verbs of change in theoretical studies, in particular with regard to the approaches to their classification based on the features of their semantic and conceptual description. The study establishes a link between the aspectual property of telicity and the notion of scale in relation to quantised change. Causation is discussed in terms of causative–inchoative verb pairs and the corresponding frames they evoke, which show similarities in their definitions and correspondences between the core frame elements involved and their semantic types and constraints. The frames discussed are connected by frame-to-frame relations (*Inheritance* and *Causativity*), which comprehensively describe the semantics of verbs of change together with the relevant aspects of the changes involved. The most frequent syntactic valency patterns associated with these verbs and their realisation in

Bulgarian are presented on the basis of observations from the Bulgarian semantically annotated corpus (Koeva et al. 2011).

Chapter 4 *On conceptualising the aspects of motion* provides an analysis of undirected and directed motion from the perspective of frame semantics by examining the semantic description and syntactic realisation of the frame elements of several FrameNet frames describing motion and drawing parallels between the syntactic properties of motion verbs in English and Bulgarian. The research questions include: Which frame elements are conceptually present in the semantics of verbs across motion-related frames, even if they are not obligatorily expressed; which frame elements are more prominent in the semantics of verbs evoking the same frame; what are the possible and preferred means of syntactic expression of these frame elements; what are the patterns of syntactic expression; and how are the investigated frame elements represented, as opposed to how they may remain syntactically implicit. The empirical evidence provided by the FrameNet corpus is validated using a sample of annotated Bulgarian examples. The second part of the chapter contains a case study of several representative frames from the domain of communication. The study illustrates how the semantics of the parent frame is further specialised, i.e. narrowed down, profiled, etc. in the inheriting frames and how this process is reflected in the configuration of the frame elements describing each frame and in their syntactic expression in English and Bulgarian.

Chapter 5 *Frame semantics and verbs of emotion* is devoted to the presentation of the peculiarities of the general verbs of emotion together with their frames and frame element representations. This is followed by a description of the semantic class of emotion verbs and an analytical overview of the typological approaches of various authors. The aim of the study is to analyse the most general emotion verbs and to group them according to their semantic and syntactic features. Thus, within this class of verbs, five major subclasses are formed, namely: basic neutral verbs of emotion, verbs for emotional attitude, causative verbs of emotion that encode an Agent or Stimulus as subject, and stative or inchoative verbs formed from the causatives using the reflexive by-form particle *se*. The main focus is on five frames, namely *Feeling*, *Experiencer_focused_emotion*, *Stimulate_emotion*, *Cause_to_experience* and *Emotion_directed*. They are all characterised in terms of the verbs they cover, their core frame elements and the possible representations they can have with regard to their syntactic realisation. The verbs that evoke the frame *FEELING* have neutral semantics. The verbs of the other four semantic frames denote a positive or negative emotion in their semantic structure.

Chapter 6 *Frame semantics and verbs of contact* provides a semantic description of verbs from WordNet that belong to the general Bulgarian vocabulary and to the semantic class of contact verbs. Taking into account the information about the semantic hierarchy from BulNet (Koeva 2021), WordNet and the semantic frames from FrameNet, the contact verbs are grouped into two main classes of predicates: Verbs of physical contact by motion and verbs of physical contact in state. The main semantic frame for verbs of contact covers events in which two or more entities come into physical contact with each other. Semantic subgroups within the verbs of physical contact by movement and the verbs of physical contact in state, based on frame semantics, are provided, along with the description of syntactic properties and the definition of more specific selectional restrictions for each verb. The study makes statements about the internal semantic organisation of verbs within the domain of verbs with the semantic attribute physical contact.

Chapter 7 *Basic verb vocabulary: An empirical approach to Argument Structure and word associations* deals with the results of a study on the degree of acquisition of a number of verbs in Bulgarian that are considered part of children's general vocabulary. The pilot experiment involved language tasks designed to test whether the verbs in the extracted set belong to the basic conceptual apparatus and whether this approach is suitable for making observations about the respondents' linguistic knowledge, experience and intuition regarding the use of these verbs. The analysis of the target verbs is based on the description of both semantic and conceptual frames. The pilot results are presented with regard to the semantic frames of the target verbs, the lexical entries in argument positions and the semantic and syntactic combinability.

The semantic frames can be used in an experiment to assess children's mastery of semantic conceptualisation and syntactic use of verbs from their basic vocabulary. These and a number of other applications: automatic assignment of semantic roles, automatic recognition of events in news, automatic recognition of scenes in images and videos are some of the applications in which the Bulgarian FrameNet can be used. In addition, the semantic and syntactic information in the Bulgarian FrameNet can be used for theoretical considerations, including comparative studies focussing on the modern state of the Bulgarian language and other languages for which a FrameNet has been developed.

References

- Fellbaum, Christiane (ed.). 1999. *WordNet: An electronic lexical database*. Cambridge: MIT Press.

References

- Fillmore, Charles J. & Collin F. Baker. 2009. A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 313–340. Oxford: Oxford University Press.
- Fillmore, Charles J., Collin F. Baker & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the Conference COLINGACL '98, Montreal, Canada*, 86–90.
- Fillmore, Charles J., Christopher R. Johnson & Miriam R. Petruck. 2003. Background to framenet. *International Journal of Lexicography* 16(3). 235–250. DOI: 10.1093/ijl/16.3.235.
- Koeva, Svetla. 2021. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla, Svetlozara Leseva, Borislav Rizov, Ekaterina Tarpomanova, Tsvetana Dimitrova, Hristina Kukova & Maria Todorova. 2011. Design and development of the Bulgarian sense-annotated corpus. In María Luisa Carrió Pastor & Miguel Ángel Candel Mora (eds.), *Proceedings of the III international congress of corpus linguistics. Information and communications technologies: Present and future in corpus analysis*, 143–150.

Chapter 1

Universality of semantic frames vs. specificity of conceptual frames

Svetla Koeva 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

FrameNet is a semantic network that links semantic frames, each evoked by a set of lexical units and consisting of frame elements (with semantic types, definitions and relations) that outline the semantic structure of the frame, as well as frame-to-frame relations and annotations that illustrate the syntactic realisation of the frame elements.

The Bulgarian FrameNet is based on the FrameNet and at the same time offers the possibility to encode language-specific semantic structures, either by replicating or reconstructing existing semantic frames or by introducing new frames. An abstract representation, called *superframe*, is developed to replicate language-independent information (at least for English and Bulgarian) from semantic frames. In the Bulgarian FrameNet, the *conceptual frames* inherit either all or part of the language-independent information from the semantic frames via the superframes and may contain additional language-specific data to represent scenarios evoked by the Bulgarian lexical units. Each conceptual frame is extended by a set of nouns that represent the lexical realisations of the frame elements corresponding to the target lexical units.

The study presents the FrameNet, the creation of FrameNets for other languages, the motivation for introducing the conceptual frames and the superframes, which combine the semantic and conceptual frames in a *multilingual network*, and the structure of the Bulgarian FrameNet, which includes Lexical, Grammatical, Frame and Syntactic sections (valence patterns).

The overall aim is to present our approach to the identification and transfer of *language-universal* knowledge from the FrameNet semantic frames, universal in the sense that it applies to both English and Bulgarian, and the definition and integration of *language-specific components* of the conceptual frames for Bulgarian (as compared to English).



Svetla Koeva. 2024. Universality of semantic frames vs. specificity of conceptual frames. In Svetla Koeva (ed.), *Universality of semantic frames and language specific Bulgarian data*, 1–63. Berlin: Language Science Press. DOI: ??

Svetla Koeva

1 Introduction

FrameNet is a semantic network that links semantic frames, each evoked by a set of lexical units and consisting of frame elements (with semantic types, definitions and relations) that outline the semantic structure of the frame (Fillmore et al. 2003, Fillmore & Baker 2009). It also includes frame-to-frame relations and contains syntactic and semantic annotations of examples that illustrate the syntactic realisation of frame elements.

The study presents the structure of the Bulgarian FrameNet, which is based on two basic principles: maintaining consistency with FrameNet and providing a mechanism for encoding semantic structures that either replicate or reconstruct the existing semantic frames or are completely new. To achieve this, an abstract level of representation, the *superframe*, is introduced, which contains the *language-independent information* inherited from the semantic frames. To represent the semantic structure of Bulgarian lexical units that evoke the same situation, property or process, an abstract structure, the *conceptual frame*, is introduced, which is influenced by the semantic frames of FrameNet. The conceptual frame:

- applies only to lexical units described by the same set of core frame elements, which in turn have the same syntactic realisation and lexical compatibility;
- is extended by nouns that can form semantically valid phrases with the verbal lexical units that evoke the frame.

A superframe is linked to exactly one semantic frame, while a superframe can be connected to one or more conceptual frames. Three models of correspondence between a conceptual frame and a superframe are described: (a) equivalence, (b) partial equivalence and (c) no equivalence. The superframe is introduced to ensure alignment with language-independent information from FrameNet that is valid for at least two languages, English and Bulgarian, while conceptual frames are used to delineate semantic and syntactic differences in conceptual descriptions of Bulgarian lexical units. This representation enables the integration of Bulgarian into a global network that captures both unique semantic and syntactic features of individual languages as well as language-independent features that may apply to a large group of languages.

1 *Universality of semantic frames vs. specificity of conceptual frames*

In the following sections, we present the structure of FrameNet, the creation of FrameNets for other languages, the motivation for the introduction of superframe and conceptual frame, and the relations between the two abstract structures. This is followed by an overview of the structure of the Bulgarian FrameNet, which comprises four sections: Lexical, Grammatical, Frame and Syntactic, all of which are integrated into a web-based data management system called BulFrame (Koeva & Doychev 2022). This system facilitates the manual evaluation and visualisation of the Bulgarian FrameNet.¹

We describe the components that make up the Lexical, Grammatical, Frame, and Syntactic section of each conceptual frame and present their components, sources and associated data. The Bulgarian lexical units are provided with additional grammatical, lexical and semantic information. The frame elements in the Bulgarian FrameNet are associated to nouns that are suitable for collocations with verbal lexical units that evoke the corresponding frame. Based on annotated examples, each frame element is linked to the relevant syntactic categories, grammatical roles and labels for implicit use related with its lexical representation.

The contributions of the study are as follows (a) formulation of an abstract structure, the superframe, to connect the semantic and conceptual frames in a cross-linguistic network; (b) identification of language-independent knowledge (for at least two languages, in our case for English and Bulgarian) in the semantic frames of FrameNet for transfer to the superframes; (c) definition of conceptual frames based on the structure of the semantic frames of FrameNet and their extension with components containing additional lexical and grammatical information; (d) associating the conceptual frame elements with sets of nouns that can be collocated contextually with the lexical units evoking the frame; and (e) developing the network of conceptual frames valid for Bulgarian, containing both language-independent information from the corresponding semantic frames and language-specific information for Bulgarian.

The BulFrame system for editing, evaluating and visualising data as well as the results of the annotation are presented in various studies, e.g. in (Koeva 2024, Koeva & Doychev 2022) and the other contributions in this volume.

2 **Semantic and syntactic representations in FrameNet**

FrameNet is based on the theory of Frame semantics (Fillmore 1982, 1976a,b, 1985, 2006, 2008, Fillmore & Baker 2009), going beyond the general semantic roles of Case Grammar (Fillmore 1968).

¹<https://dcl.bas.bg/bulframe/>

Svetla Koeva

The central idea of Frame semantics is that word meanings are described in relation to semantic frames, which are schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation for meaningful communication within a particular speech community (Fillmore et al. 2003: 235). Semantic frames are defined more concisely as schematic representations of speakers' knowledge of the situations or states of affairs that underlie the meanings of lexical items (Fillmore 2007: 130). A frame-bearing lexical unit evokes a frame, and a valency description of a specific lexical unit presents the ways the semantic valents are expressed in the sentence built around the frame-bearing unit (Fillmore 2007: 131).

FrameNet is a collection of semantic frames (each evoked by a set of lexical units associated with valency patterns) that represent conceptual-semantic and syntactic descriptions based on the annotation of examples. The semantic frame in FrameNet includes the following components: the frame name; the informal definition of the situation represented by the frame; a specification for the semantic type of the frame (optional); the set of frame elements (core and non-core: peripheral, extrathematic and core-unexpressed); a specification for the relations between frame elements, if any; a specification for frame-to-frame relations, if any; and the lexical units that evoke the frame.

The frame element information includes the name of the frame element, its informal definition, the semantic type (optional) and examples illustrating the use of the frame element (optional). The information on the lexical units includes a definition, the semantic type (optional), examples and annotation in the examples of the frame elements as well as the grammatical categories and grammatical functions of their syntactic realisations.

Frame semantics thus links lexical units with both linguistic and conceptual information. The linguistic information consists of the frames as predicate classes, the sets of frame elements associated with them and their valency patterns. The conceptual information comprises the descriptions of situations and their participants as well as the relations between the frames (Sikos & Pado 2018).

Two types of criteria were used to formulate the semantic frames (Ruppenhofer et al. 2016: 11–17): a checklist of features and other principles such as paraphrases and alternative answers to a question.

The checklist of features includes (Ruppenhofer et al. 2016: 12–14):

- the same number and type of frame elements for all lexical units;
- the same set of stages and transitions (sub-events) shared by lexical units, i.e. unlike the verb *decapitate*, the verb *shoot* can be used to report the event

1 Universality of semantic frames vs. specificity of conceptual frames

of firing and hitting at a person, but it does not entail that the person dies, thus, the two verbs should belong to different frames;

- the same participants point of view, i.e. since the verb *buy* takes the point of view of the BUYER and the verb *sell* takes the point of view of the SELLER, they belong to different semantic frames;
- the same interrelations between frame elements for all lexical units, i.e. a PURPOSE expressed with the verb *buy* relates to the BUYER, a PURPOSE expressed with the verb *sell* relates to the SELLER, and the different relations indicate participation in different semantic frames;
- the same presuppositions, expectations, and concomitants of the target lexical units, i.e. the verb *receive* presupposes a willing AGENT participating as a DONOR while the verb *take* does not;
- the similar basic denotation of the lexical units (similarity of type);
- the similar pre-specifications given to frame elements by frame-evoking lexical units, i.e. verbs as *crowd*, *flock*, *pour*, *stream*, etc., are part of the frame Mass_motion but not of the frame Self_motion since they require that the moving entity is a Mass_theme, which generally consists of many individuals.

The development of frames is also based on the *paraphrasability* (or near-paraphrasability) of lexical units: whether one lexical unit can be more or less successfully replaced by another, while evoking the same frame and the same configuration of frame elements. One and the same semantic frame can be evoked by synonyms, near synonyms, antonyms, derivationally related lexical units, hypernyms, or hyponyms. For example, the verb *hate* with the definition ‘feel intense dislike for or a strong aversion towards’ is a synonym of the verb *detest* with the definition ‘dislike intensely’ in the semantic frame *Experiencer_focused_emotion*. Both verbs have a hypernym *dislike* with the definition ‘feel distaste for or hostility towards’, the verb *resent* with the definition ‘feel bitterness or indignation at’ is also a *hyponym* of *dislike* and it has hyponyms such as *abhor*, *abominate*, and *despise*. On the other hand, verbs like *excrate*, *contemn*, *scorn*, *disdain* are not presented in the frame as of September 2024. **Multiword expressions** are also included, albeit relatively rarely.

As in dictionaries, the lexical units of FrameNet are provided with *definitions*, which were either taken from the Concise Oxford Dictionary, 10th Edition (courtesy of Oxford University Press) or written by the FrameNet developers (Ruppenhofer et al. 2016: 9).

Svetla Koeva

In FrameNet, the frame elements are classified according to how central they are in a particular frame, whereby three levels are distinguished: core, peripheral and extrathematic. A **core frame element** is an element that is necessary for the central meaning of the frame (Fillmore 2007: 133) and that represents a conceptually essential component of a frame and distinguishes the frame from others (Ruppenhofer et al. 2016: 23). **Peripheral frame elements** mark such notions as TIME, PLACE, MANNER, MEANS DEGREE and the like. They do not distinguish between different frames and can be instantiated in any semantically suitable frame (Ruppenhofer et al. 2016: 24). **Extrathematic frame elements** are understood as not conceptually belonging to the frames they appear in. They are part of other abstract frames and situate the event against the backdrop of another event (Fillmore 2007: 133). The **Core-unexpressed** property refers to frame elements that function as core frame elements but do not appear in descendants of that frame. In child frames, however, the Core-unexpressed frame element is absorbed by the lexical units in the frame and cannot be represented individually (Ruppenhofer et al. 2016: 25).

In FrameNet, some formal properties, typically co-present, are taken into account when selecting the core frame elements. The core frame element (Ruppenhofer et al. 2016: 23–24):

- should be specified openly;
- if omitted, it receives a definite interpretation (in the sentence *John arrived* a certain frame element – GOAL LOCATION – is understood; GOAL is therefore a core frame element);
- has no formal marking (its interpretation depends entirely on the target: i.e. frame elements that can be subject or object in a simple active sentence in English, or has an idiosyncratic formal marking (i.e. the preposition *on* in *depend on* has no semantic meaning).

Although some of the names of the frame elements correspond to the names of the semantic roles, the names of the frame elements only serve a mnemonic purpose (Fillmore et al. 2003: 237). The definitions of frame elements are statements that express the semantics of a particular frame element in relation to the target lexical unit (and possibly in relation to other frame elements).

It has been established that the frame elements are not necessarily independent of each other. Some groups of frame elements behave like sets (called Core Sets), since the existence of any member of the set is sufficient to fulfil the semantic valency of the predicator (Ruppenhofer et al. 2016: 25). For example, SOURCE,

1 Universality of semantic frames vs. specificity of conceptual frames

PATH and GOAL core frame elements in motion frames form a **Core Set** in the sense that only one or two (rarely all three) frame elements can occur in a sentence without violating the semantic structure.

The relation *Requires* is coded if the occurrence of a core frame element presupposes that another core frame element also occurs. The relation *Excludes* is observed if one of the frame elements from a group of conceptually related frame elements occurs and no other frame element from this group can occur (Ruppenhofer et al. 2016: 26). For example, the frame elements GOAL and ITEM complement each other in the frame Attaching and exclude the frame element ITEMS:

- (1) *The robber TIED [Harry]_{ITEM} [to the chair]_{GOAL}.*
- (2) *The robber TIED [Harry's ankles]_{ITEMS} together.*

The FrameNet frames are linked by a system of nine *frame-to-frame relations*, seven of which fall into three groups: Generalisation, Event structure, and Systematic (Fillmore & Baker 2009: 806-807). FrameNet can therefore be seen as a semantic net (or a set of small semantic nets) whose nodes represent the semantic frames and whose arcs represent the (semantic) relations between the frames.

Generalisation relations are *Inheritance*, *Perspective on* and *Using*. In the relation *Inheritance* (represented by directed (asymmetric) relations *Inherits from* and *Is Inherited by*), the frame elements of the parent frame are bound to the frame elements of the child frame, whereby the names of the child frame elements can be different. The semantics of the child frame is therefore a subtype of the semantics of the parent frame, and the child frame can contain additional frame elements (Fillmore & Baker 2009: 330). For example, the Inheritance relation exists between the frame Revenge and the frame Rewards_and_Punishment because the frame Revenge involves one person inflicting punishment on another, as in its parent frame, the frame Rewards_and_Punishment. However, the frame Revenge is explicitly different from the frame Rewards_and_Punishments as it is outside institutional or judicial control (Fillmore & Baker 2009: 330).

It is also asserted that the *Inheritance* relation corresponds to the *is-a* relation in ontologies and that every semantic fact about the parent frame must correspond to an equally specific or more specific fact about the child (Ruppenhofer et al. 2016: 80). The complexity of the Inheritance relation can manifest itself in different ways (Ruppenhofer et al. 2016: 81): parent and child frames can have different extrathematic frame elements; a child frame can have frame elements that are not present in the parent frame or such that are extrathematic in the parent frame; a child frame often does not express the parent frame elements of

Svetla Koeva

type CORE-UNEXPRESSED; a frame element of a child frame can be mapped to two frame elements of the parent frame; etc.

Table 1 illustrates the *Inheritance* relation between the frame *Experiencer_focused_emotion* and its successor frames: *Desiring* and *Mental_stimulus_exp_focus*.

As the example shows, the relations between the frame elements of the frames connected via the *Inheritance* relation are quite complex: omission of a core frame element, i.e. the frame element TOPIC in the frame *Desiring*; specification of child frame elements, which is indicated by the names of the frame elements, i.e. STIMULUS in the frame *Mental_stimulus_exp_focus*, defined as “the person, event or state of affairs that evokes the emotional response in the EXPERIENCER”, corresponding to the more general frame element CONTENT in the frame *Experiencer_focused_emotion*, defined as “what the Experiencer’s feelings or experiences are directed towards or based upon; the CONTENT differs from a stimulus because the CONTENT is not construed as being directly responsible for causing the emotion”. In addition, the CONTENT can be expressed by one or both of the frame elements FOCAL_PARTICIPANT (“the entity that the EXPERIENCER wishes to be affected by some EVENT”) and EVENT (“the change that the EXPERIENCER would like to see”) in the frame *Desiring*; etc. Although the frame elements of the parent frame are by and large retained in the child frames linked by *Inheritance*, the example shows that some frame elements of the parent frame can be omitted in the child frame.

The relation *Perspective on* (represented by directed (asymmetric) relations *Perspectivises* and *Is Perspectivised in*) encodes the different perspectives on an abstract event (Fillmore & Baker 2009: 867). The use of this relation indicates the existence of at least two different possible points of view on the neutral frame. The commercial transaction scenario, where buying and selling are seen as different perspectives on the transfer of goods (*Commerce_goods_transfer*) and paying and accepting money are seen as different perspectives on the transfer of money (*Commerce_money_transfer*), is an example that is frequently analysed in the FrameNet literature. It has also been shown that frames with perspectives are often non-lexical and abstract (Osswald & Van Valin 2014: 131).

In the relation *Using* (with its members: *Uses* and *Is Used by*), the child frame is dependent on the background knowledge provided by the parent frame; at least some of the core frame elements of the parent frame are bound to child frame elements, but not all (Fillmore & Baker 2009: 330). For example, the frame *Being_* - attached with a definition ‘An ITEM is attached by a HANDLE, via a CONNECTOR, to a GOAL, or ITEMS are attached to each other’ is *Used by* the frame *Being_* -

1 Universality of semantic frames vs. specificity of conceptual frames

Table 1: The *Inheritance* relation between the frame `Experiencer_focused_emotion` and its successor frames, expressed by frame elements

Frames	Experiencer_focused_emotion	Desiring	Mental_stimulus_exp_focus
CORE FRAME ELEMENTS (FES)			
EXPERIENCER	Yes	Yes	Yes
CONTENT	Yes	FOCAL_PARTICIPANT; EVENT	STIMULUS
TOPIC	Yes		Yes
EVENT	Yes	LOCATION_OF_EVENT	
CORE UNEXPRESSED FES			
EXPRESSOR	Yes		Yes (core)
STATE	Yes		Yes (core)
PERIPHERAL FES			
DEGREE	Yes	Yes	Yes
MANNER	Yes	Yes	Yes
TIME	Yes	Yes	Yes
EXPLANATION	Yes	Yes	Yes
CIRCUMSTANCES	Yes		Yes
PARAMETER	Yes		Yes
EMPATHY_TARGET			Yes
DURATION		Yes	Yes
PURPOSE_OF_EVENT		Yes	
ROLE_OF_FOCAL_PARTICIPANT		Yes	
TIME_OF_EVENT		Yes	
PLACE		Yes	

Svetla Koeva

detached with the definition ‘An ITEM is detached from a SOURCE, or ITEMS are detached from each other’.

- (3) *It seems that [the nits]_{ITEM} are shiny ATTACHED [to the hair]_{GOAL}.*
- (4) *I feel like [my head]_{ITEM} is DETACHED [from the rest of my body]_{SOURCE}.*

Event structure relations are *Subframe* and *Precedes* (Fillmore & Baker 2009: 867).

Subframe relation (*Subframe of* and *Has Subframe(s)*) is used when the child frame is expressed as a sub-event of a more complex parent event. For example, the frame *Criminal_process* has four subframes: *Arraignment*, *Arrest*, *Sentencing*, and *Trial*.

Precedes relation (*Precedes* and *Is Preceded by*) indicates that there is a temporal order between the frames: the parent frame precedes the child frame. For example, the frame *Employment_continue* *Precedes* the frame *Employment_end* and *Is Preceded by* the frame *Employment_start*.

Causative of and *Inchoative of* are **Syntactic** relations (Fillmore & Baker 2009: 331). In the relation *Causative of*, the parent frame represents the causative that corresponds to the child frame. In the relation *Inchoative of*, the parent frame represents the inchoative and the child represents the stative. For example, the frame *Cause_to_fragment* is related to the frame *Breaking_apart* by the relation *Causative of*. The frame *Cause_to_fragment* has an AGENT as part of its conceptual core structure, while the frame *Breaking_apart* does not and expresses the AGENT as an oblique.

Furthermore, if there are groups of frames that are similar and should be carefully distinguished, each of the frames in question has a *See Also* relation with a representative member of the group; *Metaphor* is a relation between a source frame and a target frame in which many or all of the lexical units of the target frame are at least partially understood in terms of the source frame (Ruppenhofer et al. 2016: 85).

According to Fillmore, the implementation of Frame semantics in FrameNet should lead to correct frame-to-frame relations, including generalisations about how syntactic roles are assigned to arguments that depend on the more abstract inherited schemas (Fillmore 2007: 157). Developing a consistent relational structure of frames with different degrees of abstraction is a key challenge for the FrameNet approach, as certain case studies show (Osswald & Van Valin 2014: 153). At the same time, the addition of new frame-to-frame relations together with proposals for distinguishing subtypes within existing relations (Sikos & Pado 2018:

1 Universality of semantic frames vs. specificity of conceptual frames

12–19) emphasises both the complexity of the conceptual information presented and the potential for its extension.

Figure 1 provides an overview of the connectedness between frames in FrameNet.

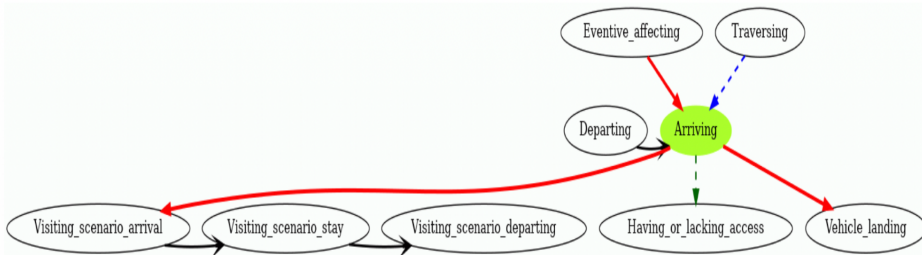


Figure 1: The immediate frame-to-frame relations of the semantic frame Arriving. Red arrows *Inheritance*, black – *Precedes*, green – *Using*, blue – *Subframe*, the direction is parent-child, the dashed lines show inverse relations.

The semantic types in FrameNet are used for (Ruppenhofer et al. 2016: 86):

- Marking of frames for their function.
- Specification of the basic typing of fillers for frame elements.
- Marking important dimensions of semantic variation between the lexical units in a frame.

Lexical units, frames and frame elements are categorised according to **ontological semantic types**. For example, the semantic type [Region] is assigned to the lexical unit *island.n* in the frame *Natural_features*, while the type [Body of water] is assigned to the lexical unit *bay.n*.

For frames, the semantic type indicates that each lexical unit of the frame can be labelled with an equivalent or more specific type. For example, the frame *Clothing* has the semantic type [Artefact]. Consequently, all its lexical units denote artefacts, i.e. *boot.n*, *cape.n*, *dress.n*, etc. (Lönneker-Rodman & Baker 2009: 422–423).

Semantic types for frame elements classify the type of filler that is to appear as a frame element. Not all frame elements (and frames) have a specific semantic type, and in general semantic types are too broad, so they lack precision when it comes to conveying actual constraints on lexical combinations. For example, certain frame elements within the semantic frame *Experiencer_focused_emotion*

Svetla Koeva

have rather general semantic types: **CONTENT** with the semantic type [Content]; **EVENT** with the semantic type [State of affairs]; **EXPERIENCER** with the semantic type [Sentient]; **DEGREE** with the semantic type [Degree]; **EXPLANATION** with the semantic type [State of affairs]; **MANNER** with the semantic type [Manner]; **TIME** with the semantic type [Time]. On the other hand, some frame elements such as **TOPIC**, **EXPRESSOR**, **STATE** are not specified with a semantic type.

Framal types are applied to frames. The type [Non-lexical] characterises frames that have no lexical units but are used to semantically connect frames in a network. The type [Non-perspectivized] is used for frames that consist of a large number of lexical units that are connected by a common scene as a background. These frames usually lack a consistent set of frame elements for the targets, a consistent assignment of time to events or players and a consistent point of view between the targets, e.g. the frame *Performers_and_roles*, which contains lexical units as different as *co-star.v*, *feature.v* and *as.prep* (Ruppenhofer et al. 2016: 87).

Annotations of examples (originally mainly from the British National Corpus) are provided for lexical units. The annotations show the variety of syntactic manifestations of individual frame elements in the corpus (including zero realisations), together with the patterns of frame element realisations in sentences (Fillmore 2007: 132).

The syntactic annotation includes the labelling of **grammatical categories** and the **grammatical functions** of sentence constituents in relation to a particular target lexical unit.

The principal grammatical functions are External, Object and Dependent; the other grammatical functions are Appositive, Modifier, Head, Genitive and Quantifier, which are particularly important for nouns (Fillmore 2007: 135). The grammatical function External corresponds not only to the subject of a finite sentence but also to the phrases that stand for the subject function of non-finite verbs, e.g., the controllers of subject roles in Raising and Equi constructions and subordinated participial constructions, and to the primary arguments of frame-bearing nouns and predicatively used adjectives (Fillmore 2007: 135). The function Dependent is used for all other dependents of a verbal predicate (other than External and Object).

The annotated examples show that some frame elements are restricted to certain parts of speech, suggesting that it might be a slightly different scene and raising questions such as: Is there a difference between frame elements of targets from different parts of speech that evoke one and the same frame, and what is the inheritance relation for targets from different parts of speech?

1 *Universality of semantic frames vs. specificity of conceptual frames*

The top-down approach to frame creation and annotation is described as follows (Lönneker-Rodman & Baker 2009: 418–419):

- Selection of a semantic domain and outline of the frames involved.
- Definition of the frames and their frame elements and selection of the lexical units, each with a short definition.
- Determination of the principal syntactic patterns and extraction of examples for each pattern from a large corpus.
- Annotation of a sufficient number of examples to prove all relevant syntactic realisations of each frame element. FrameNet has extended its annotation to continuous text. In full-text annotation, all content words are annotated, leading to the addition of new lexical units within existing frames and (less frequently) the creation of new frames.
- Development of the FrameNet annotation view and the lexical entry view.

As it has been pointed out, the semantic and syntactic descriptions in FrameNet differ from other lexical resources in several ways (Fillmore 2007: 129), including: a) its reliance on corpus evidence; b) its foundation on knowledge of the cognitive (semantic) frames that motivate and underlie the meanings of each lexical unit; c) its recognition of various types of discrepancies between lexical units on the semantic level and patterns of syntactic form; and d) its provision of the means of assigning partial interpretations to frame elements that are conceptually present but syntactically unexpressed.

3 **FrameNets for other Languages**

FrameNet has been largely extended to other languages (Boas 2009a), such as Spanish (Rüggeberg & Petruck 2003, Subirats 2009), Japanese (Ohara et al. 2004, Ohara 2012), German (Burchardt et al. 2009), Chinese (You & Liu 2005), Italian (Lenci et al. 2010), Swedish (Borin et al. 2010), Brazilian Portuguese (Torrent, Salomão, et al. 2014), French (Candito et al. 2014), Hebrew (Hayoun & Elhadad 2016), Danish (Pedersen et al. 2018), Czech (Materna & Pala 2010), and many others (Torrent et al. 2020).

When creating lexical-semantic networks, two basic approaches are usually used: the expand model and the merge model (Vossen 1996: 716). The first approach is to translate the lexical units, their definitions and (possibly) usage examples from one language (usually English) into another and to transfer (and

Svetla Koeva

manually or semi-automatically check) all the relations between the lexical units as well as the remaining semantic information.

The task of FrameNets for other languages, which are created by the expand model (i.e. by searching for translation equivalents of language units), is to encode the language-specific features that can be expressed both semantically (by the number and relations of the frame elements) and at the grammatical level. In general, it can be said that many differences at the semantic level between languages are due to their different grammatical structures and, to a lesser extent, to the encoding of different features of the real world.

For example, the Spanish FrameNet describes the meaning of lexical units by drawing directly on the frames already constructed for English and analysing the grammatical constructions in which these lexical units are instantiated (Subirats 2009: 136). If the English frames are not compatible with the Spanish language, the inconsistencies are resolved by restructuring the frames.

The second approach is to merge existing language resources for a particular language with other lexico-semantic resources for another language (usually English). One example is the Czech FrameNet, which was created by linking the independently developed Verbalex (a lexicon of verb valency for Czech) with the FrameNet (Materna & Pala 2010). The independent development of FrameNets may face the problem of achieving sufficient overlap in lexical coverage while maintaining language-specific properties.

It was found that there are two primary strategies for FrameNet development: a lemma-by-lemma strategy, that provides annotations that reflect the overall ambiguity of a given lemma within a target corpus, and a frame-by-frame strategy, that enforces the coherence of annotations within a frame (Candito et al. 2014: 1373).

The frame-by-frame approach, which is used by most FrameNets, takes into account the entire lexical diversity available for the expression of a frame (Candito et al. 2014: 1373). However, only the senses of a particular lemma related to covered frames are taken into account, and these senses are not necessarily the most frequent.

The lemma-by-lemma strategy considers different lemma senses for which there is often no frame in the English FrameNet, including rare senses. During the development of the German FrameNet (SALSA), each instance of a lemma in a corpus was annotated and tested for a FrameNet frame. Proto-frames were created for lemmas that could not be defined by existing frames. The proto-frames contain a single lexical unit and are not coupled with frame-to-frame relations (Burchardt et al. 2009: 213).

1 Universality of semantic frames vs. specificity of conceptual frames

Some FrameNets are built entirely by experts (manually), both the mapping to English and the semantic and syntactic annotation, while others rely on automatic or semi-automatic mapping or annotation, possibly using post-validation, such as the Italian FrameNet (Lenci et al. 2010).

Some of the linguistic issues that have arisen in the development of other FrameNets have been discussed in (Boas 2009a): degree of overlapping cross-lingual polysemy, differences in lexicalisation patterns, measurement of paraphrase relations (words that evoke a particular meaning may differ in different sentences) and translation equivalence.

FrameNet is used extensively for the development of multilingual resources, and two general approaches to FrameNet integration can be distinguished: either building on the English FrameNet infrastructure as a foundation (Boas 2009b, Rüggeberg & Petruck 2003) or by (semi-)automatically creating frame-based multilingual resources (Peron-Corrêa et al. 2016, Torrent, Salomao, et al. 2014). The first approach uses the semantic frames as interlingual representations to connect different parallel lexicon fragments and involves several steps:

- Removing all language-specific information for English, including lemma, parts of speech and annotated sentences, and retaining only the information that is not specific to English – frames, frame-to-frame relations, frame elements and frame element relations.
- Repopulating the database to create a non-English FrameNet (Boas 2009b: 72).

The (semi-)automatic creation of FrameNet-like resources involves the use of existing linguistic frameworks or corpora to extract semantic frames, frame elements and their relations to each other. Computational methods are employed to automatically identify frames in large datasets and annotate examples. This process includes the extraction of frame elements and the creation of frame-to-frame relations. The aim is to create a comprehensive lexical-semantic resource, similar to FrameNet, with minimal manual intervention.

The development of FrameNet for languages other than English has shown that many frames, especially those for common human behaviours such as *drinking*, *eating* and *sleeping*, are relevant in multiple languages despite the presence of numerous language-specific valency patterns (Baker & Lorenzi 2020: 78). The different languages have adhered to the Berkeley FrameNet model to varying degrees: German, French, Swedish and Chinese FrameNet have deviated further from it by either adding many new frames or/and modifying existing ones, while

Svetla Koeva

Spanish, Japanese and Brazilian-Portuguese FrameNet have closely followed the original FrameNet and used FrameNet frames as templates (Baker & Lorenzi 2020: 78). The Multilingual FrameNet project (Baker et al. 2018) investigates the relations between frames in different languages and alignments between FrameNets. There are different approaches to calculate the similarity of frames to create cross-lingual alignments: alignment by translation of lexical units, alignment by frame names, alignment by similarity of frame elements, alignment by similarity of distribution of lexical units, etc. (Baker & Lorenzi 2020: 79–80).

In this study we outline the basic principles for the development of the Bulgarian FrameNet, relying on language-independent information from the semantic frames while taking into account the language-specific features of Bulgarian. We can characterise the model for the development of the Bulgarian FrameNet as a **semi-automatic expand model**, since the automatic mapping of lexical units from semantic frames is applied to the Bulgarian WordNet (Leseva & Stoyanova 2020), but both the automatic mapping of translation equivalents and the semi-automatic compilation of extended semantic and grammatical information for Bulgarian are evaluated manually.

The most important steps in the creation of the Bulgarian FrameNet can be summarised as follows:

- Semi-automatic identification of lexical units (verbs) belonging to the general lexicon of Bulgarian;
- Identification of semantic frames suitable for describing situations evoked by the selected Bulgarian lexical units;
- Import of relevant language-independent information (valid at least for English and Bulgarian) from FrameNet semantic frames into superframes and conceptual frames within the system for the development of the Bulgarian FrameNet, BulFrame;
- Semi-automatic population of conceptual frames with relevant Bulgarian lexical units and related lexical, grammatical and semantic information;
- FrameNet-based annotations of examples to illustrate the valency patterns of the selected lexical units;
- Manual evaluation of the information in conceptual frames based on the annotation and potential reconstruction of conceptual frames, leading to the development of multiple conceptual frames associated with a superframe.

4 Introducing superframes and conceptual Frames in Bulgarian FrameNet

The endeavours to create the Bulgarian FrameNet have a history of about 20 years, the origins of which go back to predecessors such as the Bulgarian Valence Dictionary and the Semantic-Syntactic Dictionary of Bulgarian (Koeva et al. 2003). Originally, the resources focussing on frame-like semantic and syntactic descriptions were exclusively centred on Bulgarian, without establishing correspondences with FrameNet.

In the following phase, appropriate semantic frames were selected manually and language-independent information was extracted from these frames. This information was then supplemented with Bulgarian lexical units evoking the corresponding frames, and relevant examples were annotated (Koeva 2010b). However, this endeavour was fraught with challenges, as there were no suitable means of maintaining correspondence with the semantic frames while providing options for reconstructing the semantic frame structures required for an adequate representation of some Bulgarian lexical units. Further challenges were to encode the translation equivalence between Bulgarian and English lexical units and to ensure the consistency of the FrameNet-like annotation with respect to the Bulgarian grammatical structure.

In its current stage, the Bulgarian FrameNet comprises two abstract semantic structures: a superframe and a conceptual frame, and it contains lexical units (accompanied by comprehensive lexical, semantic and grammatical information) that evoke conceptual frames, as well as valency patterns derived from authentic examples.

The main motivation for introducing superframes and conceptual frames is to facilitate the inclusion of language-specific information while ensuring consistency and alignment with the relevant semantic frames.

Superframes establish abstract mappings between semantic frames in FrameNet and their counterparts in Bulgarian, thus forming a bridge between semantic resources. Conceptual frames (linked with a superframe) encode relevant information for Bulgarian, which may overlap in whole or in part with that for English (Figure 2).

4.1 Superframes

Introducing a mediating abstract layer, such as the level of superframes, enables the alignment of the appropriate components in Bulgarian FrameNet with FrameNet semantic frames, while allowing some others to retain their specificity. Su-

Svetla Koeva

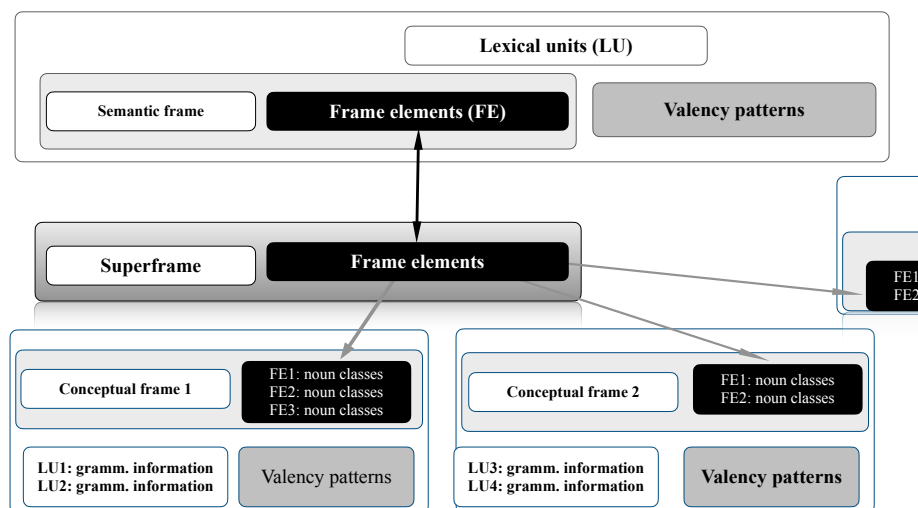


Figure 2: The correspondence between Berkeley semantic frames, superframes and conceptual frames for Bulgarian

perframes are constructed by removing all language-specific information for English, including lexical units that evoke the frames and their parts of speech, and retaining only non-specific information – semantic frames, their semantic types and definitions, frame-to-frame relations, frame elements, their semantic types and definitions, frame element relations, and administrative information such as frame and frame element names. Superframes therefore contain language-independent information that can apply to at least two languages, in this case English and Bulgarian.

In principle, a superframe may be constructed based on semantic frames for languages other than English for which a FrameNet is developed. This means that when a conceptual frame is developed based on Bulgarian data for which no appropriate superframe exists, a new superframe may be constructed, retaining only language-independent information in it.

This strategy aims to establish a seamless connection with FrameNet while enabling the identification and description of language-specific conceptualisations that are unique to Bulgarian and, if necessary, splitting a semantic frame into two or more conceptual frames, each characterised by different levels of reconstruction. An equivalence relation is established between the language-independent information in a semantic frame and the language-independent information in a superframe.

1 Universality of semantic frames vs. specificity of conceptual frames

Conceptual frames are used to introduce script-like descriptions that are relevant to Bulgarian and that may be wholly or partially analogous to the information for English or provide unique information relevant to the Bulgarian conceptual description. A superframe can therefore be linked to one or more conceptual frames. However, there can be at most one conceptual frame whose components are connected to the language-independent components of the superframe via an equivalence relation and to the semantic frame via the latter. The remaining conceptual frames are connected to the superframe by partial equivalence relations, that can be tracked to determine which components of the conceptual frames are equivalent to the corresponding components in the superframe and which are not. In some cases, only one conceptual frame for Bulgarian can be associated with a particular superframe.

The relations between superframes mirror the relations between semantic frames in FrameNet. At the current stage of development of the Bulgarian FrameNet, there is rarely a need to introduce a conceptual frame that is not linked to an existing superframe, and its mirroring as a superframe is not accompanied by the introduction of new frame-to-frame relations. Such changes to the FrameNet network, if they become necessary in the future, should be made with a high degree of consensus.

4.2 Conceptual frames

A *conceptual frame* can be defined (similarly to the semantic frame) as an abstract structure that describes a certain type of situation or event together with its actors and properties (Koeva 2020: 7). The conceptual frame is characterised by frame elements and relations between them and is complemented by a set of nouns that are compatible with the lexical units that evoke the frame.

A specific conceptual frame in the Bulgarian FrameNet is evoked by a group of lexical units, which (as of September 2024) are exclusively verbs.

Conceptual frames have a frame name, a definition, a semantic type, frame elements and relations between frames. *Frame elements* have a name, a definition, a semantic type, a core status and relations to each other: Core Sets, Excludes, Requires. This information is inherited from the semantic frames (via superframes) if they are already defined for English, and validated for Bulgarian by annotation.

Our motivation for employing superframes and conceptual frames is based on the following arguments:

Argument 1: Not all lexical units that evoke a given semantic frame exhibit the same semantic structure, which may lead to different syntactic behaviour.

Svetla Koeva

As part of the comprehensive FrameNet approach to conceptual description, we want to distinguish groups of lexical units with equivalent semantic and syntactic properties. Therefore, we adhere to the principle that the *semantic description of lexical units associated with a given conceptual frame is achieved by using the same number and type of core frame elements*. This approach does not change the structure of semantic frames, as many conceptual frames can be associated to a semantic frame (by a superframe). Therefore, there is often no one-to-one correspondence between a FrameNet semantic frame and a conceptual frame, as there are differences in conceptualisation between languages. The abstract superframe connects conceptual frames that express the same scene (one fully, the other partially) as the FrameNet semantic frame. The omission, rare addition and status change of core frame elements within the conceptual frames associated with a semantic frame is justified by the annotation of examples. As for the equivalent syntactic properties, they are only related to the equivalent semantic properties, i.e. to the number and type of core frame elements, but not to the possibilities of expressing one and the same frame element in different syntactic ways, e.g. by a prepositional phrase or a clause.

For example, the Bulgarian verbs *настанявам* (sit-IPFV, ‘am sitting’), *настаня* (sit-PFV, ‘sit’), with the definition ‘determine, show someone a place to sit or lie down and help him/her to take it’ (or comparison: the definition of *sit.v* in FrameNet is ‘cause to sit or be placed (somewhere)’), evokes the frame **Placing**, which encompasses core frame elements such as **AGENT**, **THEME** and **GOAL**. The **AGENT** is in a Core Set with the core frame element **CAUSE**, and each of them controls the **THEME** by placing it in a location, the **GOAL**. For the semantic description of the Bulgarian verbs *настанявам*, *настаня* in the conceptual frame **PLACING**, only the frame element **AGENT** is relevant, while the frame element **CAUSE** is omitted as semantically incompatible.

- (5) *Тозава* [*тя*]_{AGENT} **НАСТАНИ** [*майка си*]_{THEME}
 Then she set her mother
 [*в удобното кресло*]_{GOAL}.
 in the comfortable armchair.
 ‘Then she sat her mother in the comfortable armchair.’
- (6) **Тозава* [*вятърът*]_{CAUSE} **НАСТАНИ** [*майка ѝ*]_{THEME}
 Then the wind set her mother
 [*в удобното кресло*]_{GOAL}.
 in the comfortable armchair.
 ‘Then the wind sat her mother in the comfortable armchair.’

1 Universality of semantic frames vs. specificity of conceptual frames

Another example are the Bulgarian imperfect verbs from the frame *Self_Motion*, describing a scene in which a being moves in a certain way: *ходя* (walk-IPFV, ‘am waking’) ‘move by walking’; *разхождам се* (walk-IPFV, ‘am waking’) ‘walk somewhere outdoors’; *плувам* (swim-IPFV, ‘am swimming’) ‘for living organisms - move on the water surface or in the water by certain movements of the body’.

These lexical units imply very little in terms of source and direction, and there is no reason to include the frame elements *SOURCE*, *GOAL* and *DIRECTION* in their semantic description as core elements.² This is in contrast to the derivatively related perfective verbs, in whose semantic structure these frame elements can be core elements: *преплувам* (swim across-PFV, ‘swim across’) ‘for humans or animals – by swimming pass, cross a body of water or reach a certain place to which I am led’; *доплувам* (swim up-PFV, ‘swim up’) ‘swim to a certain place’.

- (7) [*Момчето*]_{SELF_MOVER} **ПЛУВА** [*в реката*]_{AREA}.
The boy swim-IPFV in the river.
‘The boy is swimming in the river.’

- (8) [*Момчето*]_{SELF_MOVER} **ДОПЛУВА** [*до брега*]_{GOAL}.
The boy swim_up-PFV to the shore.
‘The boy swam to the shore.’

Argument 2: In contrast to English and other languages, a large number of diatheses in Bulgarian are associated with a lexical and/or morphological change of the source verb and are part of the lexicon in dictionaries.

Our second reason relates to the inclusion of verbal diatheses in FrameNet. In FrameNet, there is no specific list of verbal diatheses that a semantic frame encompasses. However, certain details in the annotation instructions indicate that diatheses associated with a particular predicate are considered part of the frame to which the basic diathesis belongs. For example, no additional frame is formulated for the word *sell* to account for usages such as *Those boots sell well* that deprofile and make generic one or more of the prominent participants, in this case the *SELLER* (Ruppenhofer et al. 2016: 12). A similar approach also applies to passive sentences.

In cases where the semantic roles (the relation of an argument to the predicate or, in other words, of a frame element to the situation evoked by the lexical

²A deeper semantic analysis will show that verbs such as *walk*, *swim*, etc. are typical activity verbs, but when used with an explicitly expressed *GOAL*, they can be regarded as accomplishment verbs.

Svetla Koeva

unit) do not change, the diatheses can undoubtedly be interpreted within a single frame, even if some frame elements remain implicit. We refer to such diatheses as syntactic, e.g. the participial passive and syntactic reciprocals in Bulgarian. However, if the semantic role of at least one frame element changes as a result of the diathetic alternation (we call such diatheses lexical), there are reasons to reconstruct the semantic frame in a new conceptual frame.

In Bulgarian there are several lexical diatheses (*se* passive, impersonal participle passive, impersonal *se* passive, middle, anticausative, lexical reciprocal, optative, impersonal optative, “oblique” subject and property of the “oblique” subject (Koeva 2022: 153–155) and some others. The lexical diatheses can either be structure-preserving (i.e. the number of frame elements remains the same, but at least one of the frame elements is given a new semantic role) or structure-changing (whereby the number of frame elements changes).

- *Structure-preserving* diatheses in Bulgarian are optatives and lexical reciprocals. In optative diathesis, the semantic role of a core frame element, the source subject, is changed, which is accompanied by a change in its grammatical role. In lexical reciprocal diathesis, the semantic roles of two core frame elements (source subject and object) as well as the grammatical role and the syntactic category of the source object change.
- *Structure-changing* diatheses in Bulgarian: the impersonal passives (both impersonal participle and impersonal *se*-passive), the impersonal optatives, the middles, the anticausatives, the “oblique” subjects and the property of “oblique” subjects, show a reduction of semantic role as follows: both the source subject and the source object in the impersonal passives, the source subject in the middles and anticausatives and the source object in the impersonal optatives and “oblique” subjects. The reduction of the semantic role can be accompanied by a change in the semantic role, the grammatical role and the syntactic category of a remaining frame element.

The problem of the representation of lexical diatheses in the structure of the Bulgarian FrameNet is even more complicated because: a) in some of them the change of frame elements is connected with the appearance of a new frame element which is not present in the source diathesis; b) some of them have a regular character, i.e. if certain lexical, morphological and syntactic features are present in the source diathesis, the formation of a certain lexical diathesis follows. An example of a regularly occurring lexical diathesis in which a new frame element appears that is not part of the structure of the source diathesis is optative diathesis in Bulgarian (which expresses a wish or desire to carry out the state or process

1 Universality of semantic frames vs. specificity of conceptual frames

evoked by the source diathesis). The optative diathesis in Bulgarian is characterised by the following general features: The semantic role (frame element) of the canonical subject changes from AGENT (the name of the frame element can be different in FrameNet, e.g. READER) to EXPERIENCER, while that of the canonical object (if the verb is transitive) does not. At the morphological level, the optative verb is characterised by a reduction of the verb paradigm to the third person singular and plural and by the conjunction of the verb with the marker *ce* (self, ‘oneself’). The agentive source subject has the selectional restriction *person* (animate), the object – the selectional restriction *inanimate*, and the source verb should be in the imperfective aspect (primary or derived from a primary imperfective verb).

- (9) [Момчето]_{READER} **ЧЕТЕ** [книга]_{TEXT}.
 The boy read-IPFV book.
 ‘The boy is reading a book.’
- (10) **ЧЕТЕ** [му]_{EXPERIENCER} **СЕ** [книга]_{TEXT}.
 read-IPFV him self book.
 ‘He feels like reading a book.’

Although the meanings of the modified verbs in lexical diatheses differ and there are morphological (the lemma form), grammatical (the change of syntactic categories and grammatical roles in the realization of one or two frame elements) and semantic differences (the change in the number and type of frame elements or semantic roles), most lexical diatheses in Bulgarian are formed by regular language rules and can be predicted just like the syntactic ones. For those that involve the introduction of a new core frame element, such as the optative, there are only technical solutions to mark the option during annotation, i.e. by the name of the frame element: Agent-to-Experiencer and by the syntactic category and grammatical role of the word or phrase (otherwise, all conceptual frames that allow optative diathesis and other diatheses with similar regular alternations must be downgraded). However, numerous diatheses, e.g. lexical reciprocals and anticausatives, are not only formed regularly when certain conditions are met by the source diathesis, but they are also used very frequently and as such have become part of the lexicon in Bulgarian dictionaries. For such verbs there is a reason to present them in a separate conceptual frame in relation to their source verbs.

Lexical reciprocals are defined as “words with an inherent reciprocal meaning” (Nedjalkov 2007: 14). There are some unmarked reciprocal predicates in Bulgarian: *приличам на* (resemble, ‘look like’); reciprocal predicates with a reciprocal marker *се*: *състезавам се* (compete with, ‘compete with someone’), and

Svetla Koeva

reciprocal predicates that are a derived reciprocal diathesis: *прегръщам се* (hug with, ‘to hold someone at the same time as he/she holds me’) derived from the source diathesis *прегръщам* (hug, ‘to put one or two arms around someone or something and hug him/her to my chest’). The verbs *прегръщам* and *прегръщам се* are presented in two conceptual frames under the superframe Manipulation, which is connected to the semantic frame Manipulation. Since the meaning is reciprocal, but the reciprocity is not realised syntactically by a reciprocal pronoun and a plural subject as in *The boy and the girl hold each other*, the frame elements are both AGENT and ENTITY with a different focus on one of the two.

- (11) [*Момчето*]_{AGENT} ПРЕГРЪЩА [*момичето*]_{ENTITY}.
 The boy hold-IPFV the girl.
 ‘The boy hugs the girl.’
- (12) [*Момчето*]_{AGENT-ENTITY} СЕ ПРЕГРЪЩА [*с момичето*]_{ENTITY-AGENT}.
 The boy hold-IPFV with the girl.
 ‘The boy hugs with the girl.’

Another example of diathesis that is regularly listed in dictionaries is **anti-causative** diathesis, which is also known as inchoative, causative-inchoative or ergative diathesis (Levin 1993: 27). In this diathesis, the semantic role of the source subject is reduced and the semantic and grammatical role of the source object is changed. For example, the verbs *късам* (tear, ‘to cut something into pieces’) and *къса се* (is torn-3SG-3PL, ‘to fall into pieces’) are members of the causative-anticausative diathesis and are located in separate conceptual frames under the superframe Cutting, which is connected to the semantic frame Cutting.

- (13) [*Съседката*]_{AGT} КЪСА [*салфетката*]_{ITEM} [*на парчета*]_{PIE}.
 The neighbor cut-IPFV the napkin into pieces
 ‘The neighbor is tearing the napkin into pieces.’
- (14) [*Салфетката*]_{ITEM} СЕ КЪСА [*на парчета*]_{PIE}.
 The napkin itself is_cut-IPFV.3SG into pieces
 ‘The napkin is torn into pieces.’

In fact, the anticausatives in FrameNet are housed in separate semantic frames, which are linked to the corresponding causative frames via the frame-to-frame relations *Inchoative of* and *Causative of*. However, the Inchoative frames are not yet fully represented in the FrameNet. In order to maintain the FrameNet structure, no new semantic frame is introduced for the anticausative (inchoative) diathesis,

1 Universality of semantic frames vs. specificity of conceptual frames

but an inchoative conceptual frame is established which, together with the causative conceptual frame, is connected to the causative superframe and via this to the causative semantic frame. This approach maintains the structure of the semantic frames without altering it, yet effectively reflects the distinctions in both the semantic and syntactic structures of the verbs that evoke them in Bulgarian.

The so-called autocausative diathesis can be seen as a variant of the anticausative diathesis, with the difference that the verb is bound to an animate subject that causes its own activity. A large proportion of autocausative verbs have become part of the lexical structure of Bulgarian and are not perceived as a regular product of autocausative diathesis, as is actually the case.

- (15) [Бащата]_{AGENT} **ДЪРЖИ** [детето]_{ENTITY} за ръка.
 The father hold-IPFV the child for the hand
 ‘The father is holding the child’s hand.’
- (16) [Детето]_{PROTAGONIST} **СЕ ДЪРЖИ** за ръката на баща си
 The child itself is_hold-IPFV.3SG for his father’s hand.
 ‘The child is holding his father’s hand.’

In such examples, the autocausative marker *се* (se, ‘oneself’) in Bulgarian is not the short form of a reflexive pronoun, but a lexical marker that is part of the verb. For verbs such *държа се* (hold, ‘hold on to something with my hands. as a support to keep my balance so that I don’t topple over or fall’) separate conceptual frames are created (if a corresponding meaning is not lexicalised in English and there is no semantic frame), which are linked to the respective causative semantic frame by a superframe.

Argument 3: Conceptual frames differ from semantic frames in that the frame elements of the conceptual frame are associated with a number of lexical units through which they can potentially be realised.

Each core element of the conceptual frame is connected with a set of nouns that are compatible with the verbs that evoke the frame. The set can contain only one noun, several nouns or a large number of nouns linked by semantic relations at the lexical level (synonymy, antonymy) or by hierarchical conceptual relations (hyperonymy, hyponymy). For example, the verb *варя* (boil) with the definition ‘WN: cook food in very hot or boiling water’ from the frame Apply_heat is characterised by four frame elements: COOK, FOOD, CONTAINER and HEATING INSTRUMENT, and for each of these elements the synset (one or more) from the Bulgarian WordNet that dominates the nouns suitable for collocations is specified.

Svetla Koeva

- COOK: pro-drop, NP, subject, eng-30-00007846-n: *person*
- FOOD: optional, NP, object, eng-30-07555863-n: *food*; eng-30-07649854-n: *meat*; eng-30-07775375-n: *fish*; eng-30-07707451-n: *vegetable*
- CONTAINER: optional, PP, object (*в* 'in'), eng-30-03990474-n: *pot*
- HEATING INSTRUMENT: optional, PP, object (*от* 'of', *на* 'on'), eng-30-08581699-n: *hearth*; eng-30-03543254-n: *stove*; eng-30-03343560-n: *fire*

4.2.1 Levels of equivalence between conceptual frames and superframes

Three general cases can be outlined. A superframe may be suitable for adoption as a conceptual frame if it reflects the semantics of at least one Bulgarian lexical unit. Some modifications to the semantic structure of the superframe may be required to match the Bulgarian data; these changes relate to the number and type of frame elements. It may also be necessary to develop new conceptual frames to describe language-specific data. Thus, with regard to the use of the language-independent information from the semantic frames (which apply at least to English and Bulgarian), several cases may arise in relation to the superframes and conceptual frames: *equivalence*, *partial equivalence* and *no equivalence*.

An *equivalence* relation is observed when the abstract semantic representation of a superframe is copied into a conceptual frame to describe a scene evoked by a particular Bulgarian lexical unit (or units). For example, the semantic frame *Breaking_apart* with the definition 'A WHOLE breaks apart into PIECES, resulting in the loss of the WHOLE (and in most cases no piece that has a separate function)' applies to English lexical units: *break apart*, *break*, *crumble*, *fragment*, *shatter*, *snap*, *splinter*, as well as to their Bulgarian translation equivalents: *чупи се*, *счупи се*, *счупва се*, *разпада се*, *разпадне се*, *разтрошава се*, *разтроши се*, *трошава се*, *троши се*. In both languages, an equivalent situation can be represented, which is evoked by translation equivalents and expressed by the same number and type of frame elements: WHOLE and PARTS. In Bulgarian FrameNet, the respective conceptual frame is thus constructed through a superframe.

- (17) [Стъклената кана]_{WHOLE} СЕ СЧУПИ
 The glass jug break-PFV.3SG
 [на множество малки парченца]_{PIECES}.
 into many small pieces.
 'The glass jug broke into many small pieces.'

1 Universality of semantic frames vs. specificity of conceptual frames

The same procedure applies if a semantic frame is defined in FrameNet that is suitable for describing a Bulgarian verb or verbs, but its translation equivalent is not available in FrameNet. For example, the semantic frame *Breaking_apart* is also suitable for describing the verbs *пръсна се* ‘burst’, *пръсва се* ‘is bursting’ with the definition ‘WN: of a solid body or object – to split, break apart suddenly and with force into parts’, which are hyponyms of the verbs *счупи се*, *счупва се*. In this case, the lexical units are added to the Bulgarian conceptual frame.

- (18) [*Балонът*]_{WHOLE} *СЕ ПРЪСНА* *неочаквано* [*на парчета*]_{PIECES}.
 The bubble burst-PFV.3SG unexpectedly into pieces.
 ‘The bubble burst unexpectedly into pieces.’

A relation of *partial equivalence* is observed when a semantic frame defined in FrameNet is only partially suitable for describing a Bulgarian verb or verbs. In such cases, an equivalent superframe is defined for the language-independent information and the corresponding conceptual frame is reconstructed. This reconstruction can include the exclusion or addition (rarely) of a frame element. It is also possible to change the core status of a frame element. For example, the frame *Breaking_apart* is partially suitable for describing the verbs *пробива се* ‘is breaking through’, *пробие се* ‘break through’ with a definition ‘WN: of a solid body or object – suffer a breach of integrity by stabbing or piercing with a sharp object’, whereby only one frame element is realised: *WHOLE*. The new conceptual frame is connected to the same superframe as the conceptual frame evoked by verbs such as *счупи се* (breaking apart).

- (19) [*Гумата*]_{WHOLE} *СЕ ПРОБИ* *на две места*.
 The tyre puncture-IPFV.3SG in two places.
 ‘The tyre has punctured in two places.’

The abstract semantic structure *superframe* is introduced to maintain the relation to a semantic frame and to combine semantically related verbs that do not have exactly the same meaning and the same semantic, morphological and syntactic features. A superframe corresponds to a semantic frame from FrameNet and connects a group of conceptual frames that share all or part of the semantic information of the respective semantic frame in FrameNet. The conceptual frames associated with a particular superframe are identified by the name of the corresponding semantic frame and an additional unique name after one of the verbs that evoke the conceptual frame, e.g. *Breaking_apart_пробива_се*.

The relation *no equivalence* occurs when a semantic frame that is suitable for describing a Bulgarian verb is not defined in FrameNet and a new superframe and a conceptual frame must be defined. This can happen for two reasons:

Svetla Koeva

- The concept exists in English, but the corresponding semantic frame has not yet been created in FrameNet, for example, the verb *golf.v*.
- The concept is not conceptualised in English; for example, *захъркам* (am snoring-IPFV, ‘start snoring’); *захъркам* (snore-PFV, ‘start snoring’); *затъркам* (am looking for-IPFV, ‘start looking for’); *затърся* (look for-PFV, ‘start looking for’).

- (20) [Тоѝ]_{SOUND_SOURCE} *моментално* **ЗАХЪРКА** *в хотела.*
 He instantly startsnore-PST.3SG in the hotel.
 ‘He instantly started snoring in the hotel.’

The language-independent information in the semantic frames, which is inherited by the superframes, is located on the conceptual and semantic level. This includes the definitions of the frames, the relations between the frames, the number and types of the frame elements, their definitions, semantic types, core status and relations. Administrative information, such as the name of the frame and the names of the frame elements, is also inherited. In addition, conceptual frames may contain information that is language-specific or potentially language-universal but has not yet been integrated into FrameNet.

Conceptual frames also contain sets of nouns that are suitable for collocations with the target verb. In addition, some information, such as definitions of lexical units, semantic relations between the concepts they denote, semantic classes of lexical units, grammatical information such as verb aspects and administrative information such as identification numbers, is taken from WordNet.

5 Structure of the Bulgarian FrameNet

The structure in Bulgarian FrameNet associated with each lexical unit consists of the following sections: **Administrative**, **Lexical**, **Grammatical**, **Frame** and **Syntactic**.

Administrative information ensures the unambiguous interpretation of lexical units and frames. The WordNet ILI (Vossen 2004) serves several purposes: it acts as a unique identifier in both Bulgarian FrameNet and Bulgarian WordNet (BulNet), as it is the primary identifier for lexical units in Bulgarian FrameNet, it indicates the mapping to the corresponding synset (concept) in Princeton WordNet, it relates synonyms and labels the word senses associated with the lexical units. The names of the semantic frames are unique and are transferred to both

1 Universality of semantic frames vs. specificity of conceptual frames

the superframes and the conceptual frames, and the combination of frame name and frame element name is also unique.

Lexical units are provided with lexical and semantic information (lemma, part of speech, lexical type – indication whether it is a multiword expression or not, sense definition, semantic class of the lexical unit, semantic relations to other verbs, if any, and some stylistic or usage labels) in the **Lexical section** and with grammatical information (verb aspect, transitivity and the range of grammatical subjects) in the **Grammatical section**.

The **Frame section** contains information about the frame definition, the frame elements, their definitions and relations, their semantic types and the semantic classes of the noun synsets that are suitable for pairing with the lexical units that evoke the frame.

Grammatical categories and grammatical functions encode the syntactic realisation (valency pattern) in the **Syntactic section**, as supported by the annotation.

The sources of inheritance and uniqueness of information in the Bulgarian FrameNet are schematised in the Table 2.

5.1 Lexical section

Following FrameNet, a **lexical unit** is defined as a pairing of a word with a sense (Fillmore et al. 2003: 235) (expressed by lemma and definition). The FrameNet assumption that each sense of a polysemous word belongs to a different semantic frame is followed, and it also applies to homonyms. For example, the Bulgarian verb *деля* (divide) with the definition ‘WN: make a division or separation; FN: separate into parts or groups’ evokes the frame *Separating*, while the verb *деля* (share) with the definition ‘WN: use jointly or in common; FN: to use something jointly with another sentient being’ evokes the frame *Sharing*. There are five meanings in the Dictionary of Bulgarian Language³ that are close to the meaning of the verb *деля* (divide), and four meanings in the Bulgarian wordnet⁴; in both sources the granularity of meaning is thus high, suggesting that such words may belong to separate conceptual frames that are related to one or more superframes.

In contrast to English, homonymy between lemmas from different parts of speech occurs less frequently in Bulgarian, but it does exist, e.g. the verb *суша* (dry) ‘WN: to remove moisture and make dry’ and the noun *суша* (land) ‘WN: the solid part of the earth’s surface’.

³<https://ibl.bas.bg/rbe/lang/bg/деля>

⁴<https://dcl.bas.bg/bulnet/>

Svetla Koeva

Table 2: Source of information in Bulgarian FrameNet

	FrameNet (FN)	BulFrame (BF)
Admin. information		
Frame name	FN	FN or BF
FE name	FN	FN or BF
Verb ID	No	WordNet (WN)
Lexical information		
Lemma type	word, MWE	word, MWE
POS	V, N, Adj, Adv	V
Definition	FN	BF
Semantic class	No	WN
Stylistic note	No	BulNet (BWN)
Semantic type	FN	FN
Semantic relations	No	WN
Grammatical information		
Verb Aspect	No	BWN
Transitivity	No	BF
Personality	No	BF
Frame information		
Frame definition	FN	FN or BF
Frame-to-Frame relations	FN	BF
Frame elements	FN	FN or BF
FE Core status	FN	FN or BF
FE definition	FN	FN or BF
FE type	FN	FN or BF
FE relations	FN	FN or BF
V-to-N compatibility	No	BF
Syntactic information		
Grammatical category	FN	BF
Grammatical function	FN	BF
Implicitness	FN	BF

1 *Universality of semantic frames vs. specificity of conceptual frames*

In this phase of the development of the Bulgarian FrameNet, we focussed on 5,074 verbs, which were selected according to quantitative and qualitative criteria and a heuristic according to which the criteria are applied (Koeva & Doychev 2022: 207–208). The criteria include presence in the Age of Acquisition Test – the school level at which a word (the meaning of a word) must be learnt or mastered (Dale & O’Rourke 1981, Goodman et al. 2008, Morrison et al. 1997); presence in WordNet Base concepts (Vossen et al. 1998: 12–14), aiming for maximum overlap and compatibility between the wordnets of multiple languages; root distance (the number of nodes) of a synset to the root of the local tree (the hierarchical substructure in WordNet in which the corresponding synset is contained); relative frequency in the Bulgarian National Corpus (Koeva et al. 2012), in Bulgarian textbooks from first to fourth grade and in a Bulgarian dictionary for primary school children.

Table 3 shows the language-independent and language-specific information provided in the Lexical section of the Bulgarian FrameNet.

Table 3: Language-independent and language-specific information in Lexical section

Type of information	FrameNet	BulFrame
Semantic class	No	Language-independent
Stylistic and usage notes	No	Both specific or independent
Semantic relations	No	Language-independent

This information either shows the systematic semantic relations between the concepts denoted by the lexical units or serves as classifying meta-information indicating the affiliation of the lexical units to certain semantic, stylistic or usage classes.

5.1.1 **Lemma**

In Bulgarian grammar, it is assumed that the lemma is the highest unmarked word form, i.e. the form in which there are no morphematically expressed grammatical categories (with the exception of the verbs (Kutsarov 2007: 20), where the lemma is the form of the first person singular present tense), while the grammatically most bare form is the third person singular present tense).

The lemma for certain verb classes with a restricted paradigm, such as impersonal verbs, is the present tense in the third person singular. For other verb

Svetla Koeva

classes with a restricted paradigm, however, the first person singular is chosen as the lemma in dictionaries, even if it is not used with the specified word sense. For example, *тека* (flow-1SG.PRS, ‘I am flowing’), is used in dictionaries as a lemma instead of *тече* (flow-3SG.PRS, ‘it is flowing’). In the Bulgarian FrameNet, the lemma is defined as the first member of the word paradigm actively used in the language (Koeva 2008: 25), and for personal verbs the lemma is the first person singular, present tense; for impersonal and third-personal verbs the lemma is the third person singular, present tense; and for plural personal verbs the lemma is the first person plural, present tense (Koeva 2010a: 19).

5.1.2 Multiword expressions

There are many different classifications for multiword expressions Baldwin & Kim 2010, Constant et al. 2017, of which we have chosen the following classification for verbal multiword expressions in Bulgarian:

Semi-fixed: The number of constituents is fixed, but these constituents can undergo certain paradigmatic changes within certain grammatical categories; the order of constituents can change, although there is a preferred word order; and there is room for insertions from restricted groups of words, i.e., the multiword expression *гушна букета* (hug-PRS.1SG the-bouquet, ‘to kick the basket’).

In this context, we can distinguish different types of personal verbs whose lemma is formed with a “reflexive” by form particles. These include personal reflexiva tantum *se* verbs, such as *спирам се* (stop-1SG oneself, ‘am stopping’); personal reflexiva tantum *si* verbs, such as *спомням си* (remember-1SG oneself, ‘remember’); personal reciproca tantum *se* verbs, such as *състезавам се* (compete-1SG with someone, ‘compete with’); and personal reciproca tantum *si* verbs, such as *пиша си* (write-1SG with someone, ‘correspond with’).

This group also includes third-personal verbs that may or may not form their lemma with a “reflexive” particle. These are: third-personal accusativa tantum verbs, such as *мързи ме* (lazy-3SG me-ACC.1SG, ‘I am lazy’); third-personal dativa tantum verbs, such as *хрумне ми* (occur-3SG me-DAT.1SG, ‘it occurs to me’); and impersonal reflexiva dativa tantum verbs, such as *зади ми се* (sick-3SG me-ACC.1SG oneself, ‘I feel sick’). In these verbs, one of the frame elements must obligatorily be expressed by a personal pronominal clitic (accusative or dative), but it is conventionally regarded as part of the lemma, since the other forms express no meaning without it.

1 Universality of semantic frames vs. specificity of conceptual frames

Non-fixed: Its constituents can undergo morphological changes, undergo changes in word order and accommodate variable elements in their composition, e.g. *изнасям лекция* (give-3SG lecture, ‘to give a lecture’), *лекция ще изнасям утре* (give-FUT.3SG lecture tomorrow, ‘I will give a lecture tomorrow’).

Most constructions with support verbs belong to this group. In FrameNet, it is assumed that support (light) verbs are selected by a frame-bearing noun: *say a prayer* = *pray* vs. **give a prayer* and *give a speech* = *speak* vs. **say a speech* (Fillmore et al. 2003: 244). For example, *give a lecture* is part of the frame *Speak_on_Topic*, evoked by the lexical unit *lecture.n*. An important consequence of this analysis is the annotation in FrameNet of support verb subjects as frame elements relative to the noun.

The relevance of the verb to the support construction has been demonstrated as the support verbs can determine the semantic role that a particular constituent takes in a sentence (Fillmore et al. 2003: 244). For example, in the first sentence below, the grammatical subject is *PATIENT*, while in the second sentence the grammatical subject is *AGENT*.

- (21) *В миналото [той]_{PATIENT} Е ИМАЛ^{Supp}*
 In the past he has-PST.3SG
[катастрофа]_{UNDESIRABLE_EVENT} с камион.
 accident with a truck.
 ‘In the past he had an accident with a truck.’
- (22) *Днес [той]_{AGENT} НАПРАВИ^{Supp} катастрофа с камиона.*
 Today he make-PST.3SG accident with the truck.
 ‘Today he has made an accident with the truck.’

Both multiword expressions serve as synonyms for the Bulgarian verb *катастрофирам*. In the first example, it corresponds to the meaning ‘having an accident’, in the second example to the meaning of ‘making an accident’. The observations indicating that constructions with support verbs often have a single verb synonym, as well as the fact that support constructions can be considered sentence predicates, provide convincing evidence for the inclusion of support constructions as verbal multiword expressions in the Bulgarian FrameNet. The solution is that the multiword expressions with support verbs are considered evoking the respective noun frame and are annotated in the same way as in English. The difference is that the entire multiword expression is added as a lexical unit with its own meaning. Thus, the first multiword expression with a support

Svetla Koeva

verb becomes part of the conceptual frame *Catastrophe*, which is linked to the superframe and the semantic frame *Catastrophe*, while a new conceptual frame and superframe is created for the second, since there is currently no suitable semantic frame that can be copied.

5.1.3 Definition

The definition serves to explain the meaning of a verb in a way that clearly distinguishes it from other meanings of the same word. These definitions are taken from the Bulgarian WordNet (Koeva 2021: 57–58), in which verbs of the imperfective and perfective aspect are intentionally presented as synonyms, accompanied by a common definition to preserve the structure of the Princeton WordNet. An appropriate definition should reflect the category verb aspect and the morphological features of the verbs (the limited person paradigm as third person, impersonal and plural personal); therefore, some of the definitions in the Bulgarian FrameNet need to be modified. For example, the following two verbs: *излитам* (take off-IPFV.1SG, ‘am taking off’) and *излети* (take off-PFV.1SG, ‘take off’) are described with one definition in Bulgarian WordNet: ‘за летателни и космически апарати или под. – отделям се от земята и започвам да летя’ (of an aircraft and spacecraft or sub. – leave-IPFV.1SG the ground and begin to fly).

The definition was modified to describe the meaning of the imperfective and perfective verbs that are used in the third person only:

- (23) *излита* take off-IPFV.3SG *It is taking off.*
 ‘за летателни и космически апарати или под. – отделя се от земята и започва да лети’
 ‘of an aircraft, spacecraft, etc., leaves the ground and begins to fly’
- (24) *излети* take off-PFV.3SG *It takes off.*
 ‘за летателни и космически апарати или под. – да се отдели от земята и започне да лети’
 ‘of an aircraft, spacecraft, etc., to leave the ground and begin to fly’.

5.1.4 Semantic type, semantic class, stylistic and usage labels

Lexical units can be tagged with three different types from various sources: *semantic type* from FrameNet, *semantic class* from WordNet, and *stylistic and usage labels* from Bulgarian WordNet.

1 Universality of semantic frames vs. specificity of conceptual frames

FrameNet applies a number of **semantic types** to lexical units, with most types reserved for nouns and adjectives; however, all types available for verbs can be added to Bulgarian lexical units where appropriate. For example, *see.v* with the definition ‘COD: perceive with the eyes’, which evokes the frame *Perception_experience*, and *glance.v* with the definition ‘COD: take a brief or hurried look’, which evokes the frame *Perception_active*, have the semantic type *Visual_modality*, which can be transferred to the corresponding Bulgarian lexical units.

The synsets (or the individual words that make them up) in Princeton WordNet are organised into *semantic classes* (primitives) that represent basic concepts that serve as distinct roots of different hierarchies. Verbs within these hierarchies are grouped under common semantic classes (Fellbaum 1993: 47), such as *bodily care and functions*, *change*, *cognition*, *communication*, *competition*, *consumption*, *contact*, *creation*, *emotion*, *motion*, *perception*, *possession*, *social interaction*, *weather verbs*, *state*. Each verb in the Bulgarian FrameNet is assigned a semantic class from the WordNet. For example, the verb *искам* (wish, ‘order politely; express a wish’) has the semantic class *verb.emotion* and the verb *моля* (request, ‘ask a person to do something’) has the semantic class *verb.communication*.

Labels or notes (also from the Bulgarian WordNet) are assigned to the corresponding lexical units to indicate various features, such as non-standard usage, figurative meanings, obsolete terms, informal usage and more. This labelling scheme reflects various distinctions in language usage: **belonging to non-standard lexis**, which includes dialectal words, slang or vernacular terms; **words with unfavourable connotations**; **use in a specific functional style**, such as colloquial, poetic, literary or technical terms; **historical period of use**, which distinguishes between obsolete, historical and newly coined words; **the expressive properties of the word**, such as pejorative meanings, augmentative or diminutive forms; **frequency of use**, which indicates whether a word is rare; **the nuances in use**, such as figurative meanings (Koeva 2021: 55). Stylistic marking usually excludes words from the core vocabulary, and although the labelling comes from the Bulgarian WordNet, the number of marked verbs is not large.

While the stylistic and usage labels are language-specific, the semantic classes are largely language-independent. A single lexical unit can be given several labels if they characterise it in different ways.

5.1.5 Semantic relations

The **lexical units** that evoke a conceptual frame can be one or more linked to each other by lexical relations (synonymy, antonymy) and/or hierarchical semantic

Svetla Koeva

relations at the conceptual level (hypernymy, troponymy, entailment).

The *semantic relations* are inherited from the (Bulgarian) WordNet. Taxonomic relations for verbs are inverse and transitive (*has a troponym* and *has a hypernym*, or (*has a subevent* and *ї*). Non-hierarchical relations are: symmetrical, irreflexive and non-transitive (*antonymy*); symmetrical, irreflexive and Euclidean (*also see*, *verb group*).

The semantic relations in WordNet are defined between synsets, and in Bulgarian WordNet verbs with different lexico-grammatical aspect are unnaturally grouped in one synset. The following general rules have been implemented to split verbs with different aspect in the Bulgarian FrameNet, adopting the semantic relations for the concepts they express:

- *Troponymy*, *Hypernymy* and *Antonymy* connect either imperfective or perfective verbs.
- *Also see* and *Verb group* link verbs regardless of their aspect.

Semantic relations between lexical units are usually language-independent, with the exception of relations that reflect culturally and historically specific concepts. By integrating the semantic relations of WordNet into the Bulgarian FrameNet, the scope of the semantic description is extended, which enables the application of certain evaluation heuristics.

Figure 3 shows an example of the Lexical section for the verb *варя* (boil, ‘FN: cook by immersing in boiling water’), evoking the conceptual frame *Apply_heat_варя*.

5.2 Grammatical section

The conceptual frame integrates morphological details specific to each verb: the **verb aspect**; the **range of grammatical subjects** that a predicate can select, such as nouns, all personal pronouns, third person pronouns only, subject clauses or none; and the **range of grammatical objects** that a predicate can select, including nouns, personal accusative pronouns, complement clauses or none. The verb aspect is determined by the morphological structure of verbs, which includes lexico-grammatical prefixes and grammatical suffixes. The choice of grammatical subject is related to the person of the verb, while the presence of grammatical object depends to a certain extent on the transitivity of the verb.

These features (the ranges of grammatical subjects and objects) are closely related to the semantic and grammatical structures of a particular language. Even though some of the grammatical features are common to typologically related

1 Universality of semantic frames vs. specificity of conceptual frames

Word

варя

Characteristic	Value
POS	v
Group ID	eng-30-00328128-v
Sense	1
Semantic class	verb.change
Multiword Expression	< Empty >
Note	< Empty >
Definition	подлагам храна, суровина или под. на действието на много гореща или в
Has Hypernym	готвя (eng-30-00322847-v)
Has Hypernym	сготвям (eng-30-00322847-v)

Figure 3: Illustration of Lexical section in the BulFrame system.

languages, they cannot be considered language-independent. In the Slavic languages, including Bulgarian, the lexical manifestation of the verb aspect is evident, in contrast to English, where the continuous and perfect aspect are indicated by an auxiliary verb together with a present participle and a past participle respectively. The range of grammatical subjects and objects in FrameNet is illustrated by annotation and summaries of valency patterns, provided that the annotation covers all contextual realisations within a language. In the Bulgarian FrameNet, the verb aspect and the ranges of grammatical subject and object selection are explicitly indicated for each verbal lexical unit. In addition, these features are also confirmed by the annotation and the valency patterns.

Table 4 summarises the language-specific information in the Grammatical section of the conceptual frame.

Table 4: Language-specific information in Grammatical section

Type of information	FrameNet	BulFrame
Verb aspect	No	language-specific
Range of grammatical subjects	language-specific	language-specific
Range of grammatical objects	language-specific	language-specific

The combination of the values of the three categories (as well as some basic

Svetla Koeva

semantic information for the frame elements AGENT and EXPERIENCER such as human and animate, and their grammatical roles) determines the formation of verbal diatheses in Bulgarian.

5.2.1 Verb aspect

As in other Slavic languages, it is disputed whether perfective and imperfective verbs express different lexical meanings or whether they are different forms of the same word. Some scholars such as (Andreychin 1944: 137,193), (Kutsarov 2007: 546), (Nitsolova 2008: 247) favour the former view, while others such as (Maslov 1982: 189), (Stankov 1980: 8–9) argue for the latter. However, the prevailing evidence favours the first assumption and suggests that perfective and imperfective verbs in Bulgarian express different lexical and grammatical characteristics and thus represent different lexical units.

Bulgarian perfective and imperfective verbs typically exhibit overt morphological distinctions. Aspectual derivation in Bulgarian involves two main processes: perfectivisation and imperfectivisation. These derivations are formed by adding a prefix, a suffix or both to the verb root.

Certain verbs have a morphologically non-derived imperfective form (*imperfectiva tantum*) and can form perfective counterparts through prefixation. Conversely, perfective verbs can become imperfective through suffixation, a process known as secondary imperfectivisation. As a result, many Bulgarian verbs have aspect triplets.

- | | | | |
|------|----------------|---------------|-----------------|
| (25) | <i>водя</i> | <i>изведа</i> | <i>извеждам</i> |
| | take-IPFV | take_out-PER | take_out-IPFV |
| | ‘am taking to’ | ‘take out’ | ‘am taking out’ |

A smaller group of verbs are morphologically non-derived perfective verbs (*perfectiva tantum*), which produce imperfective counterparts through suffixation. Both non-derived perfective and derived imperfective verbs can form new prefixed verbs, whereby the lexico-grammatical aspect of the source verb is retained.

- | | | |
|------|-------------|---------------|
| (26) | <i>кажа</i> | <i>казвам</i> |
| | say-PFV | say-IPFV |
| | ‘say’ | ‘am saying’ |
-
- | | | |
|------|---------------|-----------------|
| (27) | <i>изкажа</i> | <i>изказвам</i> |
| | express-PFV | express-IPFV |
| | ‘express’ | ‘am expressing’ |

1 Universality of semantic frames vs. specificity of conceptual frames

In Bulgarian WordNet, each verb is assigned a label that indicates its lexico-grammatical aspect. For example: perfective verb – *запея* ‘start singing’; imperfective verb – *запявам* ‘starting singing’; simultaneous perfective and imperfective verb – *пенсионирам* ‘retire’; imperfective verb without perfective equivalent – *вали* ‘it is raining’; perfective verb without imperfective equivalent – *повярвам* ‘believe’. The labels indicating the aspect of the verbs in the Bulgarian WordNet are transferred to the Bulgarian FrameNet.

The difference in verb aspect is reflected both lexically, morphologically and syntactically, which determines the explicit definition of the aspect values in the Bulgarian FrameNet. Here are some examples of morphological differences (Nitsolova 2008: 256):

- The verbs with perfective aspect lack the so-called independent present tense (the present tense forms can only be used in subordinate clauses and their temporal meaning depends on the main verb), the present participles (both active and passive) and the negative imperative forms.
- The derivational potential also differs between perfective and imperfective verbs. Perfective verbs cannot form certain types of deverbal nouns or nouns that denote occupations.

The syntactic differences can be summarised as follows (Koeva 2022: 56):

- The perfective aspect has a direct influence on the syntactic realisation of verb complements. Direct objects of perfective verbs cannot remain implicit, and perfective verbs cannot serve as complements of phase predicates.
- The perfective aspect of the verbs also reflects the restrictions in the formation of diatheses, as the perfective verbs in Bulgarian do not form middles, optatives or impersonals.

5.2.2 Range of grammatical subjects

With regard to the categories person and number, each verb in the Bulgarian FrameNet is categorised in one of four ways: **personal** (for first, second and third person singular and plural), **impersonal** (restricted to third person singular), **third-personal** (for third person both singular and plural) and **plural-personal** (for first, second and third person plural) (Koeva 2010a: 33). This categorisation

Svetla Koeva

is compatible with transitivity: Personal verbs can be both transitive and intransitive, while impersonal and third-personal verbs are typically intransitive, with the exception of accusativa tantum verbs.

Furthermore, the range of grammatical subjects corresponds to the possibilities of occupying the subject position: personal verbs require a subject (noun, noun phrase, substantive, pronoun, or subject clause); impersonal verbs do not require an argument in the subject position; third-personal verbs admit only a third-person subject (noun, noun phrase, substantive, pronoun); and plural-personal verbs require a plural subject (noun, noun phrase, substantive, pronoun).

5.2.3 Verb transitivity

Transitivity (and intransitivity) result from a specific syntactic realisation of a core frame element: a noun phrase that fulfils the grammatical role of a direct object.

Most languages have a number of transitivity classes of verbs. A typical pattern (which occurs in English, Bulgarian and many other languages) is (Dixon & Aikhenvald 2000: 4):

- strictly intransitive verbs that only occur in an intransitive clause;
- strictly transitive verbs that only occur in a transitive clause (in such a case, the position of the direct object in the Bulgarian FrameNet is marked with the label obligatory);
- ambitransitive verbs (or labile) that occur in both an intransitive and a transitive clause (in such a case, the position of the direct object in the Bulgarian FrameNet is marked with the label optional).

Within the two main groups of verbs (transitive and intransitive), the subclasses in Bulgarian are further subdivided on the basis of their lexical properties. This subdivision takes into account whether the verbs form a multiword expression with a “reflexive” particle and/or whether they obligatorily combine with a pronominal “accusative” or “dative” clitic (Koeva 2010a: 34).

In diatheses, it is common that the number of core frame elements (subjects and objects) either decreases or that the semantic roles of these core frame elements change. Many diatheses mainly concern transitive verbs and their transformation into intransitives, in which the original object takes on the grammatical role of the subject.

1 *Universality of semantic frames vs. specificity of conceptual frames*

5.3 Frame section

The parts of the Frame section have different origins: some are inherited from FrameNet (in our case through a superframe), others are built according to the FrameNet structure, and another part is specific to the organisation of conceptual frames.

The correspondence with the language-independent semantic description in FrameNet (which is proven to be valid for at least two languages: English and Bulgarian) is documented in the conceptual frames by the superframes. There are three primary scenarios: two of them involve some form of equivalence (equivalence or partial equivalence) with a superframe, another does not. In the case of equivalence, the superframe is copied into a conceptual frame. In the case of partial equivalence, the language-independent information from FrameNet is reconstructed in a conceptual frame. If there is no equivalence, a new conceptual frame is developed and the language-independent information is integrated as a new superframe.

The FrameNet-related parts (inherited, (re)constructed or newly constructed) include frame elements together with their names, status (core, non-core and extrathematic), definitions, semantic types and relations.

The reconstruction of an inherited language-independent part may involve the reduction of a (core) frame element, the occasional insertion of a new frame element, or the change of status of a core or a peripheral frame element. Some relations between frame elements can either be reduced or redefined.

In addition to specifying semantic types for frame elements, a set of nouns is defined by using one or more noun synsets in WordNet that dominate the hypernym trees in which the nouns compatible with the target lexical units are represented.

The Table 5 contains both language-independent and language-specific information in the Frame section, which is explained in more detail below.

Table 5: Language-independent (LI) and language-specific information in Frame section

Type of information	FrameNet	BulFrame
Frame elements	LI or specific	LI or specific
FE relations	LI or specific	LI or specific
FE noun fillers	No	LI or specific

Svetla Koeva

5.3.1 Frame elements

Conceptual frames consist of frame elements equipped with name, definition, semantic type, core status, internal relations between the frame elements and information about the nouns with which these frame elements can be expressed. The names, definitions and semantic types of frame elements are adopted by FrameNet without additional specifications if they are suitable for Bulgarian. Only in rare cases, when a new frame (or a new frame element) is proposed, are they created from scratch.

So far, only the core frame elements relevant to Bulgarian have been included in the conceptual frames. In most cases where there is overlap between the core frame elements in the conceptual descriptions for both languages (English and Bulgarian), this correlation also extends to the peripheral frame elements.

5.3.1.1 Core vs. peripheral frame elements

When constructing the Bulgarian FrameNet, it is necessary to assess whether (core) frame elements are applicable to the semantic structure of the Bulgarian lexical units. The task is to determine whether a particular frame element is relevant, and if so, whether it is a core or a peripheral, and in some cases there is a core status shift of frame elements in Bulgarian descriptions. A (core) frame element may be omitted or modified to better fit the description of the Bulgarian lexical units.

The distinction between core and peripheral frame elements in FrameNet is based on the following properties of core frame elements: overtly specified, unambiguous interpretation when omitted, and without formal marking.

If we compare Bulgarian with English, the overt specification of frame elements cannot be regarded as a formal feature of subjects. In contrast to English, subject omission in Bulgarian can occur in combination with all verbal categories, not only with the imperative. Bulgarian is a null subject language and allows subject omission due to its rich verb inflectional morphology, which indicates person, number and in some verb categories also the gender of the omitted subject. The use of null subjects in Bulgarian in the first and second person is not grammatically and contextually restricted, while the choice between explicit or implicit subject in the third person may depend on the context of the discourse.

The conditions for direct null objects differ from the conditions for null subjects. Similar to some other languages, direct null objects in Bulgarian can only be observed with transitive imperfective verbs. Since the imperfective verb may not imply the result of an activity, the object can be omitted. A null object is permissible if it is understood by the lexical meaning of the verb, mentioned earlier

1 Universality of semantic frames vs. specificity of conceptual frames

in the discourse or implied by the context. In cases where support verbs are used, the object cannot be omitted even with transitive imperfective verbs:

- (28) *Влязох в стаята, където Иван четеше **книга**, преди да заспи.*
 ‘I entered the room where Ivan was reading a book before falling asleep.’
- (29) *Влязох в стаята, където Иван четеше, преди да заспи.*
 ‘I entered the room where Ivan was reading before falling asleep.’
- (30) *Влязох в стаята, където Иван вземаше **хапчета за сън**, преди да заспи.*
 ‘I entered the room where Ivan was taking sleeping pills before falling asleep.’
- (31) * *Влязох в стаята, където Иван вземаше, преди да заспи.*
 ‘I entered the room where Ivan was taking before falling asleep.’

Indirect objects and nominal or prepositional adverbials that express a core element of the frame can also remain implicit if they can be derived from the wider context.

It has been pointed out that arguments at the semantic level can be obligatory or non-obligatory, and truly optional semantic arguments are distinguished from obligatory semantic arguments. For example, in *He kicked the pumpkin (down the stairs)* the phrase *down the stairs* is a realisation of an optional semantic argument, while in *He threw the pumpkin (down the stairs)* it is an obligatory semantic argument (Culicover & Jackendoff 2005: 174-176). Both verbs allow the optional expression of the *path of motion*. If such a *path* is not expressed with *kick*, the direct object does not have to experience any movement. With *throwing*, on the other hand, something is set in motion. Even if the *path* expression is omitted, it is therefore semantically implicit. In the verb *throw*, the *path of motion* is therefore an obligatory semantic argument that remains implicit at the syntactic level.

In Bulgarian, there is no general formal marking for grammatical roles such as subject and object. Only the personal pronominal clitics have forms for the nominative, accusative and dative. For indirect objects and prepositional adverbials, the range of permissible prepositions can be specified. Given the many options for omitting sentence constituents and the limited use of formal case markers, the unambiguous semantic interpretation of omitted core constituents therefore remains the primary formal feature for their differentiation in Bulgarian.

The challenges in determining the core status of frame elements for Bulgarian verbs can be twofold: firstly, the categorisation of indirect objects that do not

Svetla Koeva

serve as core frame elements, and secondly, the categorisation of prepositional and noun adverbial phrases that serve as core frame elements.

It has been pointed out that adverbial phrases as core frame elements typically accompany verbs describing situations in which temporal or spatial elements represent frame-internal information – “information that fills in details of the internal structure of an event ... as opposite to the information about the setting of incidental attending circumstances of that event, the frame-external information” (Fillmore 1994: 159).

- (32) *The show started **at five o'clock**.*
- (33) *The performance lasted **five hours**.*
- (34) *John lives **in Sofia**.*
- (35) *The vase stays **at the table**.*

5.3.1.2 Basic instances of restructuring frame elements

Since the core frame elements were taken from the FrameNet, the main task in building the Bulgarian FrameNet is to assess whether these frame elements are applicable to the semantic structure of the Bulgarian lexical units. In rare cases, a peripheral frame element may be elevated to the status of a core frame element in the Bulgarian FrameNet. Conversely, a core frame element can be omitted or modified in order to adapt it to the semantic description of the situation evoked by the Bulgarian lexical units.

An example of the lack of a core frame element in the Bulgarian FrameNet is the semantic frame Awareness, which contains core frame elements such as COGNIZER, CONTENT and TOPIC, the latter two forming a Core Set. The Bulgarian verb *вярвам* (believe, ‘COD: feel sure of the truth of’) evokes the conceptual frame Awareness, in which the frame element TOPIC is not involved. In other words, when describing the verb *вярвам*, only the frame element CONTENT is relevant within the conceptual frame Awareness, while the frame element TOPIC is omitted.

- (36) *Вчера [той]_{EXPERIENCER} **ВЯРВАШЕ** [в своята гениалност]_{CONTENT}.*
 Yesterday he believe-PST in his genius.
 ‘Yesterday he believed in his genius.’
- (37) *[ДНІ]_{EXPERIENCER} **ВЯРВАМ** [че това е верният път]_{CONTENT}.*
 I-dropped believe-PRS that this is the right way.
 ‘I believe that this is the right way.’

1 Universality of semantic frames vs. specificity of conceptual frames

Another example in the Bulgarian FrameNet that illustrates the transformation of a peripheral frame element into a core frame element is the frame element GOODS in the frame Robbery, which is defined as follows: A PERPETRATOR wrongs a VICTIM by taking something (GOODS) away from him. Bulgarian verbs such as *обирам* (rob-IPFV, ‘am robbing’), *обера* (rob-PFV, ‘rob’) evoke the frame Robbery, and in their syntactic realisation the element GOODS receives a unique interpretation and is therefore considered a core frame element.

- (38) *Тозава [скитникът]_{PERPETRATOR} ОБРА [дома на капитана]_{SOURCE}.*
 Then the tramp rob-PST the captain’s house.
 ‘Then the tramp robbed the captain’s house.’
- (39) *Тозава [скитникът]_{PERPETRATOR} ОБРА [накитите]_{GOODS}*
 Then the tramp rob-PST the ornaments
[от къщата]_{SOURCE}.
 from the house.
 ‘Then the tramp stole the ornaments from the house.’

The Bulgarian verbs resulting from lexical reciprocal diathesis, in which the reciprocal meaning is expressed at the lexical level in both singular and plural forms, are an example of a case in which the semantic roles of the core elements of the frame involved in the source are shifted within the derived diathesis. For example, the verbs *пиша* (write-PER, ‘provide information to someone through letters’) and *пиша си* (write-PER oneself-REFL, ‘exchange information with someone through letters’) are both parts of the frame Text_creation, which is defined as follows: An AUTHOR creates a TEXT, either written, such as a letter, or spoken, such as a speech, that contains meaningful linguistic tokens and may have a particular ADDRESSEE in mind. The source diathesis evokes a situation with the core frame elements AUTHOR and ADDRESSEE, while the derived diathesis with the reciprocal meaning evokes a situation with the core frame elements AUTHOR1_ADDRESSEE2 and AUTHOR2_ADDRESSEE1.

- (40) *[Поетът]_{AUTHOR} ПИШЕ [на своята любима]_{ADDRESSEE}.*
 The poet write-PRS to his beloved.
 ‘The poet writes to his beloved.’
- (41) *[Поетът]_{AUTHOR1_ADDRESSEE2} СИ ПИШЕ*
 The poet oneself_write-PRS
[със своята любима]_{AUTHOR2_ADDRESSEE1}.
 with his beloved.
 ‘The poet corresponds with his beloved.’

Svetla Koeva

5.3.1.3 Frame element relations

The relations between frame elements are inherited from FrameNet. If a semantic description of a scene is suitable for English and Bulgarian, then the generalisations for the relations between the frame elements should also apply to both languages (as a general rule). For example, in the frame *Manipulation*, which describes the manipulation of an ENTITY by an AGENT, the *Core Set* is defined between the frame elements AGENT and BODYPART_OF_AGENT.

- (42) *Бабата започна да* *го* **МАСАЖИРА**
 The_old lady start.3SG-PST to he-DAT
 [с *ръце*]_{BODYPART_OF_AGENT}
 massage with hand-PL.
 ‘The old lady started massaging him with her hands.’
- (43) *Бавно* [*ръцете*]_{BODYPART_OF_AGENT} **МАСАЖИРАХА** *врата и гърба*
 Slowly the_hand.PL massage-3PL neck and back
 му.
 his.
 ‘Slowly, the hands massaged his neck and back.’

The definition of Core Sets (each member of the set is sufficient to fulfil the semantic valency of the predicator) allows the inclusion of diatheses *Oblique* subject, as members of one and the same conceptual frame. The alternations that fall into this group are: Natural force subject, Instrument subject, Locatum subject, Raw Material Subject (Levin 1993: 79–83). The “oblique subject” diatheses are realised when the semantic role of the subject does not change and the semantic role of the prepositional object is reduced, but the source noun of the prepositional phrase is realised as a derived subject (Koeva 2022: 148).

In many cases, it is sufficient to define a Core Set because it means the realisation of one frame element instead of another, but it also defines their co-occurrence, which is not possible with the relation *Exclusion*.

The frame elements CONTENT (The CONTENT is the entity that evokes a cognitive reaction for the EXPERIENCER) and EXPERIENCER (The EXPERIENCER is newly aware of the CONTENT) from the frame *Enter_awareness* are in the relation *Requires*. The frame element CONTENT cannot appear without the frame element EXPERIENCER. The example is for with the lexical unit *хрумне ми* (it occurs to me, ‘FN: suddenly become known (to someone)’).

1 Universality of semantic frames vs. specificity of conceptual frames

- (44) *ХРУМНА* [*му*]_{EXP} [*една идея*]_{CONT}.
 Occur I-DAT an idea.
 ‘An idea occurred to me.’
- (45) *ХРУМНА* [*му*]_{EXP} [*да излезе от стаята*]_{CONT}.
 Occur he-DAT to leave from the_room.
 ‘It occurred to him to leave the room.’

An example of the additional coding of relations between frame elements are the relations *Requires* between the frame elements WHOLE and PARTS as well as between PART and PART2, while PART1 *Excludes* PARTS and PART2 *Excludes* PARTS in the frame *Becoming_separated* with the definition: A WHOLE separates into PARTS, or one part of a whole, called PART1, becomes separate from the remaining portion, PART2.

5.3.2 Noun frame elements fillers

FrameNet enables the characterisation of ‘role fillers’ by semantic types of frame elements, which should be largely constant across all uses. However, not all frame elements are provided with a semantic type or the semantic types are too general. For this reason, we have decided to specify as far as possible the set of nouns that are suitable to represent a particular frame element in a sentence and combine with the target verb.

Several attempts have already been made in this direction. Compared to FrameNet, another lexical-semantic resource based on frame semantics, VerbAtlas, uses fewer semantic roles (frame elements) (25) and many more semantic types (selectional preferences) (116), expressed in terms of WordNet synsets (Di Fabio et al. 2019: 627). Selectional preferences were manually chosen from a set of 116 macro-concepts defined by WordNet synsets whose hyponyms are considered likely candidates for the corresponding argument slot (Di Fabio et al. 2019: 627), a strategy similar to the previous one based instead on algorithms (Agirre & Martínez 2001). The comparison of the interpretation of the verb *hear* in FrameNet and VerbAtlas shows that the semantic frame in FrameNet *Perception_experience* is more general (it includes any kind of perception), while the frame in VerbAtlas includes verbs with a different but narrower meaning, limited to auditory perception. In FrameNet there is a frame element that refers to the body part through which the perception occurs, and in VerbAtlas there are two semantic roles: *Stimulus* and *Source*, with an equal selectional preference: *entity* (the difference remains unclear). Both resources contain frame elements (semantic roles)

Svetla Koeva

for the *Perceiver* and the perceived auditory *Phenomenon*. It is very difficult to define sets of verb-noun combinations, regardless of whether automatic or corpus observation methods are used, due to the objective difficulties that figurative but acceptable usage may always entail.

In another approach, for a dictionary containing FrameNet-based data for English, Brazilian Portuguese and Spanish, domain-specific ontologies are used to impose semantic constraints on the frame elements (Hauck et al. 2015). In the context of the Brazilian FrameNet, each core frame element undergoes an analysis based on the aspect of the scene it represents, resulting in the mapping of one or more frames to the frame element (Torrent et al. 2022). Only frames that stand for events, states, attributes and relations are eligible for frame element to frame relations. The information available from the definition or semantic type is used to determine the type of concept it refers to (e.g. person, place, event) and to identify the highest level frame that represents it. In this way, the FrameNet is enriched with additional semantic information by linking the conceptual structures that make it up. The approach seems to be similar to the introduction of morphosemantic relations between verb and noun synsets in WordNet; however, this extension is not applicable as nouns have not yet been included in the Bulgarian FrameNet.

The accepted approach is to select the topmost synset (or conjunction of topmost synsets) from the Bulgarian WordNet that dominates all corresponding noun synsets for a given frame element. The nouns in WordNet are divided into 25 semantic classes (Miller 1990: 16), which, being general, can be subdivided into subclasses. For example, within the semantic class food, the subclass beverage can be introduced for nouns associated with verbs such as *stir*, *sip*, *drink*, *lap*, and so on. Such a representation aims to specify the organisation of concepts in an ontological structure that allows inheritance between semantic classes in the hierarchy and ensures a more precise specification of compatibility between verbs and nouns. One way to extend the WordNet semantic class repository is to map the WordNet synsets to an existing hierarchy of semantic types, e.g. the Corpus Pattern Analysis (CPA) types (Hanks 2004). The extension of the WordNet semantic classes with the CPA semantic types is done manually by matching the CPA semantic types with the WordNet synsets and selecting the most suitable ones (Koeva et al. 2018). Initially, less than 100 semantic classes were used; however, as the number of lexical units and conceptual frames increased, so did the number of semantic classes defined, reaching 377 in September 2024.

The aim is to provide representative information about the collocations between verbs and nouns by extracting corpus evidence. Both extremely vague and overly specific descriptions are avoided. For example, the verbs *bomb.v* and

1 Universality of semantic frames vs. specificity of conceptual frames

attack.v are part of the frame **Attack** (An **ASSAILANT** physically attacks a **VICTIM**), and the frame element **VICTIM** with the verb *bomb.v* is characterised by the hyponyms region: *eng-30-08630985-n* and *building: eng-30-02913152-n*. Combinations such as *bombed the island, village, meadow, bank, building* etc. are therefore supplied, while collocations such as *bombed the ship, boat and army*, which have a lower frequency, are not supplied. For the verb *attack.v*, the specification for: *settlement: eng-30-08672562-n, building: eng-30-02913152-n, military unit: eng-30-08198398-n, defensive structure: eng-30-03171356-n*. In FrameNet, the frame element **VICTIM** is represented by the semantic type *Sentient*, which is not particularly representative for the verbs *bomb.v* and *attack.v*. In the VerbAtlas, the verbs *bomb* and *attack* are part of the frame **bomb-attack**, and the corresponding semantic role is **patient** with the selective preference **object**, which is an over-generalised representation. Instead, our approach is to identify patterns of nouns that are combined with specific verbs from semantically relevant frames.

Another example are the verbs used in Bulgarian to convey information to the recipient (indirect object), such as *казвам, съобщавам* ('say, 'convey an information, an opinion, an instruction, etc. '); *разказвам* (tell; narrate, 'communicate a story, a fairy tale, etc. '); *обяснявам, разяснявам* (explain, 'clarify (something) to someone by describing it in more detail'), etc. All these verbs are assigned to the semantic class *verb.communication* in WordNet and belong to the semantic frame **Statement** with the definition: This frame contains verbs (and nouns) that communicate the act of a **SPEAKER** to address a **MESSAGE** to some **ADDRESSEE** using **LANGUAGE**. Similarly, a **TOPIC** can be specified instead of a **MESSAGE**. To illustrate the proposed approach, the corresponding noun fillers for the frame elements **SPEAKER** and **MESSAGE** are shown in the frame **Statement** evoked by the verb *обяснявам* (explain).

The fillers for the frame element **SPEAKER** are either nouns that are assigned to the semantic class *noun.person* in WordNet, or non-sentient nouns whose meaning can express unions of persons, such as *party, ministry, organisation, company*, etc., which denote organisations that are responsible for specific functions, policies or services. In this context, such nouns embody abstract concepts of administrative powers, policy formulation, regulatory oversight, etc., which do not refer to physical, tangible entities, but to the collective functions and responsibilities related to human activities. Therefore, in this case, appropriate nouns in the WordNet synsets should be classified based on their ability to express the collective functions of people.

Regarding the noun fillers of the frame element **MESSAGE**, it should be noted that these are nouns that are classified as *noun.communication* or *noun.cognition* in WordNet. However, these nouns differ in how they express communication

Svetla Koeva

and cognition. Therefore, it is important to develop a technique to eliminate those nouns that cannot be collocated with the verb *обяснявам* (explain) as direct objects. The synset communication with the definition ‘something that is communicated by or to or between people or groups’ is at the top of the hierarchy of nouns labelled with the semantic class *noun.communication*. However, this meaning is too abstract to serve as a filler for the frame element MESSAGE. The same applies to the top synset in the hierarchy, which is labelled with the semantic class *noun.cognition*: cognition; knowledge; noesis with the definition ‘the psychological result of perception, learning and reasoning’. Although the hyponyms of these synsets are suitable in most cases, some inappropriate synsets appear in the respective subtrees: receipt – ‘an acknowledgment (usually tangible) that a payment has been made’; mail – ‘the bags of letters and packages that are transported by the postal service’; and publication – ‘the communication of something to the public; making information generally known’, among others. All non-combinable nouns are concrete and contained in synsets labelled *noun.communication*. One possible strategy to restrict them is to add another level of noun classification: abstract and concrete nouns.

Two approaches must therefore be combined to correctly determine the appropriate noun classes to fill the positions of the frame elements of a given verb:

- Selection of the most appropriate synset or the most suitable combination of synsets in the hypernym hierarchy that semantically dominate the corresponding nouns.
- Introduction of additional elementary semantic types and classification of nouns based on these types so that correct generalisations can be made. These types include *collective*, *abstract*, *concrete* and *agentive*.

Figure 4 shows an example of how the information for conceptual frames is stored in Bulgarian FrameNet.

5.4 Syntactic section (Valency patterns)

The Syntactic section comprises two components: the specification of *grammatical categories* and the assignment of *grammatical functions* to syntactic phrases, which are realisations of frame elements. In addition, it contains specifications for implicit realisations and sets of suitable prepositions for prepositional phrases as well as sets for suitable clause linking components.

1 Universality of semantic frames vs. specificity of conceptual frames

Apply_heat	
Attribute	Value
Frame (saps)	<input checked="" type="checkbox"/> Included + -
Lemma	saps (eng-30-00328128-v)
Origin	FrameNet
Definition	A Cook applies heat to Food, where the Temperature_setting of the heat may be specified.
Semantic type	
Frame element	<input checked="" type="checkbox"/> Included + -
Element name	Cook
FE-to-FE relation	
Status	core
Semantic type	Sentient
Element definition	The Cook applies heat to the Food.

Figure 4: An extract from the conceptual frame Applied_heat_варя.

Valence patterns associated with frame elements of specific conceptual frames are language-specific, although a comparison with English may show that some patterns are applicable to both languages.

Table 6 represents language-specific information in the Syntactic section.

Table 6: Type of information in the Syntactic section.

Type of information	Semantic frame	Conceptual frame
Implicitness	language-specific	language-specific
Grammatical category	language-specific	language-specific
Grammatical function	language-specific	language-specific

The main source for the annotation and extraction of the valency patterns is the Bulgarian sense-annotated corpus (Koeva 2012). It was selected due to its relatively extensive coverage, which includes 17,041 verbs with 6,612 unique senses. The high variance in the corpus is achieved by selecting about 100 word excerpts from each of the 500 texts of the Bulgarian Brown corpus. These samples are selected based on the density of the most frequently occurring open-class lemmas, applying heuristics to ensure a fair representation across different parts of speech and a broader coverage of lemmas. In addition, the corpus is characterised by the fact that all the words it contains have been manually annotated for their senses from the Bulgarian WordNet. So far, the full annotation includes verbs related to communication, contact, change, perception, emotion and movement, with a total of about 2,500 unique meanings (Koeva 2024). The annotation led to

Svetla Koeva

the grouping of lexical units (as of September 2024) into 551 conceptual frames, which are linked to 247 semantic frames.

For the realisations of each frame element, a **syntactic description** is provided on the basis of the annotation. For example, the verb *режа* (cut, ‘WN: separate with or as if with an instrument’) from the frame *Cutting* is characterised by the following syntactic description of the frame elements:

- AGENT: pro-drop, NP, Subject
- OBJECT: optional, NP, Object
- PARTS: optional, PP, Object (*на* ‘of’)
- INSTRUMENT: optional, PP, Object (*с* ‘with’)

Individual words are annotated for the part of speech and dependency relations corresponding to grammatical functions (roles) are automatically assigned by the Natural Language Processing Pipeline (Koeva et al. 2020), so that the grammatical categories and universal dependencies are automatically annotated before the correspondences between grammatical categories, grammatical roles and frame elements are manually annotated.

5.4.1 Grammatical categories

The **grammatical categories** in the syntactic realisation of the core frame elements in Bulgarian are: NP (noun phrase), PP (prepositional phrase), S (clause), AP (adjective phrase), ACCCL (obligatory accusative clitic) and DATCL (obligatory dative clitic). Noun phrase stands for nouns, substantives (lexical and syntactical), pronouns or noun phrases, also coordinative. A prepositional phrase can have a simple structure (a preposition and a noun, a pronoun or a substantive) or a complex structure (a preposition and a noun phrase). The adjective phrase, the noun phrase and the prepositional phrase can be a realisation of a small clause. Obligatory accusative or dative clitics are frame element fillers of multiword verbs, although as such they are part of the complex lemma. Since in such cases the accusative or dative clitics are instances of core frame elements, we repeat the clitics both in the lemma and in the valency patterns.

There can be more than one valency pattern for a single verb with a unique meaning. To account for this, two strategies are used:

- A given frame element can have more than one type of realisation, e.g. a noun phrase or a clause;

1 Universality of semantic frames vs. specificity of conceptual frames

- In some conceptual frames, special relations may exist between the competing realisations of frame elements

For example, the lexical unit *убеждавам* (persuade) is part of the frame *Sua-sion* which has a frame element *TOPIC* with a definition: The general item or items that are the focus of the *CONTENT* of the *SPEAKER*'s message. The frame element *TOPIC* can have different syntactic realisations: a prepositional phrase or a clause.

- (46) *Днес [все повече предприемачи]_{SPKR} УБЕЖДАВАТ [клиентите]_{ADDR} [PP в предимствата на търговската марка]_{TOP}.*
'Nowadays, more and more entrepreneurs are convincing customers in the advantages of the trademark.'
- (47) *Днес [все повече предприемачи]_{SPKR} УБЕЖДАВАТ [клиентите]_{ADDR}, [S че търговската марка има предимства]_{TOP}.*
'Nowadays, more and more entrepreneurs are convincing customers that the trademark has advantages.'

In the annotation, various prepositions and clause-linking phrases can be collected for a specific frame element, which is expressed as a prepositional phrase. The types of clauses vary according to the type of linking, whether by interrogative pronouns or conjunctions; the complex linking phrases or complementisers are annotated and summarised accordingly. The combinations of prepositions and clause types form clusters within different verb classes, each associated with conceptual frames belonging to different semantic domains.

We have implemented an approach that complements the annotation to encode possible valency patterns, even if they do not yet occur in the corpus examples. A different approach would significantly reduce their number. For example, the lexical unit *разказвам* ('tell, narrate; communicate a story, tale, etc.') from the conceptual frame *Statement* is linked to 22 valency patterns that only take the core frame elements into account. These patterns include options for expressing the frame element *SPEAKER* with a Definite Null Instantiation or a noun (a noun, a noun phrase or a pronoun, with the exception of possessive and reflexive pronouns); the frame element *MESSAGE* with Indefinite Null Instantiation, a noun (a noun, a noun phrase, an accusative personal pronoun clitic) or a complement clause; the frame element *ADDRESSEE* with an Indefinite Null Instantiation or a prepositional phrase introduced by the preposition *на* ('to') or replaced by a dative personal pronoun clitic, and the frame element *TOPIC* with Indefinite Null Instantiation or a prepositional phrase introduced by the preposition *за* ('about').

Svetla Koeva

5.4.1.1 Null instantiations of frame elements

The phrases expressing the frame elements are: a) (in rare cases in Bulgarian) obligatorily explicit or b) non-obligatory explicit, which means that the potential for syntactic realisation of the phrase is present, but its explicitness is not obligatory because it is understood from the context in a broader sense (verb morphology, preceding text, extra-linguistic information, etc.). A special case is a pronominal drop in the subject position.

In FrameNet, the annotation for zero instantiation corresponds to the alternatives for the omission of core frame elements in Bulgarian FrameNet: pro-drop subjects and implicit (optional) objects. If the missing part is understood in the linguistic or conversational context, this is called Definite Null Instantiation in FrameNet and corresponds to pro-drop subjects. Indefinite Null Instantiation is indicated by the absence of objects in verbs such as *eat*, *read*, *drink*, etc., i.e. in cases where transitive verbs are used intransitively. As in FrameNet, construct-related omitted constituents can also occur here, such as the omitted subject of imperative sentences, the omitted agent of passive sentences, etc.

5.4.2 Grammatical functions

The **grammatical functions** used in Bulgarian FrameNet are subject, direct object, indirect object, adverbial, subject clause, object clause, adverbial clause and small clause. Compared to FrameNet, they are more detailed, especially with regard to the clauses.

The frame elements associated with the subject of Bulgarian verbs can be characterised as follows: They can have an explicitly or implicitly expressed subject with a complete paradigm; alternatively, they can have a subject explicitly or implicitly expressed in the third person; or they can have no subject at all. The frame elements corresponding to the objects of Bulgarian verbs can be classified as follows: with a single NP object; with an NP object and a complement clause; with an NP object and PP objects, regardless of their number; with an NP object, PP objects (regardless of their number) and a complement clause; with PP objects, regardless of their number; with PP objects (regardless of their number) and a complement clause; with a complement clause; and without objects. In addition, an AdvP predicate modifier, SC (small clause) NP, SC (small clause) PP and SC (small clause) AP can occur as realisations of some frame elements.

While the grammatical functions are largely predictable due to the nature of the frame elements and the grammatical categories of their syntactic realisations, the encoding of this information mainly illustrates the manifestation of different types of diatheses registered for a lexical unit.

1 *Universality of semantic frames vs. specificity of conceptual frames*

Our approach aims to maximise coverage by explicitly encoding the grammatical functions for potential syntactic realisations of a given frame element alongside the annotation process.

Figure 5 provides an overview of the encoding of syntactic information (valency patterns) in the Bulgarian FrameNet.

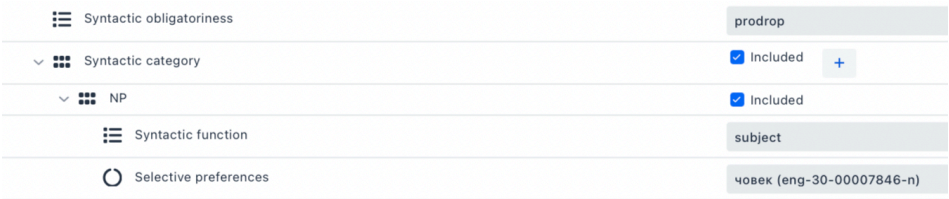


Figure 5: Illustration for a part of the Syntactic section in the BulFrame system.

6 Conclusions

The study presents the semantic frames of FrameNet on the basis of Charles J. Fillmore’s frame semantics theory and outlines the main assumptions underlying the development of the Bulgarian FrameNet.

A core premise of our research is that the semantic frames developed for English as part of the Berkeley FrameNet project can also be applied to the semantic analysis of Bulgarian. In other words, we claim that semantic frames represent a language-independent repository of semantic descriptions in which the language-independent components of the semantic frames are integrated into abstract structures known as superframes.

Most semantic frames can actually be used for analysing Bulgarian, especially through the use of superframes and conceptual frames that facilitate the alignment of semantic frames with Bulgarian data, including idiosyncratic differences and lexical diatheses. The main components of this description, lexical units and frame elements, are enriched in the following way: lexical units are enriched with grammatical, lexical and semantic information, such as semantic classes and semantic relations, while frame elements are associated with noun phrases that represent how they can be realised and combined with the target lexical units.

The Bulgarian FrameNet will be integrated into the network of equivalent or identical linguistic descriptions for other languages by using the language-universal information in the FrameNet to describe Bulgarian lexical units.

Svetla Koeva

The unified representation of semantic and syntactic information is important for the system analysis, description and use of Bulgarian, both for natural language processing and as a source of linguistic knowledge. The syntactic patterns resulting from the realisation of frame elements can be studied in order to draw classificatory and typological conclusions about Bulgarian verb classes. This also applies to the patterns of prepositions and clause types and, to the highest degree, to the semantic classes of nouns that coexist with the target lexical units.

Abbreviations

A or a	Adjective	LI	Language-independent
ACCCL	Obligatory accusative clitic	LS	Language-specific
Addr	ADDRESSEE	MWE	Multiword expression
AdvP	Adverbial phrase	N or n	Noun
AP	Adjectival phrase	NP	Noun phrase
BF	BulFrame	PFV	Perfective
CPA	Corpus Pattern Analysis	PP	Prepositional phrase
DATCL	Obligatory dative clitic	S	Subordinate clause
FE	Frame element	SC	Small clause
FN	FrameNet	V or v	Verb
ILI	Inter-lingual index	WN	WordNet
IPFV	Imperfective		

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

References

- Agirre, Eneko & David Martínez. 2001. Learning class-to-class selectional preferences. In *Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning (ConLL)*. <https://aclanthology.org/W01-0703>.
- Andreychin, Lyubomir. 1944. *Basic Bulgarian grammar (in Bulgarian)*. Sofia: Hemus Publishing House.

1 Universality of semantic frames vs. specificity of conceptual frames

- Baker, Collin F., Michael Ellsworth, Miriam R. Petruck & Swabha Swayamdipta. 2018. Frame semantics across languages: Towards a multilingual FrameNet. In *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts*, 9–12. Santa Fe, New Mexico, USA: Association for Computational Linguistics. <https://aclanthology.org/C18-3003>.
- Baker, Collin F. & Arthur Lorenzi. 2020. Exploring crosslinguistic frame alignment. In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, 77–84. Marseille, France: European Language Resources Association. <https://aclanthology.org/2020.framenet-1.11>.
- Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkha & Fred J. Damerau (eds.), *Handbook of Natural Language Processing, Second edition*, 267–292. Boca Raton: CRC Press.
- Boas, Hans C. (ed.). 2009a. *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin, New York: De Gruyter Mouton.
- Boas, Hans C. 2009b. Semantic frames as interlingual representations for multilingual lexical databases. In Hans C. Boas (ed.), *Multilingual FrameNets in computational lexicography: Methods and applications*, 59–100. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110212976.1.59.
- Borin, Lars, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj & Dimitrios Kokkinakis. 2010. The past meets the present in Swedish FrameNet++. In *Proceedings of the 14th EURALEX international congress*, 269–281.
- Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado & Manfred Pinkal. 2009. Multilingual FrameNets in computational lexicography: Methods and applications. In Hans C. Boas (ed.), *Framenet for the semantic analysis of German: Annotation and representation and automation* (Trends in Linguistics 200), 209–244. Berlin: Mouton de Gruyter.
- Candito, Marie, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël de Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathieu, Philippe Muller, Benoît Sagot & Laure Vieu. 2014. Developing a French FrameNet: Methodology and first results. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1372–1379. Reykjavik, Iceland: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/496_Paper.pdf.
- Constant, Mathieu, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner & Amalia Todirascu. 2017. Survey: Multiword expression processing: A survey. *Computational Linguistics* 43(4). 837–892. DOI: 10.1162/COLI_a_00302.
- Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Svetla Koeva

- Dale, Edgar & Joseph O'Rourke. 1981. *The Living Word Vocabulary: A National Vocabulary Inventory*. Chicago: World Book-Childcraft International.
- Di Fabio, Andrea, Simone Conia & 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 (EMNLP-IJCNLP)*, 627–637. Hong Kong, China: Association for Computational Linguistics. DOI: 10.18653/v1/D19-1058. <https://aclanthology.org/D19-1058>.
- Dixon, Robert & Alexandra Aikhenvald. 2000. Introduction. In Robert Dixon & Alexandra Aikhenvald (eds.), *Changing valency: Case studies in transitivity*, 1–29. Cambridge: Cambridge University Press.
- Fellbaum, Christiane. 1993. English verbs as a semantic net. *International Journal of Lexicography* 3(4). 40–51.
- Fillmore, Charles J. 1968. The case for case. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–88. New York: Holt, Rinehart & Winston.
- Fillmore, Charles J. 1976a. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* 280(1). 20–32.
- Fillmore, Charles J. 1976b. Frame semantics and the nature of language. In *Annals of the New York academy of sciences: Conference on the origin and development of language and speech*, vol. 280, 20–32.
- Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Publishing.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di semantica* 6. 222–254.
- Fillmore, Charles J. 1994. Under the circumstances (Place, Time, Manner, etc.) In Kevin E. Moore, David A. Peterson & Comfort Wentum (eds.), *Proceedings of the 20th annual meeting of the Berkeley Linguistics Society: General session dedicated to the contributions of Charles J. Fillmore*, 158–172. Berkeley, CA: Berkeley Linguistics Society.
- Fillmore, Charles J. 2006. Frame semantics. In Dirk Geeraerts (ed.), *Cognitive linguistics: Basic readings*, 373–400. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110199901.373.
- Fillmore, Charles J. 2007. Valency issues in FrameNet. In Thomas Herbst & Katrin Götz-Votteler (eds.), *Valency: Theoretical, descriptive and cognitive issues*, 129–160. Berlin: Mouton de Gruyter.

1 *Universality of semantic frames vs. specificity of conceptual frames*

- Fillmore, Charles J. 2008. The Merging of “Frames”. In R. Rossini Favretti (ed.), *Frames, Corpora, and Knowledge Representation*, 1–12. Bononia University Press. http://www.icsi.berkeley.edu/pubs/ai/ICSI_mergingframes08.pdf.
- Fillmore, Charles J. & Collin F. Baker. 2009. A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 313–340. Oxford: Oxford University Press.
- Fillmore, Charles J., Christopher R. Johnson & Miriam R. Petruck. 2003. Background to framenet. *International Journal of Lexicography* 16(3). 235–250. DOI: 10.1093/ijl/16.3.235.
- Goodman, Judith C., Philip S. Dale & Ping Li. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language* 35(3). 515–531. DOI: 10.1017/S0305000907008641.
- Hanks, Patrick. 2004. Corpus pattern analysis. In *Proceedings of the 11th EURALEX international congress*, 87–97. Lorient, France.
- Hauck, Paulo, Fernanda Campos, Tiago T. Torrent, Ely Matos & José Maria David. 2015. FrameNet-based automatic suggestion of translation equivalents. In *ONTOBRAS 2015 Brazilian Ontology Research Seminar*.
- Hayoun, Avi & Michael Elhadad. 2016. The Hebrew FrameNet project. In *Proceedings of the International Conference on Language Resources and Evaluation*. <https://api.semanticscholar.org/CorpusID:31290>.
- Koeva, Svetla. 2008. Bulgarian FrameNet. Semantic-Syntactic dictionary of Bulgarian – Conceptual Model (in Bulgarian). In *Bulgarian FrameNet: Semantic-Syntactic Dictionary of Bulgarian*, 5–51. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2010a. *Bulgarian FrameNet (in Bulgarian)*. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2010b. Lexicon and grammar in Bulgarian FrameNet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, Malta: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/705_Paper.pdf.
- Koeva, Svetla. 2012. Bulgarian sense-annotated corpus: Between the tradition and novelty. *Études Cognitives/Studia Kognitywne* 12. 181–198.
- Koeva, Svetla. 2020. Semantic Relations and Conceptual Frames. In Svetla Koeva (ed.), *Towards a Semantic Network Enriched with a Variety of Semantic Relations*, 7–20. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2021. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.

Svetla Koeva

- Koeva, Svetla. 2022. The system of diatheses in Bulgarian and stative predicates. In Svetla Koeva, Elena Ivanova, Yovka Tisheva & Anton Zimmerling (eds.), *Ontology of stative situations – linguistic modeling. A contrastive Bulgarian-Russian study*, 117–160. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Koeva, Svetla (ed.). 2024. *Semantichni studii. Sintaktichno i semantichno opisanie na balgarski glagoli*. Sofia: Publishing House of the Bulgarian Academy of Sciences. <https://dcl.bas.bg/wp-content/uploads/2024/06/SemanticStudies2024.pdf>.
- Koeva, Svetla, Tsvetana Dimitrova, Valentina Stefanova & Dimitar Hristov. 2018. Mapping WordNet concepts with CPA ontology. In *Proceedings of the 9th Global Wordnet Conference*, 69–76. Nanyang Technological University (NTU), Singapore: Global Wordnet Association. <https://aclanthology.org/2018.gwc-1.9>.
- Koeva, Svetla & Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, 203–213. Sofia, Bulgaria: Department of Computational Linguistics, IBL – BAS. <https://aclanthology.org/2022.clib-1.23>.
- Koeva, Svetla, Emil Doychev & Georgi Cholakov. 2003. Formal representation of syntactic environment and syntactic features of words (SYNText – a WEB base system designed for frame lexicon). In *Balkan language resources and tools*, 41–47.
- Koeva, Svetla, Nikola Obreshkov & Martin Yalamov. 2020. Natural language processing pipeline to annotate Bulgarian legislative documents. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 6988–6994. Marseille, France: European Language Resources Association. <https://aclanthology.org/2020.lrec-1.863>.
- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Rositsa Dekova, Tsvetana Dimitrova & Ekaterina Tarpomanova. 2012. The Bulgarian national corpus: Theory and practice in corpus design. *Journal of Language Modelling* 0(1). 65–110. DOI: 10.15398/jlm.v0i1.33.
- Kutsarov, Ivan. 2007. *Theoretical grammar of Bulgarian. Morphology (in Bulgarian)*. Plovdiv: Publishing house of Plovdiv university.
- Lenci, Alessandro, Martina Johnson & Gabriella Lapesa. 2010. Building an Italian FrameNet through semi-automatic corpus analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC*

1 Universality of semantic frames vs. specificity of conceptual frames

- 2010). Valletta, Malta: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/313_Paper.pdf.
- Leseva, Svetlozara & Ivelina Stoyanova. 2020. Beyond lexical and semantic resources: Linking WordNet with FrameNet and enhancing synsets with conceptual frames. In Svetla Koeva (ed.), *Towards a semantic network enriched with a variety of semantic relations*, 21–48. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Lönneker-Rodman, Birte & Collin F. Baker. 2009. The FrameNet model and its applications. *Nat. Lang. Eng.* 15(3). 415–453. DOI: [10.1017/S1351324909005117](https://doi.org/10.1017/S1351324909005117).
- Maslov, Yuriy. 1982. *Gramatika na balgarskiya ezik*. Sofia: Nauka i izkustvo.
- Materna, Jiří & Karel Pala. 2010. Using ontologies for semi-automatic linking VerbaLex with FrameNet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, Malta: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/867_Paper.pdf.
- Miller, George A. 1990. Nouns in WordNet: A lexical inheritance system. *International Journal of Lexicography* 3(4). 245–264.
- Morrison, Catriona, Tameron Chappell & Andrew Ellis. 1997. Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology Section A* 50(3). 528–559. DOI: [10.1080/027249897392017](https://doi.org/10.1080/027249897392017).
- Nedjalkov, Vladimir. 2007. Overview of the research: Definitions of terms, framework, and related issues. In Vladimir Nedjalkov (ed.), *Reciprocal constructions* (Typological Studies in Language 71), 3–114. Amsterdam: John Benjamins.
- Nitsolova, Ruselina. 2008. *Bulgarian grammar. Morphology (In bulgarian)*. Sofia: Publishing house of Sofia University.
- Ohara, Kyoko. 2012. Semantic annotations in Japanese FrameNet: Comparing frames in Japanese and English. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, 1559–1562. Istanbul, Turkey: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/1053_Paper.pdf.
- Ohara, Kyoko, Seiko Yamaguchi Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito & Shun Ishizaki. 2004. The Japanese FrameNet project: An introduction. In <https://api.semanticscholar.org/CorpusID:15338426>.
- Osswald, Rainer & Robert D. Van Valin. 2014. FrameNet, frame structure, and the syntax-semantics interface. In Thomas Gamerschlag et al. (eds.), *Frames and concept types* (Studies in Linguistics and Philosophy 94), 125–156. Springer.

Svetla Koeva

- Pedersen, Bolette, Sanni Nimb, Anders Søgaard, Mareike Hartmann & Sussi Olsen. 2018. A Danish FrameNet lexicon and an annotated corpus used for training and evaluating a semantic frame classifier. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA). <https://aclanthology.org/L18-1378>.
- Peron-Corrêa, Simone, Alexandre Diniz, Meire Lara, Ely Matos & Tiago Torrent. 2016. FrameNet-based automatic suggestion of translation equivalents. In João Silva, Ricardo Ribeiro, Paulo Quaresma, André Adami & António Branco (eds.), *Computational processing of the Portuguese language*, 347–352. Cham: Springer International Publishing.
- Rüggeberg, Carlos Subirats & Miriam R. Petruck. 2003. Surprise: Spanish FrameNet! In *International congress of linguists workshop on frame semantics*, 6. <https://api.semanticscholar.org/CorpusID:868992>.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin. F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.
- Sikos, Jennifer & Sebastian Pado. 2018. FrameNet’s using relation as a source of concept-based paraphrases. *Constructions and Frames* 10(1). 38–60. DOI: 10.1075/cf.00010.sik.
- Stankov, Valentin. 1980. *Verb aspect in the Bulgarian language (in Bulgarian)*. Sofia: Nauka i izkustvo.
- Subirats, Carlos. 2009. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In Hans C. Boas (ed.), *Multilingual FrameNets in computational lexicography: Methods and applications*, 135–162. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110212976.2.135.
- Torrent, Tiago T., Collin F. Baker, Oliver Czulo, Kyoko Ohara & Miriam R. Petruck (eds.). 2020. *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*. Marseille, France: European Language Resources Association. <https://aclanthology.org/2020.framenet-1.0>.
- Torrent, Tiago T., Maria Margarida M. Salomao, Fernanda C. A. Campos, Regina M. M. Braga, Ely E. S. Matos, Maucha A. Gamonal, Julia A. Gonçalves, Bruno C. P. Souza, Daniela S. Gomes & Simone R. Peron. 2014. Copa 2014 FrameNet brasil: A frame-based trilingual electronic dictionary for the football world cup. In Lamia Tounsi & Rafal Rak (eds.), *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, 10–14. Dublin, Ireland: Dublin City University & Association for Computational Linguistics. <https://aclanthology.org/C14-2003>.

1 Universality of semantic frames vs. specificity of conceptual frames

- Torrent, Tiago T., Maria Margarida M. Salomão, Ely Edison da Silva Matos, Maucha Andrade Gamonal, Júlia Gonçalves, Bruno Pereira de Souza, Daniela Simões Gomes & Simone Rodrigues Peron-Corrêa. 2014. Multilingual lexicographic annotation for domain-specific electronic dictionaries: The Copa 2014 FrameNet Brasil project. *Constructions and Frames* 6. 73–91.
- Torrent, Tiago Timponi, Ely Edison Da Silva Matos, Frederico Belcavello, Marcelo Viridiano, Maucha Andrade Gamonal, Alexandre Diniz da Costa & Mateus Coutinho Marim. 2022. Representing context in FrameNet: A multi-dimensional, multimodal approach. *Frontiers in Psychology* 12. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.838441/full>.
- Vossen, Piek. 1996. Right or wrong: Combining lexical resources in the EuroWordNet project. In Jarborg M. Gellerstam, S. Malmgren, K. Noren, L. Rogstrom & C. R. Papmehl (eds.), *Proceedings of the Euralex Workshop*, 715–728. Göteborg, Sweden.
- Vossen, Piek. 2004. EuroWordnet: A multilingual database of autonomous and language-specific wordnets connected via an inter-lingual index. *International Journal of Lexicography* 17(2). 161–173. DOI: 10.1093/ijl/17.2.161.
- Vossen, Piek, Laura Bloksma & Rodriquez Horacio. 1998. *The EuroWordNet Base Concepts and top ontology*. <https://research.vu.nl/ws/portalfiles/portal/74100710/D017>.
- You, Liping & Kaiying Liu. 2005. Building Chinese FrameNet database. In *2005 international conference on natural language processing and knowledge engineering*, 301–306. <https://api.semanticscholar.org/CorpusID:10064243>.

Chapter 2

Language-independent and language-specific properties of semantic description: A case study on verbs of communication

Svetlozara Leseva 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

Ivelina Stoyanova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

The study focuses on the properties of verb conceptual description in view of their linguistic universality and transferability of conceptual information across languages. Further, we present the semantic class of verbs of communication, the hierarchical organisation of frames and the corresponding frame elements. We consider the most prominent FrameNet frames evoking verbs of communication of higher frequency and make observations on the syntactic realisation of the frame elements in different valence patterns both in English and Bulgarian.

1 Introduction

In this paper we focus on combining the semantic description available for verbs in different lexical semantic resources (WordNet and FrameNet) which contain complementary semantic information (Baker & Fellbaum 2009). We discuss the



aspects of universality of conceptual knowledge that enable the transfer of semantic and to a lesser extent syntactic information across resources and languages. Further, we analyse the language-specific properties of the semantic and syntactic description. We illustrate our findings in a case study on verbs of communication in English and Bulgarian.

For the purposes of the study we employ: (a) the Princeton WordNet, PWN (Fellbaum 1999b), and the Bulgarian WordNet (Koeva 2021a), and (b) FrameNet (Fillmore et al. 1998, Ruppenhofer et al. 2016). In particular, we centre on the information included in them and how they complement each other in terms of coverage of lexical units and with respect to the semantic and syntactic features of the description. While we use resources for English and Bulgarian, the principles adopted in this work are applicable to other languages for which a wordnet aligned with PWN is developed.

There are several other resources relevant to our study, which provide background on the approaches for the extensive language-specific description of verb classes in comparison to developing cross-lingual and multilingual lexical and semantic resources. Further, their brief review sheds light on the possibilities for combining resources aiming at comprehensive description of lexical units. The functionalities and the additional information contained in these resources are summed up below.

VerbNet (Kipper-Schuler 2005, Kipper et al. 2008) provides substantial coverage of the English verb inventory and defines syntactic-semantic relations in an explicit way by means of predicate-argument structures (defined as configurations of thematic roles) with one-to-one linking to the syntactic category (type of phrase) and grammatical function (subject, object, etc.) of each argument expressed in terms of a relatively small number of syntactic frames. Selectional restrictions are defined for the thematic roles assigned to a verb's arguments; these restrictions capture the semantic/ontological class of the nouns that express the arguments. However, although the verb classes describe the syntactic behaviour of verbs, many of the traditional thematic roles employed may be too general for an exhaustive semantic description and appropriate handling of the syntax-semantics interface and is often biased towards English. Moreover, the overlap (and hence, the coverage of the existing mappings) between the WordNet synsets and the VerbNet classes is not large enough to provide sufficient data for analysis.

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) integrating lexicographic and encyclopaedic knowledge from WordNet

2 *Language-independent and language-specific properties*

and Wikipedia (Navigli & Ponzetto 2010). Each verb synset in VerbAtlas is assigned a frame corresponding to its prototypical predicate-argument structure. Obligatory components are described using 26 semantic roles and the semantic restrictions governing their compatibility (116 types). A semantic annotation API with the frames described in it is also provided with the resource.

Predicate Matrix (de Lacalle et al. 2014) is a lexical resource resulting from the integration of several sources of predicate information: FrameNet, VerbNet, PropBank and WordNet, that have been previously aligned in Semlink¹ (Palmer 2009). Predicate Matrix is compiled using advanced graph-based algorithms to extend the mapping coverage between resources. Additionally, by exploiting SemLink, new role mappings are inferred among the different predicate schemas.

The alignments of WordNet and FrameNet have been proposed for different languages, such as Danish (Pedersen et al. 2018), Dutch (Horak et al. 2008), Korean (Gilardi & Baker 2018), among others. One of the challenges in mapping resources developed according to different methodologies is the coverage of the alignment between the units represented in them. For instance, the alignment between lexical units evoking particular frames in FrameNet and corresponding verbs in synonym sets in WordNet, achieves coverage of 30.5% (Leseva & Stoyanova 2019). New methods have been proposed to increase the coverage by discovering suitable literals based on semantic relations with literals already described in semantic frames (Burchardt et al. 2005).

Combining the semantic description of verbs from different resources has been proposed by (Urešová et al. 2020a,b). The result is a multilingual dictionary encoding a comprehensive description of the semantic classes of verbs and the semantic roles and syntactic properties of their arguments². The project is also aimed at creating an ontology of events, processes and states, and for this purpose each dictionary entry is linked to its correspondences in FrameNet, WordNet, VerbNet, Ontonotes and PropBank, as well as the Valence Dictionary of Czech Verbs (Lopatková et al. 2016), which represents the predicate-argument structure of each verb, its semantic class and the syntactic transformations (diatheses) in which it participates.

Our work on aligning conceptual resources relies on the notion of universality. We side with the idea that the conceptual description provided in the FrameNet frames is to a considerable degree language-independent, which makes it possible for it to be transferred and/or adapted from one language to another. We map the conceptual knowledge contained in FrameNet onto the Princeton WordNet

¹<https://verbs.colorado.edu/semliink/>

²<https://ufal.mff.cuni.cz/synsemclass>

and through it, onto the Bulgarian WordNet. We then go on to examine the feasibility of transferring the valence information described for English to Bulgarian and the language-specific features that need to be addressed. The combination of semantic and syntactic information is seen as a possible way of transferring knowledge across languages (especially underresourced ones) by relying on the universality of the semantic description.

The study is organised as follows. Section 2 briefly presents the lexical-semantic resources involved in the work as well as the corpora used for extracting examples illustrating the various syntactic realisations in English and Bulgarian. Section 3 discusses the mapping of FrameNet frames onto WordNet synsets with a view to the universality of conceptual description as the main principle for cross-lingual transfer. Section 4 offers a detailed analysis of the semantic class of verbs of communication in terms of their conceptual structure and frame elements involved in the relevant frames. This analysis serves as a case study illustrating the main principles of universality as well as the language-specific features of syntactic realisation of frames. Section 5 draws conclusions based on the analysis and gives some directions for future work.

2 Resources

Below we describe in brief the lexical semantic resources used in the study, focusing on their strengths and the ways of overcoming their possible limitations through integrating the information contained in them. We also describe the corpora serving as a source of examples, the methodology for extracting suitable examples and the annotation of frame elements and their syntactic realisation.

2.1 Lexical-semantic resources

2.1.1 WordNet

WordNet³ (Miller 1995, Fellbaum 1999b) 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. WordNet provides extensive lexical coverage; the verbs presented in it are organised in 14,103 synsets (including verb synsets specific for Bulgarian). In this work, we use both the Princeton WordNet and the Bulgarian WordNet

³<https://wordnet.princeton.edu/>

2 Language-independent and language-specific properties

(Koeva 2021a), which are aligned at the synset level by means of unique synset identifiers.

WordNet provides the most coarsely-grained semantic division in terms of a set of language-independent semantic primitives assigned to all the nouns and verbs in the resource. The verbs fall into 15 groups, such as *verb.change* (verbs describing change in terms of size, temperature, intensity, etc.), *verb.cognition* (verbs of mental activities or processes), *verb.motion* (verbs of change in the spatial position), *verb.communication* (verbs describing communication and information exchange), etc.⁴

Verb synsets are interrelated and form a hierarchical structure based on a troponymy relation which represents a manner relation and is to a great degree analogous to hypernymy; for example, in *talk.v – whisper.v* the second member of the pair refers to a particular, semantically more specified, manner of performing the action referred to by the first verb (Fellbaum 1999a).

2.1.2 FrameNet

FrameNet⁵ (Fillmore et al. 1998, Baker 2008) is a lexical semantic resource which couches lexical and conceptual knowledge in the apparatus of frame semantics. Frames are conceptual structures describing types of objects, situations, or events along with their components (frame elements) (Fillmore et al. 1998, Ruppenhofer et al. 2016). Depending on their status, frame elements may be core, peripheral or extra-thematic (Ruppenhofer et al. 2016). We deal primarily with core frame elements, which instantiate conceptually essential components of a frame, and which in their particular configuration make a frame unique and different from other frames.

FrameNet frames are organised into a hierarchical network by means of a number of hierarchical and non-hierarchical frame-to-frame relations (Ruppenhofer et al. 2016: 81–84). Here we list the hierarchical relations, which bear most relevance to the internal structure of verb classes. These are: *Inheritance* – a relationship between a parent frame and a more specific (child) frame, such that the child frame elaborates the parent frame; *Uses* (also called “weak inheritance”) – a relationship between two frames where the first one makes reference in a very general kind of way to the structure of a more abstract, schematic frame; *Perspective* – a relation indicating that a situation viewed as neutral may be specified by

⁴The division of the nouns and verbs into WordNet lexicographic files (reflecting the semantic primitive distinction) along with short definitions of the primitives are available at: <https://wordnet.princeton.edu/documentation/lexnames5wn>.

⁵<https://framenet.icsi.berkeley.edu/fndrupal/>

means of perspectivised frames that represent different possible points-of-view on the neutral state-of-affairs; *Subframe* – a relation between a complex frame referring to sequences of states and transitions, each of which can itself be separately described as a frame, and the frames denoting these states or transitions.

2.2 Corpora

2.2.1 Semantically annotated corpora: SemCor and BulSemCor

In order to explore the syntactic expression of the verbs and their participants, we study the usage examples available in two semantically annotated corpora – the English SemCor and the Bulgarian semantically annotated corpus, BulSemCor, both of which are annotated with WordNet senses.

SemCor (current version 3.0) (Miller et al. 1993, 1994, 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). The corpus is the largest manually annotated corpus of this kind and amounts to a total of 226,040 sense annotations.

BulSemCor (Koeva et al. 2006, 2011) is designed according to the general methodology of the original SemCor and criteria for ensuring an appropriate coverage of contemporary general lexis. In addition to open-class words, BulSemCor includes annotation of closed-class words: prepositions, conjunctions, particles, pronouns and interjections; for that purpose the Bulgarian WordNet has been expanded with closed-class words (Koeva et al. 2011). The size of the corpus is close to 100,000 annotated units.

The size of the two corpora is not sufficient to provide enough evidence for many of the studied verbs so examples from other corpora have also been employed.

2.2.2 Bulgarian-English parallel corpus

The Bulgarian-English Sentence- and Clause-Aligned Corpus (BulEnAC)⁶ (Koeva, Rizov, et al. 2012) is a parallel corpus of aligned Bulgarian and English sentences and clauses with annotation of the syntactic relation between clauses. The corpus contains 366,865 tokens (176,397 tokens in Bulgarian and 190,468 tokens in English).

⁶https://dcl.bas.bg/en/resources_list/bulenac/

2 *Language-independent and language-specific properties*

The syntactic annotation of BulEnAC involves: a) sentence and clause splitting; b) annotation of the type of syntactic relation (coordinate or subordinate) between clauses. c) marking of the elements that introduce the clause: conjunctions, complementisers, and punctuation.

BulEnAC is suitable for extracting parallel sentences illustrating the use of particular verbs evoking the frames under study. Further, it facilitates the identification of corresponding translation equivalents within aligned clauses.

2.2.3 The Bulgarian National Corpus

The Bulgarian National Corpus is the largest corpus for Bulgarian: it consists of a monolingual (Bulgarian) part and 47 parallel corpora and amounts to 5.4 billion words. The Bulgarian part includes about 1.2 billion words of running text distributed in 240,000 text samples. The texts in the corpus reflect the state of the Bulgarian language predominantly in its written modality from the middle of the 20th century (1945) until the present day (Koeva, Stoyanova, et al. 2012). The search engine developed for the exploration of the corpus allows the extraction of information according to complex grammatical criteria. We use the corpus to study the syntactic expression and the validity of the valence patterns described in Section 4 in addition to the examples extracted from the semantically disambiguated part of the corpus (BulSemCor).

2.3 Motivation for combining WordNet and FrameNet

It has long been acknowledged that combining WordNet with conceptual resources such as FrameNet results in more comprehensive semantic and syntactic representation of the lexical entries (Baker & Fellbaum 2009, Schneider et al. 2012, Das et al. 2014), thus expanding the possible applications of the resources for the purposes of syntactic and semantic parsing. Elaborating a bit on the discussion of the strengths and shortcomings of the different kinds of lexical semantic resources offered by Shi & Mihalcea (2005), we may point out the following motivation for putting effort into their alignment.

FrameNet provides a rich semantic description of the predicates using schematic representations (frames) of the configurations of “participants and props” (elements corresponding to the surrounding circumstances or other supporting facets of meaning, in the sense of Ruppenhofer et al. 2016: 7) that define the situation described. The corpus of sentences annotated with explicit and implicit frame elements supplies empirical evidence about the syntactic realisations of semantic frames that is particularly valuable not only for linguistic generalisations about the target language (English) but also as a point of departure for

making observations cross-linguistically. Besides the explicit syntactic expression, the annotators have marked non-overt but conceptually present frame elements retrievable from the immediate or the more general context (so-called null instantiations). However, while formulating ontological semantic types that classify lexical units, frames and frame elements and in the latter case denote the selectional restrictions imposed on the fillers of frame elements (Ruppenhofer et al. 2016: 86), FrameNet does not explicitly define the content of these semantic types (see Section 3.2, which provides the authors' suggestions regarding that). In addition, FrameNet's coverage is limited both in terms of the lexical units included in the frames (i.e. there are lexical units pertaining to a frame that are not listed in it) and in terms of the parts of the lexicon encompassed by the system of frames, i.e. there are lexical units that cannot be described properly by the existing frames. Finally, as some of the frame elements are too finely-grained, certain generalisations across frames and frame elements might be missed.

WordNet ensures vast lexical coverage of the English lexicon structured and enriched with lexical and semantic information in the form of synset glosses, usage examples, notes on the usage or grammatical specificities, and a rich network of semantic relations. However, WordNet encodes no explicit semantic information about the participants in the situations described by the predicates and only limited information about their syntactic behaviour.

The combination of the resources requires: (i) mapping of the units that correspond to each other in the resources, i.e. discovering the counterparts of the synsets' members among the lexical units in FrameNet and linking them to the frames they evoke; (ii) expanding the mapping by discovering new candidates in WordNet to be matched to the relevant frames.

The combination of the resources helps tackling their shortcomings with respect to the comprehensive conceptual description. A particular deficiency to the optimal use of the rich semantic information provided by FrameNet is the relatively small coverage of the lexical units included in it. One way to alleviate this is to expand the coverage of FrameNet against the WordNet sense inventory by providing procedures for mapping WordNet synsets whose members evoke an existing frame but have not been matched with one until now. Such mapping procedures are discussed in Section 3. The limitations stemming from the lack of appropriate frames to describe certain parts of the lexicon need to be addressed by defining new frames.

The greater granularity of the frame elements in FrameNet (as compared with VerbNet, VerbAtlas and other resources) is handled, where necessary, by applying a shallow hierarchy derived from the hierarchical organisation of the frames

2 *Language-independent and language-specific properties*

and the inheritance relations defined between them (Litkowski 2014)⁷. Consider for instance the taxonomy of frame elements AIR > FLUID > THEME derived from the frame hierarchy Breathing > Fluidic motion > Motion built on the frame-to-frame relation of *Inheritance* between the three frames. In certain contexts and for certain tasks it may be more appropriate to make reference not to the most specific AIR but to FLUID or even to THEME, or vice versa. The maintaining of the different levels of granularity provides a more robust semantic description that is relatively resource- and theory-independent.

While genuinely beneficial, the mutual enrichment of WordNet and FrameNet is by no means trivial, as senses of the synsets and the lexical units that may be thought as equivalent may in fact not correspond well. The use of corpus occurrence and especially the study of annotated examples help in elucidating both theoretical and pragmatic aspects of the alignment between the resources and informs the judgments made in the course of the manual validation of the automatic assignment of frames to synsets. The case study presented in Section 4 may be viewed as the result of such analysis.

3 Mapping between WordNet and FrameNet based on universal principles

Both resources have shown to be sufficiently language-independent as to provide an approximation at a description across typologically distinct languages. Both models have been transferred and adapted cross-linguistically. These include coordinated attempts to build multilingual resources or link existing independent resources through projects such as EuroWordNet (Vossen 2004) or Global WordNet (McCrae et al. 2021), as well as Multilingual FrameNet (Gilardi & Baker 2018), among others.

Our work expands on the notion of universality and cross-lingual applicability of lexical-semantic resources by linking the resources to each other and then transferring language-independent (semantic and conceptual) description of English verbs in WordNet onto the Bulgarian lexical units in the Bulgarian WordNet.

⁷<https://www.cires.com/clr/fetax.php>

3.1 Universality of semantic inheritance relations between synsets and between frames

The two resources have been aligned automatically by employing existing mappings (Tonelli & Pighin 2009, Palmer et al. 2014, among others) with additional implemented procedures for expansion and validation (Leseva et al. 2018) and later refined (Leseva & Stoyanova 2019, 2020); these procedures involve 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.

Figure 1 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 (Cooking_creation being an elaboration of the mother frame Intentionally_create).

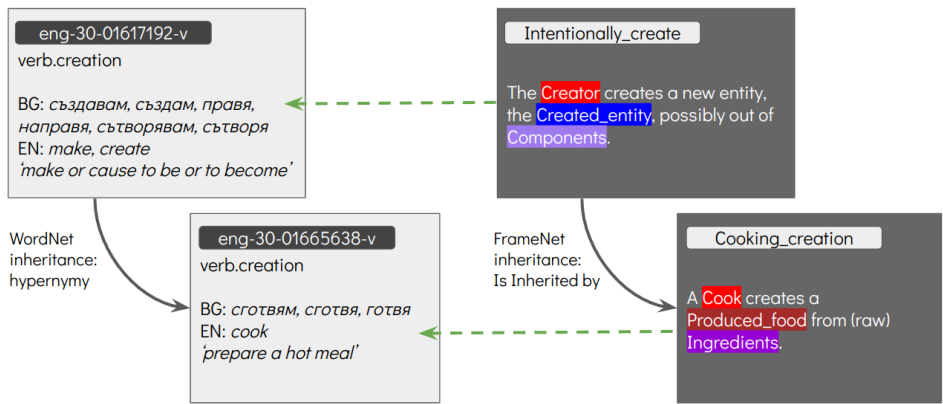


Figure 1: Frame inheritance (Intentionally_create → Cooking_creation) as reflected in the hypernym relation (make, create → cook).

3.2 Universality of selectional restrictions

Part of the FrameNet frame elements are supplied with ‘semantic types’ defining noun classes that narrow down the set of possible nouns that may be realised in the respective positions in the semantic frame. These semantic types are to

2 Language-independent and language-specific properties

a great degree relevant cross-linguistically, as they define ontological distinctions that underlie human cognition. To the best of our knowledge, the list of the FrameNet types and the pertaining definitions have not been made available, but their semantic content can be intuitively construed by speakers from the relevant designations, such as Sentient, Physical object, etc.). As noted in Ruppenhofer et al. (2016: 86) most ontological semantic types “correspond directly to synset nodes of WordNet, and can be mapped onto ontologies, e.g. Cyc or the Knowledge Graph”. The FrameNet semantic types form a semantic type hierarchy, which, however, does not necessarily correspond to that of WordNet or any other resource. Most of the frame-to-frame relations enable the propagation of the ontological semantic types of the parent frame and its frame elements down to the child frame and its frame elements (Ruppenhofer et al. 2016: 99) as well as to the lexical units in the respective frame (Ruppenhofer et al. 2016: 86). Using a linguistic taxonomy (moreover one implemented for numerous languages such as WordNet) to describe the selectional restrictions imposed by verbs on the nouns that fill the positions of their arguments has been proposed in different frameworks (Agirre & Martínez 2002, Koeva 2010). While the particulars differ, the general idea is the same as the one adopted in FrameNet, i.e. to represent semantic constraints in the form of taxonomically definable classes.

3.3 Universal and language-specific aspects of valence frames and syntactic realisation

Through the alignment between frames and synsets, each verb in WordNet is associated with a number of valence patterns defined for the lexical units evoking a given frame in FrameNet. While the semantic component of the description is language-independent, the syntactic component is more language-specific as the realisation of the frame elements depends on the syntactic properties of each language. Even so, we assume that the valence patterns that underlie the syntactic expression are valid cross-linguistically to a considerable degree as they are grounded in human cognition and the conceptualisation of situations. More precisely, valence patterns describe “the semantic and syntactic combinatory possibilities”, or valences of lexical units (Ruppenhofer et al. 2016: 7). They thus refer to the co-occurrence combinations of frame elements (both core and non-core) attested for each annotated lexical unit in the FrameNet annotated corpus.

The second, more language-specific level of syntactic description consists of the *syntactic categories and grammatical functions* by which a particular frame element for a given lexical unit is expressed. Even at this level, for many (related) languages one can observe similar syntactic expression especially with respect to

the participants that are selected as the subject and the object. A great degree of differentiation may be found at the level of certain grammatical peculiarities and constructions – for instance, unlike English, Bulgarian lacks *-ing* and infinitive clauses, so propositional complements will be realised as finite clauses; Bulgarian has impersonal verbs and subjectless sentences and does not make use of pleonastic subjects. Of course, there may be mismatches in the syntactic categories across languages, e.g. a certain frame element may be a direct object in one language and a prepositional object in another. Languages may also differ in terms of the overtiness of syntactic information, i.e. the possibility to leave an obligatory element non-explicit (null instantiations retrievable from the context or the grammatical construction); the language-specific diatheses, constructions, word order, morphosyntactic features, etc. The inventory of means that introduce certain frame elements such as prepositions, conjunctions, *wh*-words, etc. may also vary across languages.

The linking from the semantic level of the frame elements to the syntactic level of patterns of co-occurrence and syntactic categories in FrameNet is implemented in a straightforward manner by associating each frame element with a syntactic category and possibly a grammatical function – e.g. subject (NP.Ext) and object (NP.Obj).

Example 1 shows a partial representation of the valence patterns and the syntactic realisation of the verb *teach* in the FrameNet frame *Education_teaching*.

- (1) a. TEACHER INSTITUTION
NP.Ext PP[*at*]
- b. TEACHER STUDENT SUBJECT
NP.Ext NP.Obj PP[*about*]
- c. TEACHER STUDENT SKILL
NP.Ext NP.Obj Sinterrog/VPto

To sum up, even though there may be typological cross-linguistic differences in the conceptualisation and expression of situation for many language pairs, English and Bulgarian including, there are also parallels that facilitate the transfer of information across languages at the semantic and possibly at the syntactic level. Even where direct transfer of the syntactic description is not justified, the valence patterns and the syntactic realisation lattices taken from FrameNet may serve as a point of departure in the analysis of the Bulgarian syntactic data: they help establishing what is valid or invalid in Bulgarian by comparing the syntactic properties of the Bulgarian verbs to those of their English counterparts and the example sentences in the resources.

4 A case study: Verbs of communication

Below we offer an analysis of a selection of verbs of communication as an illustration of the universal principles and the language-specific features of the adopted linguistic description.

The domain of speech act verbs and their classification have been discussed by many authors (Wierzbicka 1987, Levin 1993: 202–211, Levin et al. 1997, Urban & Ruppenhofer 2001, Boas 2010, among others), including for Bulgarian (Nitsolova 2008, Penchev 1998, Tisheva 2000, 2004, Koeva 2021b, among others). While previous work in this area has served to inform the current state of the linguistic knowledge about the semantic and syntactic properties of communication verbs, the analysis below is based primarily on our observations on the descriptions proposed in FrameNet for English and exploring and extending them to Bulgarian.

First, we identify the “basic” frame which describes the general scenario or situation characterising the domain of communication in terms of the participants and circumstances involved and the relations among them (Johnson et al. 2001: 16). This general scenario is then elaborated in various ways in more specific frames. The semantic generalisations among such frames exhibiting different levels of abstraction and specialisation are typically cast in the form of frame-to-frame relations based on the inheritance among the semantic descriptions or parts of them.

The hierarchical organisation of the domain of communication verbs is presented in Figure 2.

Starting from this basic, or prototypical frame, we delve into several of the frames inheriting from it in order to show what kinds of processes are involved in the semantic specialisation and how this is reflected in the semantic description. The frames are selected based on the frequency of the verbs evoking them in the annotated data or with the objective to illustrate particular aspects of the analysis. For each such frame (including the prototypical one), we consider: (i) its semantics in terms of the frame definition, constellation of core frame elements that represent the main participants in the situation, and the relations among them, (ii) the syntactic expression of the frame elements, and (iii) the specifics of their realisation in Bulgarian as compared to English. The semantic and syntactic aspects referred to in (i) and (ii) are mostly taken for granted as represented in the FrameNet annotated corpus. In presenting each frame inheriting from the prototypical one, we do discuss how the conceptualisation of the basic frame is specialised or narrowed down and how this is reflected in the number of frame

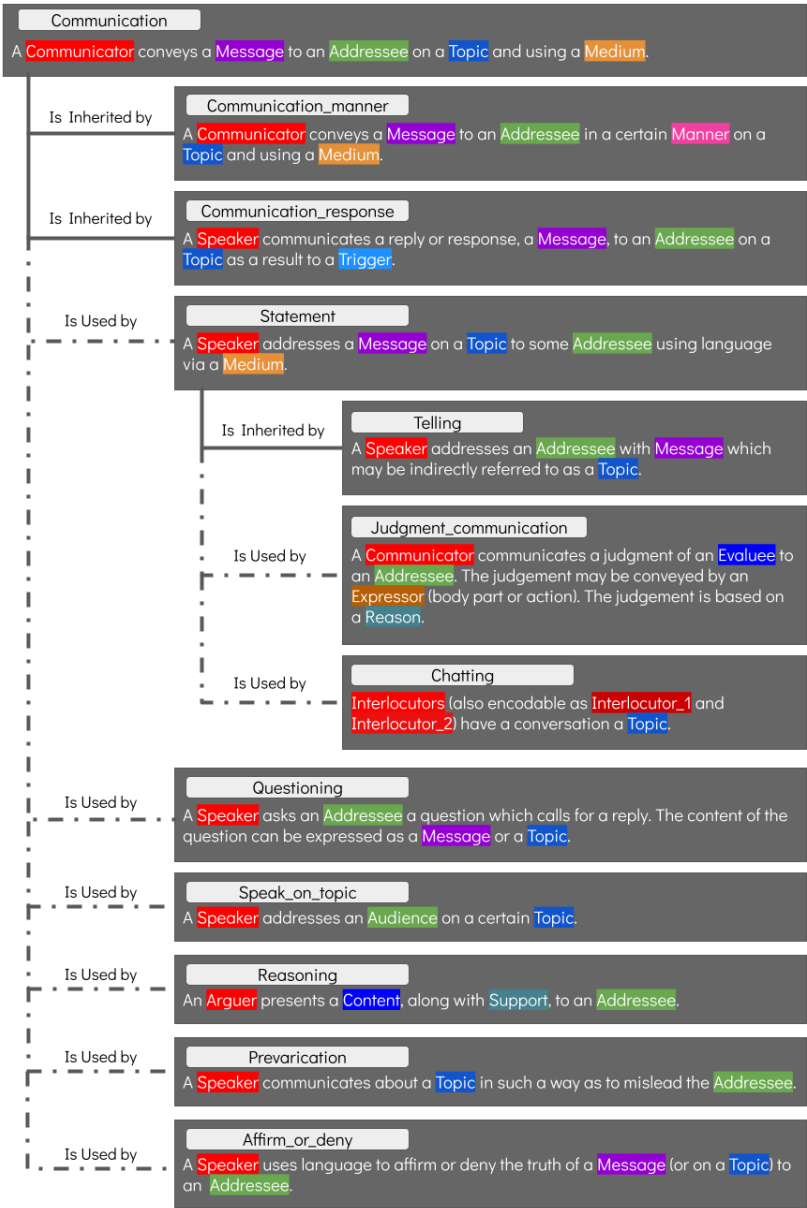


Figure 2: The hierarchical organisation of FrameNet frames describing the verbs of communication.

2 *Language-independent and language-specific properties*

elements and the relationships among them. The main burden of our work is focused on (iii), i.e. the analysis of the syntactic expression of the frame elements as attested in the corpus compiled for Bulgarian. The valence patterns emerge from the annotated examples and are thus specified independently from the English data. The same holds for the syntactic information (syntactic function and syntactic category of the expressed frame elements). The tagset of categories is adapted from the FrameNet corpus so that the notations in the two annotated datasets are unified.

Although there may be differences in the conceptualisation of situations across languages, we expect the semantic properties of the description to be largely shared between English and Bulgarian, as it has been shown by efforts undertaken for other languages (Section 1). Based on our preliminary observations, we also expect that at least part of the valence patterns will be relevant for both languages, i.e. the frame elements that tend to be expressed and the particular configurations in which they co-occur will be similar, even allowing for cross-lingual differences (such as the fact that Bulgarian, unlike English, is a pro-drop language). We then look at the syntactic expression of the patterns in terms of the grammatical function and the syntactic categories of the core frame elements and, where relevant, the possibility for their contextual construal (null instantiations).

We take as a point of departure the lattices of the frame elements and their syntactic realisations for certain verbs and the valence patterns of frame elements as described in the annotated FrameNet examples⁸ (Burchardt & Pennacchiotti 2008). In addition, below we also use examples from SemCor in order to illustrate the applicability of the FrameNet description independently of the annotation undertaken in the FrameNet corpus.

After analysing this information for English, we go on to observe to what extent it is applicable to Bulgarian. For this purpose, we have constructed a corpus of manually annotated examples extracted from BulSemCor and, where the number of examples is not sufficient, from the Bulgarian National Corpus.

Each example sentence in the English and the Bulgarian dataset is annotated as shown in Example 2. The English dataset consists of 93 verbs (lexical units in FrameNet) to which an appropriate communication-related frame is assigned. The verbs are aligned to 72 WordNet synsets. Each verb is supplied with a number of examples from the FrameNet corpus illustrating its valence patterns; the dataset contains a total of 4,525 illustration examples representing 863 different valence patterns. The annotation of each sentence in the Berkeley FrameNet cor-

⁸<http://framenet.icsi.berkeley.edu/>

Svetlozara Leseva & Ivelina Stoyanova

pus includes explicit annotation of the target word (in this case a verb) and the syntactic realisation of the frame elements.

The Bulgarian dataset covers 112 communication verbs (including aspectual pairs) across 63 WordNet synsets. As the corpus of annotated examples for Bulgarian is still work in progress, it is considerably smaller than the one for English: it contains 890 annotated sentences representing 136 different patterns. The annotation consists in labelling the sentence components with the frame elements they realise in a way consistent with the annotation in the Berkeley FrameNet.

- (2) a. FrameNet description: *ask.v* ‘say something in order to obtain an answer or some information from someone’, frame: Questioning
 WordNet alignment: {*ask*:4} ‘address a question to and expect an answer from’, synset ID: eng-30-00897746-v
 BulNet alignment: {*numam*:2, *nonumvam*:1, *nonumam*:1, *zanumvam*:3, *zanumam*:3}, synset ID: eng-30-00897746-v
- b. An adapted example from the FrameNet corpus with the relevant pattern:
[They]_{COM} ASKED [Rubbie]_{ADDR} [what she ate]_{MSG}.
[NP.Ext]_{COM} VERB [NP.Obj]_{ADDR} [Sinterrog]_{MSG}
- c. An annotated example from BulSemCor with the relevant pattern:
[Престъпникът]_{COM} ПОПИТАЛ [полицая]_{ADDR}
 Criminal-DEF asked policeman-DEF
[дали може да си купи цигари]_{MSG}.
 whether he could buy cigarettes.
 ‘The criminal asked the policeman whether he could buy cigarettes.’
[NP.Ext]_{COM} [NP.Obj]_{ADDR} [Sinterrog]_{MSG}

4.1 The prototypical frame: Communication

As noted by Johnson et al. (2001: 108), the frames in the domain of communication describe “verbal communication between people and inherit structure and frame elements from the higher-level frame Communication”. Communication is thus the prototypical frame that represents the basic conceptual structure of the activity of communication as a configuration of five main interacting frame elements. This basic structure will be further elaborated (narrowed down, profiled or otherwise specialised) in the frames that inherit it⁹.

⁹By “inherit” we mean the relationships between the more general and the more specific frames between which the following hierarchical frame-to-frame relations hold: *Inheritance*, *Using*, *Perspectivises*, *Subframe*.

2 Language-independent and language-specific properties

Definition of the frame Communication: A COMMUNICATOR conveys a MESSAGE to an ADDRESSEE; the TOPIC and MEDIUM of the communication may also be expressed.

As described in the definition, the Communication frame does not itself involve specification of the method of communication (speech, writing, gesture, etc.) but only the fact of it. The frames that inherit Communication can add elaboration to the general idea in several ways:

- (i) by specifying the MEDIUM in a variety of ways, such as the particular language (in French, in Russian), or the physical entity or channel, e.g. a medium, technology, form, etc. (*on the radio, in a letter, through the Messenger, in writing*).
- (ii) by specifying the manner of verbal communication according to various criteria such as loudness (e.g. *shout.v, whisper.v*); volubility and/or mood (e.g., *babble.v, rant.v*), distinctness (e.g., *slur.v, stutter.v, mutter.v*), among many others;
- (iii) specialisation may also mean that the more concrete frames inherit only part of the Communication frame elements or do not inherit them in a straightforward manner. For example, Judgment_communication (which inherits from Statement, in turn inheriting from Communication according to the *Using* relation, see Figure 2 above) reinterprets the frame element MESSAGE as a judgement on an EVALUEE according to a REASON.

The prototypical and the inheriting frames might exhibit a different construal of the relationship between certain frame elements. For instance (as pointed out in the description of Communication), in the frame Chatting, the COMMUNICATOR and ADDRESSEE alternate their roles, and are often expressed by a single, plural NP, i.e. the relationship between them is not asymmetrical but reciprocal as they participate in the situation in the same way.

Another aspect of specialisation is the inability for overt expression of all the frame elements (Johnson et al. 2001: 16). For example, the lexical units *talk.v* and *speak.v* in the Statement frame (which inherits Communication according to the *Using* relation) usually block the overt expression of MESSAGE, although its existence is implied at the conceptual level (in their meaning). This is shown by the fact that in the annotated examples available for the two verbs the frame element TOPIC is much more frequently expressed than MESSAGE, although it is dependent on it (the topic characterises the message).

Svetlozara Leseva & Ivelina Stoyanova

Another kind of elaboration is represented by the incorporation of frame elements (Jackendoff 1990: 164–165) whereby a certain frame element is integrated in the meaning of a verb as a result of which this frame element is usually left unexpressed (Ruppenhofer et al. 2016: 30). In the domain of Communication the frame *Communication_means* describes situations that specify the concrete means with the aid of which communication takes place; the various *MEANS* are thus incorporated in the meaning of the respective verbs, e.g. *fax.v*, *telephone.v*, *email.v*.

The frame *Communication* is evoked by a small number of verbs – *communicate.v*, *convey.v*, *indicate.v*, *share.v*. Although pertaining to the prototypical frame, these verbs are not the most frequent ones associated with the activity of communicating, which are in fact described in more elaborate frames.

4.1.1 Prototypical frame elements in the domain of communication

Below we present the prototypical frame elements of the *Communication* frame as defined in FrameNet.

COMMUNICATOR (Semantic type: Sentient) The sentient entity that uses language in the written or spoken modality to convey a *MESSAGE* to the *ADDRESSEE*.

MEDIUM The physical or abstract setting in which the *MESSAGE* is conveyed.

MESSAGE (Semantic type: Message) A proposition or set of propositions that the *COMMUNICATOR* wants the *ADDRESSEE* to believe or take for granted; in other words it is the content which is communicated.

TOPIC The subject matter to which the *MESSAGE* pertains. It is thus a property of *MESSAGE* (Johnson et al. 2001: 17) and as a result its syntactic expression is also predetermined by the expression of the *MESSAGE*.

ADDRESSEE (Semantic type: Sentient): The *ADDRESSEE* is typically a person or organisation, etc. that receives a *MESSAGE* from the *COMMUNICATOR*.¹⁰

In the remainder of the chapter the data in the annotated corpora that are subject to analysis are organised as follows. We first show and discuss how each of the considered frame elements is realised at the level of the individual verbs evoking a given frame (the odd-numbered tables). This kind of presentation allows us

¹⁰In the FrameNet frame *Communication* the *ADDRESSEE* is specified as a non-core element. However, we consider it is nonetheless implied in all examples from the FrameNet annotated corpus and thus analyse it in the set of prototypical frame elements.

2 *Language-independent and language-specific properties*

to observe the expression of each frame element for each verb and the differences among verbs in the same frame. The data shown in the pairs of odd-numbered tables enable the comparison between English and Bulgarian and help in drawing conclusions about the correspondences and differences in the syntactic realisation between the two languages. These tables, however, do not represent the configurations of frame elements that actually occur in the annotated corpora. To illustrate those, we give a summarised list of the most characteristic valence patterns for each frame (i.e. the best-represented patterns in terms of numbers of examples) and the verbs that are observed in these configurations in the two languages (the even-numbered tables). The information in the subsequent odd- and even-numbered tables is thus complementary. Due to the currently insufficient number of examples even for many English verbs, we represent the valence patterns as an aggregate of the valence patterns for all verb¹¹, thus obtaining what we call generalised valence patterns. These give us an overall idea of the distribution of valence patterns across verbs and a point of departure for a more in-depth evidence based analysis.¹²

4.1.2 Syntactic realisation of the **Communication** frame elements

The syntactic expression of the basic configuration of frame elements in the Communication frame is exemplified in Table 1.

COMMUNICATOR is expressed as the external argument, i.e. as a subject of the respective sentence or clause; as it is a sentient entity, it is realised as an NP. In a number of cases the frame element is realised as a definite null instantiation (DNI), i.e. it is retrievable from the previous context, or as a constructional null instantiation (CNI), where it is the grammatical construction that allows it to remain non-overt, e.g. in passive or infinitive clauses, etc.

Here and below, unless the distinction is specifically relevant, we consider INIs (indefinite null instantiations), CNIs (constructional null instantiations) and DNIs (definite null instantiations) as one category – NI (null instantiations), together with the category INC (incorporated frame element) (Petruck 2019). The null instantiations are a very interesting category that merits a separate in-depth study. In particular, they may be considered as exponents of distinct properties,

¹¹In theory, the differences among the individual verbs are lost in this way, but since we do not have at our disposal large samples of annotated data for each verb, in practice, this is not relevant as the sparseness of data prevents us from making such detailed observations.

¹²The numbers in the tables for English are based on a version of the Berkeley FrameNet obtained in XML format in 2019.

Table 1: Syntactic expression of the Communication frame elements of selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>communicate</i>									
COMMUNICATOR	39				5				44
ADDRESSEE			27		16			1	44
MESSAGE	3	23			14	1			41
TOPIC		1	3		1			1	6
MEDIUM	2		3					1	6
<i>indicate</i>									
COMMUNICATOR	7								7
ADDRESSEE					8				8
MESSAGE		3				6			9
MEDIUM	4								4
<i>say</i>									
COMMUNICATOR	5				6				11
MEDIUM	5				1				6
MESSAGE	6					9	2		17
TOPIC			1		1				2

may stand for different syntactic categories and constituents with different grammatical functions, and respectively – may participate in different valence patterns. However, the distinction among them is not trivial and especially the one between DNIs and INIs may require a broader context to be interpreted accurately. In addition, as this has not been the focus of study, sufficient number of examples and broad enough context has not been provided in the Bulgarian data.¹³

With the verbs in this frame, MESSAGE is typically realised as an object NP, as a complement clause (Example 3a) or as a quote. Quotes represent the content of the MESSAGE as directly stated by the COMMUNICATOR in his or her own words, while clauses denote it as being retold by someone (such as in reported speech). A MESSAGE realised as an NP constitute a nominalisation which rephrases its content in a more concise way or as a generalised idea. In about a third of the examples available for *communicate.v* the MESSAGE is annotated as an indefinite null instantiation (INI). This means that the verb is used intransitively: the MESSAGE remains syntactically unexpressed and receives a certain typical interpre-

¹³The category ‘Other’ encompasses examples where a frame element is otherwise expressed. Due to the limited number of such instances, we omit them here.

2 Language-independent and language-specific properties

tation without a specific discourse referent (Ruppenhofer et al. 2016) (Example 3b). The INIs correspond to the activity use of certain types of verbs where the object remains implicit (Van Valin & LaPolla 1997).

The FrameNet examples show that TOPIC is rarely expressed, with only several instances in the FrameNet corpus even for *communicate.v*. Extrapolating from examples from other sources and the definition of the frame element, we may conclude that the TOPIC is usually expressed as a prepositional phrase headed by the preposition *about*. An alternative way of realising the TOPIC is as a modifier of a noun expressing the MESSAGE (Example 3c); such cases corroborate syntactically its semantic dependence on the MESSAGE communicated. In the absence of an overt MESSAGE, the TOPIC may be expressed as an independent phrase (Example 3d); this is one of the typical patterns of its realisation as attested in the more specific communication frames.

MEDIUM is expressed either as a prepositional phrase, or as the subject in the case of a non-overt COMMUNICATOR.

ADDRESSEE is either realised as a prepositional phrase or is left unexpressed, although its presence is always required conceptually as every act of communication is addressed to someone. Predominantly, the non-overt ADDRESSEE frame elements are indefinite null instantiations (INI).

- (3) a. [*Iranian officials*]_{COM} **INDICATE** [*that Iran would honor its safeguards agreement with the IAEA*]_{MSG} [_{ADDR-INI}].
- b. [*They*]_{COM} *can easily* **COMMUNICATE** [_{MSG-INI}] [*with one another*]_{ADDR}.
- c. [*The letter*]_{COM} **COMMUNICATED** [*nothing*]_{MSG} [*of her pleasure and love*]_{TOP}.
- d. [*I*]_{COM} **COMMUNICATED** [*with the Minister*]_{ADDR} [*on that issue*]_{TOP}.

4.1.3 Communication valence patterns

Communication valence patterns are presented in Table 2.

The most common valence pattern found in the data is represented as an expressed subject NP COMMUNICATOR, an object NP MESSAGE and an ADDRESSEE PP. The MESSAGE is usually expressed and when it is not – the TOPIC may be realised (Example 3). Due to the small number of examples, this last observation is not included in the table, but it is supported by the expression of the relevant frame element in the more specific frames.

Table 2: FrameNet valence patterns of Communication verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{COM} [PP] _{ADDR} [NP] _{MSG}	11	<i>communicate, signal</i>
[NP.Ext] _{COM} [PP] _{ADDR} [_{MSG} -INI	7	<i>communicate</i>
[NP.Ext] _{MSG} [_{COM} -CNI [Clause] _{MSG}	5	<i>say</i>
[NP.Ext] _{COM} [_{ADDR} -INI [NP] _{MSG}	5	<i>communicate</i>
[NP.Ext] _{COM} [_{ADDR} -INI [_{MSG} -INI	4	<i>communicate</i>
[NP.Ext] _{COM} [Clause] _{MSG}	4	<i>indicate, say, signal</i>
[NP.Ext] _{MSG} [PP] _{ADDR} [_{COM} -CNI	3	<i>communicate</i>
[NP.Ext] _{COM} [_{ADDR} -DNI [NP] _{MSG}	3	<i>communicate, indicate</i>
[NP.Ext] _{COM} [PP] _{ADDR} [NP] _{MSG} [PP] _{TOP}	2	<i>communicate</i>
[NP.Ext] _{MED} [_{ADDR} -INI [Clause] _{MSG}	2	<i>indicate</i>

4.1.4 Syntactic realisation of the Communication frame in Bulgarian

The core frame elements are expressed in a similar way as in English: the COMMUNICATOR is realised as a subject, the MESSAGE is an NP object or more rarely (although varying from verb to verb) a complement clause or a quote; if overt, the ADDRESSEE is expressed as a prepositional phrase. The TOPIC is syntactically explicit in about 20% of the cases and, similarly to English, is realised as either a prepositional phrase that modifies a MESSAGE head noun (Example 4a) or independently in the absence of an overt MESSAGE (Example 4b); the number of examples is too small to make definitive conclusions, but both languages support this observation.

- (4) a. [Te]_{COM} СЪОБЩАВАТ [съответната информация]_{MSG} [за
 They communicate relevant information about
 дейността си]_{TOP}.
 activity-DEF REFL.
 ‘They communicate relevant information about their activity.’
- b. [Te]_{COM} СЪОБЩАВАТ [_{MSG}-INI [за пристигането си на
 They communicate about arrival-DEF REFL at
 гарата]_{TOP}.
 station-DEF.
 ‘They communicate about their arrival at the station.’

2 Language-independent and language-specific properties

- c. [Те]_{COM} **СЪОБЩАВАТ** [на Комисията]_{ADDR} [текста на
They communicate to Commission-DEF text-DEF of
разпоредбите]_{MSG}.
measures-DEF.
'They communicate to the Commission the text of the measures.'
- d. [Органите]_{COM} **СЪОБЩАВАТ** [цялата съществена
Authorities-DEF communicate all essential
информация]_{MSG} [_{ADDR}].
information.
'The authorities communicate all essential information.'
- e. [Страните]_{COM} **ПОСОЧВАТ**, [че информацията не може да
Parties-DEF indicate that information-DEF cannot to
бъде резюмирана]_{MSG} [_{ADDR}].
be summarised.
'The parties indicate that the information cannot be summarised.'

The syntactic realisation of the Communication frame elements in Bulgarian is shown in Table 3.¹⁴

The valence patterns in Bulgarian (Table 4) show similar preferences for the co-occurrence of frame elements; with both MESSAGE and ADDRESSEE expressed (Example 4c) or with a realised MESSAGE and a non-overt ADDRESSEE (Example 4d).

In the Bulgarian data we have found only rare instances where there is an expressed ADDRESSEE with non-overt MESSAGE or TOPIC, but this observation needs further corroboration from the data for this frame as well as for other related frames.

The MESSAGE can also be expressed as a quote or a clausal complement; however, as Bulgarian lacks infinitives and *-ing* clauses, clausal complements are realised as finite clauses (Example 4e).

4.2 Frame Communication_manner

Definition of the frame Communication_manner: The words in this frame describe MANNERS of verbal communication. Core frame elements: SPEAKER, MESSAGE, TOPIC, ADDRESSEE.

¹⁴In the Bulgarian annotated data the verbs are assigned a WordNet sense, so the corresponding Princeton WordNet synset serves as an English translation equivalent. As this information is not available to the readers, henceforth we have provided translation equivalents for the Bulgarian verbs.

Table 3: Syntactic expression of the Communication frame elements in Bulgarian.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>споделям/споделя</i> (share)									
COMMUNICATOR	14								14
MESSAGE		11			2	1			14
ADDRESSEE			12		2				14
TOPIC			1						1
<i>съобщавам/съобща</i> (communicate)									
COMMUNICATOR	29								29
MESSAGE		22			6	2			30
ADDRESSEE			22		8				30
MEDIUM			1						1
TOPIC		1	6						7
<i>предавам/предам</i> (convey)									
COMMUNICATOR	48								48
MESSAGE	3	42			1	1	1		48
ADDRESSEE			28		19				47

The **SPEAKER** is a specific type of **COMMUNICATOR** who uses his or her voice to produce the **MESSAGE**. Thus, apart from being a sentient being, it needs to be able to produce speech, e.g. is typically a person (Example 5a). The type of communication involves characteristics of individual organisms, so organisations are not typically realised as **SPEAKERS**, but small groups of people can be (Example 5b).

In particular, the verbs in the **Communication_manner** frame describe various manners of speaking or vocalising whereby a **SPEAKER** conveys a **MESSAGE** to the **ADDRESSEE**. The focus is on the specifics of the articulation or vocalisation such as clarity, speed, loudness, etc. Thus, the **MANNER** of the communication is incorporated in the lexical meaning of the verb, e.g. *whisper.v* (‘speak very softly using one’s breath’), *babble.v* (‘talk rapidly and continuously’), etc.; the **MANNER** can appear overtly when expressing additional manner meaning than the one incorporated by the verb (Example 5e).

The **MEDIUM** of communication is peripheral to the conceptualisation of the frame and thus has a non-core status.

The remaining core frame elements, i.e. **MESSAGE** and **TOPIC**, have the same specifics as in the **Communication** frame.

2 Language-independent and language-specific properties

Table 4: FrameNet valence patterns of Communication verbs, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *предавам/предам* (convey), *споделям/споделя* (share), *съобщавам/съобща* (communicate).

Pattern	#	verbs
[NP.Ext] _{COM} [NP.Obj] _{MSG} [PP] _{ADDR}	50	<i>предавам/предам,</i> <i>споделям/споделя,</i> <i>съобщавам/съобща</i>
[NP.Ext] _{COM} [NP.Obj] _{MSG} [] _{ADDR-INI}	13	<i>предавам/предам,</i> <i>съобщавам/съобща</i>
[NP.Ext] _{COM} [NP.Obj] _{MSG} [] _{ADDR-DNI}	9	<i>споделям/споделя,</i> <i>предавам/предам</i>
[NP.Ext] _{COM} [PP] _{ADDR} [PP] _{TOP} [] _{MSG-INI}	4	<i>споделям/споделя,</i> <i>съобщавам/съобща</i>
[NP.Ext] _{COM} [PP] _{ADDR} [] _{MSG-DNI}	3	<i>споделям/споделя,</i> <i>съобщавам/съобща</i>
[NP.Ext] _{COM} [Clause] _{MSG} [] _{ADDR-INI}	2	<i>съобщавам/съобща</i>
[NP.Ext] _{COM} [NP.Obj] _{MSG} [PP] _{ADDR} [PP] _{TOP}	1	<i>съобщавам/съобща</i>

4.2.1 Syntactic realisation of the Communication_manner frame elements

The syntactic expression of the basic configuration of frame elements in the Communication_manner frame is similar to the one in the Communication frame, but there are differences that we point out below. Like COMMUNICATOR, the SPEAKER is the external argument and is realised as the subject, which, under some contextually or constructionally grounded circumstances can be left implicit.

Similarly to the same frame element in the Communication frame, the MESSAGE can be expressed as a subordinate clause (Example 5c), a quoted expression (Example 5d), or an NP object that generalises over the type of information (Example 5a). In some cases the MESSAGE can be unexpressed (Example 5e).

The TOPIC is typically expressed as a prepositional phrase complement headed by ‘about’ (Example 5f). An alternative type of pattern is for it to be left implicit (a null instantiation), especially in the presence of a MESSAGE. As shown above, the two frame elements co-occur overtly primarily as an NP and a PP, where the TOPIC PP should be treated as a modifier of the MESSAGE NP (Example 5g).

The ADDRESSEE is typically left non-overt but is always implied; otherwise it is expressed as a prepositional phrase (Examples 5a, 5d, 5e).

Svetlozara Leseva & Ivelina Stoyanova

Among the verbs in this frame, certain differences may also be found. For instance, *rave.v* and *rant.v* tend to express overtly the TOPIC more often than the MESSAGE as compared with the purely manner verbs, which give preference to the MESSAGE itself.

- (5) a. [Ann]_{COM} **WHISPERED** [the question]_{MSG} [to Harry]_{ADDR}.
- b. [The crowd]_{COM} **CHANTED** [my name]_{MSG} [_{ADDR-INI}].
- c. [He]_{COM} **MUMBLED** [that he was in a state of shock]_{MSG} [_{ADDR-INI}].
- d. ['Change of plan,']_{MSG} [Peter]_{COM} **SHOUTED OUT** [to Kelly]_{ADDR}.
- e. [I]_{COM} was **SINGING** [_{MSG-INI}] [happily]_{MANR} [to myself]_{ADDR}.
- f. [He]_{COM} was **RAVING** [_{MSG-INI}] [about Armageddon]_{TOP} [_{ADDR-INI}].
- g. [He]_{COM} **MUMBLED** [something]_{MSG} [about something or other]_{TOP}.

The specifics of the syntactic expression of the basic configuration of frame elements in the Communication_manner frame is exemplified in Table 5.

4.2.2 Communication_manner valence patterns

Table 6 shows the prevalent valence patterns found with the verbs evoking the Communication_manner frame in the FrameNet annotated corpus.

The most frequent patterns include a canonical expression of the SPEAKER as a subject NP and a MESSAGE realised as either a direct quote, an object NP or a clausal complement. The valence patterns also show most frequently a non-overt or less often an expressed ADDRESSEE.

4.2.3 Syntactic realisation of Communication_manner frame in Bulgarian

In a similar manner, in Bulgarian the SPEAKER is realised as the external subject NP, while the MESSAGE is expressed as a direct quote (Example 6a), a finite complement clause (Example 6b) or an NP Object (Example 6c). The TOPIC and the ADDRESSEE are expressed as prepositional complements (Example 6d).

- (6) a. [_{SPKR-DNI}] *Едџа ПРОВОИВБАМ:* [– *За какво*
Hardly whisper.1sg: – About what
става дума?]_{MSG} [_{ADDR-INI}]
take place word?
'I hardly whisper: – What is it about?'

2 Language-independent and language-specific properties

Table 5: Syntactic expression of the Communication_manner frame elements in selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>mutter</i>									
SPEAKER	89								89
ADDRESSEE			21		68				89
TOPIC			20		14				34
MESSAGE		32				7	26		65
<i>rave</i>									
SPEAKER	27								27
ADDRESSEE			3		24				27
TOPIC		3	13		8				24
MESSAGE		1	1				3		5
<i>shout</i>									
ADDRESSEE	3		38		83				124
SPEAKER	116				5				121
TOPIC			3		34				37
MESSAGE	2	38	5			15	26		86
<i>sing</i>									
SPEAKER	59		6		2				67
ADDRESSEE			8		58				66
MESSAGE	8	33			24		2		67
TOPIC			14					1	15
<i>whisper</i>									
SPEAKER	47				5			1	53
ADDRESSEE			16		37				53
MESSAGE	6	14				8	9		37
TOPIC			7		12				19

- b. []_{SPKR-DNI} **ПРОМЪРМОРВАМ**, [че отчаяно искам да си
Mutter.1sg that desperately want.1sg to REFL
го върна]_{MSG} []_{ADDR-INI}.
it get back.
'I mutter that I desperately want to get it back.'
- c. [Аз]_{SPKR} **ИЗМЪНКАХ** [някакъв отговор]_{MSG} []_{ADDR}.
I stammered some reply.
- d. [Тя]_{SPKR} **ДРЪНКА** [на всички]_{ADDR} [за мен]_{TOP}.
She babbles to everyone about me.

Svetlozara Leseva & Ivelina Stoyanova

Table 6: FrameNet valence patterns of Communication_manner verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{SPKR} [] _{ADDR-INI} [Quote] _{MSG}	166	<i>rant, chant, slur, stutter, stammer, babble, chatter, rave, mumble, mutter, whisper, sing, shout</i>
[NP.Ext] _{SPKR} [] _{ADDR-INI} [] _{TOP-INI}	156	<i>rant, chant, slur, stutter, stammer, babble, chatter, rave, mumble, mutter, whisper, shout</i>
[NP.Ext] _{SPKR} [] _{ADDR-INI} [NP.Obj] _{MSG}	146	<i>rant, chant, slur, stutter, stammer, babble, chatter, mumble, mutter, whisper, sing, shout</i>
[NP.Ext] _{SPKR} [] _{ADDR-INI} [PP] _{TOP}	70	<i>rant, babble, chatter, rave, mumble, mutter, whisper, shout</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [] _{TOP-INI}	48	<i>rant, chant, babble, chatter, rave, mumble, mutter, whisper, shout</i>
[NP.Ext] _{SPKR} [] _{ADDR-INI} [Clause] _{MSG}	41	<i>rant, chant, mumble, mutter, stutter, stammer, whisper, shout</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [NP.Obj] _{MSG}	34	<i>mumble, mutter, stutter, whisper, sing, shout</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [Quote] _{MSG}	31	<i>rant, chant, mumble, mutter, whisper, shout</i>
[NP.Ext] _{SPKR} [] _{ADDR-INI} [NP.Obj] _{MSG} [PP] _{TOP}	21	<i>mumble, mutter, stammer, babble, sing, shout</i>

2 Language-independent and language-specific properties

- e. [Капитанът]_{SPKR} продължи да **КРЕЩИ** [заповедите си]_{MSG} [за Captain-DEF continued to shout orders REFL about разни платна и въжета]_{TOP}.
some sails and ropes.
'The captain continued shouting his orders about sails and ropes.'
- f. [– Здравейте]_{MSG} – **ИЗМЪНКВАМ** [аз]_{SPKR} [–]_{ADDR}
– Hello – mumble I
нерешително.
hesitantly.

Table 7: Syntactic expression of the Communication_manner frame elements in Bulgarian.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>шепна</i> (whisper)									
MESSAGE		3			1		3		7
ADDRESSEE			6		2				8
SPEAKER	8								8
<i>промърморвам/промърморя</i> (mumble, mutter)									
MESSAGE		2					8	3	13
ADDRESSEE			3		10				13
MEDIUM			1						1
SPEAKER	13								13
<i>викам</i> (shout)									
MESSAGE		3			1		23	4	31
ADDRESSEE			2		33				35
MEDIUM			1						1
SPEAKER	35								35
<i>прошепвам/прошепна</i> (whisper)									
MESSAGE		8					7	1	16
ADDRESSEE			6		10				16
SPEAKER	17								17

Table 7 shows a selection of verbs in Bulgarian evoking the frame Communication_manner, while Table 8 presents the most frequent valence patterns. The syntactic realisation is similar to English: strong preference for the overt expression

Svetlozara Leseva & Ivelina Stoyanova

of the MESSAGE either together with the ADDRESSEE or in its absence; realising the TOPIC most often either in the absence of (Example 6d) or as a modifier to the MESSAGE (Example 6e).

We can also note that at least for some manner verbs such as *мърморя* ‘mumble, mutter’, *мънкам* ‘stutter’, there is a marked trend of expressing the MESSAGE as a quote rather than as a complement clause.

4.3 Frame Statement

Definition of the frame Statement: A SPEAKER addresses a MESSAGE to some ADDRESSEE using language. Instead of (or in addition to) a SPEAKER, a MEDIUM may also be mentioned. Likewise, a TOPIC may be stated instead of a MESSAGE. Core frame elements: SPEAKER, MESSAGE, MEDIUM, TOPIC; Non-core: ADDRESSEE.

This frame represents the greatest number of verbs of speech, including many general lexis verbs such as *say.v*, *state.v*, *declare.v*, *speak.v*, *report.v*, *note.v*, etc.

4.3.1 Syntactic realisation of the Statement frame elements

The frame Statement is an elaboration of the prototypical frame Communication which specifies verbs for communication involving language. This is reflected by the fact that the COMMUNICATOR is conceptualised as the more specific SPEAKER, which denotes the person who produces the message. Likewise, this The frame element SPEAKER is realised as the external NP.

The MESSAGE is typically expressed either as a subordinate clause, an NP object, or a direct quote that represents the content being conveyed (Example 7a, 7b, 7c, respectively). There is a range of preferred realisations of the MESSAGE with the different verbs in this frame: some of them have a stronger tendency to take a complement subordinate clause (e.g., *claim.v*, *suggest.v*, *note.v*), while others show preference for an NP object (e.g., *profess.v*, *reiterate.v*, *relate.v*) or a quote (e.g., *exclaim.v*); in some cases the three realisations are equally likely (e.g., *caution.v*).

The TOPIC is typically expressed as a prepositional phrase headed by different prepositions depending on the verb, e.g. (*speak about him*, *speak of him*, *preach of heaven*, *comment on the protests*, *comment upon the economic conditions*), a trend inherited from the Communication frame. Similarly to the frames discussed above, usually either the MESSAGE or the TOPIC is expressed; as expected, they may also occur together in a phrase (Example 7b), where the TOPIC is syntactically

2 Language-independent and language-specific properties

Table 8: FrameNet valence patterns of Communication_manner verbs, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *бъбря* (babble, prattle), *викам* (shout), *дрънкам* (rattle, jabber), *заеквам/заекна* (stammer, stutter), *креця* (shout, yell), *мърморя, промърморвам/промърморя* (mumble, mutter), *пошушвам/пошушна, прошепвам/прошепна, шепна, шушна, шушукам* (whisper), *смотолевям/смотолевя* (mumble, falter).

Pattern	#	verbs
[NP.Ext] _{SPKR} [Quote] _{MSG} [] _{ADDR-INI}	48	<i>бъбря, викам, заеквам/заекна, креця, пошушвам/пошушна, промърморвам/ промърморя, прошепвам/прошепна, смотолевям/смотолевя, шепна</i>
[NP.Ext] _{SPKR} [NP.Obj] _{MSG} [] _{ADDR-INI}	13	<i>бърбора, викам, дрънкам, креця, мърморя, прошепвам/прошепна, шепна, промърморвам/ промърморя</i>
[NP.Ext] _{SPKR} [] _{ADDR-INI}	12	<i>бъбря, викам, дрънкам, заеквам/заекна, креця, мърморя, шушукам</i>
[NP.Ext] _{SPKR} [NP.Obj] _{MSG} [PP] _{ADDR}	11	<i>викам, дрънкам, креця, мърморя, прошепвам/прошепна, шепна, шушукам</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [Quote] _{MSG}	10	<i>бъбря, викам, прошепвам/прошепна, шепна, шушна, промърморвам/ промърморя</i>
[NP.Ext] _{SPKR} [Clause-that] _{MSG} [] _{ADDR-INI}	6	<i>викам, дрънкам, шушна, промърморвам/ промърморя</i>
[NP.Ext] _{SPKR} [PP] _{TOP} [] _{ADDR-INI}	2	<i>дрънкам</i>

Svetlozara Leseva & Ivelina Stoyanova

dependent on the MESSAGE. In addition, some verbs co-occur more readily with a TOPIC rather than with a MESSAGE, e.g. *explain.v* (Example 7d).

As a peripheral frame element the ADDRESSEE is often left non-overt although implied. When present, it is expressed as a prepositional phrase most frequently with the preposition ‘to’ (Example 7d). In some cases it may be realised as an indirect object (Example 7e).

- (7) a. [*North Korea*]_{SPKR} **CLAIMED** [*it had no intention of producing nuclear weapons*]_{MSG}.
- b. [*He*]_{SPKR} **SAID** [*little*]_{MSG} [*about the case*]_{TOP}.
- c. [*He*]_{SPKR} **ADDED**: [*‘Eldorado is a brave venture’*]_{MSG}.
- d. [*Doc*]_{SPKR} **EXPLAINED** [*the injuries*]_{MSG} [*to the police*]_{ADDR}.
- e. [*The agency*]_{SPKR} **WROTE** [*me*]_{ADDR} [*that you had moved*]_{MSG}.
- f. [*The letter*]_{MEDIUM} **ALLEGED** [*serious breaches of the law*]_{MSG}.

The various specific configuration of frame elements as expressed by verbs in the Statement frame are shown in Table 9.

4.3.2 Statement valence patterns

The prevalent valence patterns for verbs in the FrameNet frame Statement are shown in Table 10. The most typical ones include the canonical expression of the SPEAKER as the external NP and the MESSAGE as a subordinate clause, an object NP, or a quote.

Alternatively, the MEDIUM may occupy the position of the external argument with an implied generalised reading of the SPEAKER which is left unexpressed (Example 7f). Similarly to many of the frames describing verbs of communication, instead of the MESSAGE the TOPIC may be realised, most often as a prepositional phrase.

The patterns involving the expression of an ADDRESSEE are quite infrequent.

4.3.3 Syntactic realisation of the Statement frame in Bulgarian

The syntactic realisation of the frame element configurations in Bulgarian closely resembles that in English. The SPEAKER is usually realised as the external NP and can be a person, a group or an organisation (Example 8a, 8b). In some cases the MEDIUM can take the position of the external argument (Example 8c).

The MESSAGE is either a finite clause (Example 8a), an object NP (Example 8b) or a direct quote (Example 8f). The TOPIC rarely occurs together with the

2 Language-independent and language-specific properties

Table 9: Syntactic expression of the Statement frame elements in selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>announce</i>									
SPEAKER	44		3		5			1	53
ADDRESSEE			6					1	7
MESSAGE	8	20				24	6		58
MEDIUM	3		2						5
<i>declare</i>									
SPEAKER	58				7				65
ADDRESSEE			7						7
MESSAGE	7	32	6			17	15	7	84
<i>report</i>									
SPEAKER	54		1		19				74
ADDRESSEE			8						8
MESSAGE	19	20	2	1	1	44	2		89
MEDIUM	9		5		1			1	16
TOPIC	2		5		1				8
<i>say</i>									
MESSAGE	14	22	1	4	2	49	33		125
ADDRESSEE			8						8
SPEAKER	90	1			14				105
MEDIUM	9		10		1				20
TOPIC			10		1			1	12
<i>state</i>									
SPEAKER	38								38
ADDRESSEE			3						3
MESSAGE	3	8	2			19	13		45
MEDIUM	3		1		3				7
<i>suggest</i>									
SPEAKER	27		2		4				33
ADDRESSEE			5						5
MESSAGE	3	5		3		21	5		37
MEDIUM	4		4						8
<i>talk</i>									
SPEAKER	32		1		3				36
TOPIC	3		29	2		2			36
MESSAGE	1	3							4
<i>write</i>									
SPEAKER	42				1				43
ADDRESSEE		2	4			1			7
MESSAGE	1	5			2	10	13		31
TOPIC	1		22						23
MEDIUM	1		8						9

Table 10: FrameNet valence patterns of Statement verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{SPKR} [Clause] _{MSG}	281	<i>explain, note, declare, maintain, remark, mention, conjecture, reiterate, assert, preach, claim, attest, state, caution, write, add, allege, exclaim, say, suggest, insist, propose, announce, confirm, acknowledge, proclaim, reaffirm, report, pronounce</i>
[NP.Ext] _{SPKR} [NP.Obj] _{MSG}	191	<i>explain, note, declare, tell, conjecture, reiterate, assert, preach, claim, speak, talk, state, caution, write, add, allege, exclaim, say, suggest, propose, announce, confirm, acknowledge, refute, proclaim, reaffirm, report</i>
[NP.Ext] _{SPKR} [Quote] _{MSG}	143	<i>explain, gloat, declare, remark, observe, mention, reiterate, hazard, assert, preach, speak, attest, state, caution, write, add, allege, exclaim, say, pout, suggest, insist, propose, announce, proclaim, reaffirm, report</i>
[NP.Ext] _{SPKR} [PP] _{TOP}	83	<i>explain, gloat, preach, report, comment, remark, speak, talk, write</i>
[NP.Ext] _{MEDIUM} [Clause] _{MSG}	39	<i>note, declare, allege, say, suggest, propose, announce, confirm, acknowledge, proclaim, report, claim, state</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [NP.Obj] _{MSG}	28	<i>reiterate, declare, report, say, speak, state, suggest, propose, announce, mention</i>
[NP.Ext] _{SPKR} [PP] _{MSG}	28	<i>profess, declare, preach, say, speak, describe, insist, caution</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [Clause] _{MSG}	25	<i>add, explain, declare, allege, suggest, insist, propose, announce, mention, confirm, preach</i>
[NP.Ext] _{SPKR} [PP] _{MEDIUM} [Clause] _{MSG}	20	<i>explain, note, acknowledge, allege, claim, say, state, suggest, write, mention</i>
[NP.Ext] _{MEDIUM} [NP.Obj] _{MSG}	20	<i>explain, note, proclaim, tell, allege, reaffirm, say, state, propose, announce, mention</i>

2 Language-independent and language-specific properties

MESSAGE, and it is usually a modifier of the MESSAGE (Example 8d). The non-core ADDRESSEE is mostly optional and is realised as a prepositional complement (Example 8b).

- (8) a. [Панайотов]_{SPKR} **ДОБАВИ**, [че лидер на бъдещата партия ще
Panayotov added that leader of future-DEF party will
е Симеон]_{MSG}.
be Simeon.
'Panayotov added that Simeon will be the leader of the future party.'
- b. [Кредитните институции]_{SPKR} **ДЕКЛАРИРАХА** [пред
Credit institutions declared to
властите]_{ADDR} [нарасналите печалби]_{MSG}.
authorities-DEF increased-DEF profits.
'Credit institutions declared increased profits to the authorities.'
- c. [Неофициалните статистики за 1999 г.]_{MEDIUM} **СОЧАТ** [5000
Unofficial-DEF statistics for 1999 state 5000
посетители]_{MSG}.
visitors.
'The unofficial statistics for 1999 state 5,000 visitors.'
- d. [Тези лица]_{SPKR} **ИЗКАЗВАТ** [пред нас]_{ADDR} [неприятни
These persons state to us unpleasant
истини]_{MSG} [за смъртните ни врагове]_{TOP}.
truths about mortal-DEF our enemies.
'These people state to us unpleasant truths about our mortal enemies.'
- e. [В интервюто]_{MEDIUM} [Симеон]_{SPKR} **ОБЯВИ** [промяна на
In interview-DEF Simeon announced change of
политическата посока]_{MSG}.
political-DEF direction.
'In the interview Simeon announced a change in the political direction.'
- f. [- Тя го каза просто така]_{MSG} - **ДОБАВИ** [Джени]_{SPKR}.
- She it said just so - added Jenny.
'- She said it just like that - added Jenny.'

Table 11 shows some of the most frequent verbs in Bulgarian evoking the frame Statement. The Bulgarian examples show similar patterns to the realisation of frame elements to the examples in the English dataset.

Svetlozara Leseva & Ivelina Stoyanova

Table 12 presents the most frequent valence patterns typical of the verbs evoking the Statement frame in Bulgarian. Like in English, the most preferred realisations involve a subject **SPEAKER** and a **MESSAGE** expressed as an object NP, a clause or a quote.

4.4 Frame Telling

Definition of the Telling frame: A **SPEAKER** addresses an **ADDRESSEE** with a **MESSAGE**, which may be indirectly referred to as a **TOPIC**. Instead of (or in addition to) a **SPEAKER**, a **MEDIUM** may also be mentioned. Core frame elements: **SPEAKER**, **ADDRESSEE**, **MESSAGE**, **MEDIUM**, **TOPIC**.

The frame Telling is evoked by a small number of frequently occurring verbs such as *tell.v*, *advise.v*, *inform.v*, *notify.v*, etc. The frame inherits from Statement and its specialisation consists in the fact that it describes speech acts directed to a specific **ADDRESSEE**. As a result this frame element is promoted to core status and with most verbs (*inform.v*, *advise.v*, *confide.v*, *notify.v*) is favoured for the direct object position.

4.4.1 Syntactic realisation of the Telling frame elements

The frame elements generally have the same characteristics as the ones in the Statement frame from which they are inherited. The **SPEAKER** usually takes the position of the external NP (Example 9a). Most often the **ADDRESSEE** is expressed as an NP object (Example 9b) or in the case of *tell.v* as an indirect object NP or a PP.

The **MESSAGE** is most often realised as a prepositional phrase, a subordinate clause or a quote (Example 9b, 9c, 9a, respectively). It may also take the position of an NP object, while the **ADDRESSEE** is represented by a PP (Example 9d), a pattern which is actually favoured by the verb *confide.v*. Instead of the **MESSAGE**, its **TOPIC** may be realised as a prepositional phrase (Example 9e).

- (9) a. [*'Take your bag and go.'*]_{MSG} [*Jake*]_{SPKR} **TOLD** [*her*]_{ADDR}.
 b. [*The police*]_{SPKR} *didn't* **INFORM** [*the British Consulate*]_{ADDR} [*about his disappearance*]_{MSG}.
 c. [*We*] *have* **NOTIFIED** [*Benoit*]_{ADDR} [*that Tweed is wanted*]_{MSG}.
 d. [*She*]_{SPKR} **CONFIDED** [*her sadness*]_{MSG} [*in Beth*]_{ADDR}.
 e. [*He*]_{SPKR} *will* **ADVISE** [*you*]_{ADDR} [*on the inheritance tax*]_{TOP}.

2 Language-independent and language-specific properties

Table 11: Syntactic expression of the Statement frame elements in Bulgarian lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>обявявам/обявя</i> (announce)									
SPEAKER	17				1				18
MESSAGE		4				12	1	1	18
<i>твърдя</i> (claim)									
SPEAKER	11				1				12
MESSAGE						10	2		12
<i>коментирам</i> (comment)									
SPEAKER	8								8
MESSAGE		4				2	1	1	8
<i>добавам/добавя</i> (add)									
SPEAKER	10								10
MESSAGE						5	5		10
<i>съобщавам/съобщя</i> (announce)									
SPEAKER	10				1				11
MESSAGE		4				2	5		11
ADDRESSEE			1						1
<i>казвам/кажа</i> (say)									
SPEAKER	47				1				48
MESSAGE	1	10				18	19		48
ADDRESSEE			4						4
<i>обяснявам/обясня</i> (explain)									
SPEAKER	14				2				16
MESSAGE	1	2			1	5	6	1	16
ADDRESSEE			6						6
<i>заявявам/заявя</i> (state)									
SPEAKER	17								17
MESSAGE						10	7		17
ADDRESSEE			4						4

Table 12: FrameNet valence patterns of Statement verbs, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *добавям/добавя* (add), *заявявам/заявя* (state), *казвам/кажа* (say), *коментирам* (comment), *обявявам/обявя*, *оповестявам/оповестя*, *съобщавам/съобщия* (announce), *обяснявам*, *обясня* (explain), *отбелязвам/отбележа* (note), *пиша* (write), *повтарям/повторя* (reiterate), *посочвам/посоча* (state), *предлагам/предложа* (suggest).

Pattern	#	verbs
[NP.Ext] _{SPKR} [Clause] _{MSG}	67	добавям/добавя, заявявам/заявя, казвам/кажа, коментирам, обявявам/обявя, обяснявам/обясня, отбелязвам/отбележа, пиша, посочвам/посоча, предлагам/предложа, твърдя
[NP.Ext] _{SPKR} [Quote] _{MSG}	48	добавям/добавя, заявявам/заявя, казвам/кажа, коментирам, обявявам/обявя, обяснявам/обясня, отбелязвам/отбележа, пиша, повтарям/повторя, съобщавам/съобщия, твърдя
[NP.Ext] _{SPKR} [NP.Obj] _{MSG}	29	казвам/кажа, коментирам, обявявам/обявя, оповестявам/оповестя, повтарям/повторя, посочвам/посоча, предлагам/предложа, съобщавам/съобщия
[NP.Ext] _{SPKR} [Clause] _{MSG} [PP] _{ADDR}	9	заявявам/заявя, обяснявам/обясня, предлагам/предложа, съобщавам/съобщия
[NP.Ext] _{SPKR} [PP] _{ADDR} [Quote] _{MSG}	5	заявявам/заявя, казвам/кажа
[NP.Ext] _{SPKR} [NP.Obj] _{MSG} [PP] _{ADDR}	2	обяснявам/обясня

The various specific configurations of frame elements as expressed by verbs in the Telling frame are shown in Table 13.

4.4.2 Telling valence patterns

The prevalent valence patterns for the verbs in the FrameNet frame Telling are illustrated in Table 14. These include the prototypical expression of the **SPEAKER** as the external NP, usually with a direct object **ADDRESSEE**, which may be left implicit and/or a **MESSAGE** realised as a subordinate clause, a prepositional phrase or a quote; the **MESSAGE** may also be implicit. A **PP TOPIC** may co-occur with the **ADDRESSEE** but usually not with the **MESSAGE**.

4.4.3 Syntactic realisation and patterns in Bulgarian

In a similar manner, in Bulgarian the **SPEAKER** is realised as the external subject NP, while the **MESSAGE** takes the position of an object NP, a subordinate clause or a quote (Example 10a, 10b, 10c).

2 Language-independent and language-specific properties

Table 13: Syntactic expression of the Telling frame elements in selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>tell</i>									
SPEAKER	90	1	9		14				114
ADDRESSEE	18	59	3	1	36	1		2	120
TOPIC		3	31			4		1	39
MESSAGE	5	11	9	3	13	35	6	8	90
MEDIUM	10		2						12
<i>inform</i>									
SPEAKER	39				8				47
ADDRESSEE	8	37			2				47
MESSAGE			10		7	20	6		43
MEDIUM			3						3
TOPIC			4						4
<i>advise</i>									
SPEAKER	59		1		6				66
ADDRESSEE	8	31	1		27				67
MESSAGE		3	7			29	8		47
TOPIC			19		1				20
<i>confide</i>									
SPEAKER	45				1				46
ADDRESSEE			23		23				46
MESSAGE	1	23			4	14	4		46
MEDIUM			1						1

Table 14: FrameNet valence patterns of Telling verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [Clause] _{MSG}	53	<i>inform, advise, tell, assure, notify</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [PP] _{TOP}	30	<i>apprise, inform, advise, tell, notify</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [Clause] _{MSG}	26	<i>advise, confide, tell, assure</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [] _{MSG-DNI}	20	<i>inform, tell, assure, notify</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [PP] _{MSG}	20	<i>inform, advise, tell, notify</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [PP] _{TOP}	17	<i>advise, tell</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [NP.Obj] _{MSG}	16	<i>advise, confide, tell</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [NP.Obj] _{MSG}	16	<i>advise, confide, tell, notify</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [Quote] _{MSG}	14	<i>advise, confide, assure</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [Quote] _{MSG}	11	<i>inform, tell, assure</i>

Svetlozara Leseva & Ivelina Stoyanova

With some of the verbs in this frame, such as *казвам*, *съобщавам* (tell, let know) the ADDRESSEE assumes the position of the indirect object as the receiver to whom the message is directed (Example 10b), while with verbs such as *уведомявам* (notify, inform), *информирам*, *осведомявам* (inform) it is realised as an NP object (Example 10d); the ADDRESSEE may also be null instantiated (Example 10e).

- (10) a. []_{SPKR-DNI} *Искам да [ви]_{ADDR} СЪОБЩА [една тъжна*
 Want.1sg to you.2pl-DAT tell one sad
вест]_{MSG}.
 news.
 ‘I want to tell you some sad news.’
- b. [*Всеки българин*]_{SPKR} *ще [ти]_{ADDR} КАЖЕ [каквото е чул*
 Every Bulgarian will you.2sg-DAT tell whatever has heard
от майка си]_{MSG}.
 from mother REFL.
 ‘Every Bulgarian will tell you whatever he has heard from his mother.’
- c. [*Не са намерили Санса*]_{MSG} – *учтиво [го]_{ADDR} УВЕДОМИ*
 Not have found Sansa – politely him informed
[чичо му]_{SPKR}.
 uncle his.
 ‘They have not found Sansa – his uncle informed him politely.’
- d. []_{SPKR-DNI} *Трябва да ОСВЕДОМЯ [читателя]_{ADDR} [за*
 Need.1sg to inform reader-DEF about
тайната интрига]_{TOP}.
 secret-DEF plot.
 ‘I need to inform the reader about the secret plot.’
- e. [*Пенсионерите да избягват навалиците*]_{MSG}, *СЪВЕТВА [г-жа*
 Elderly-DEF to avoid crowds, advises Mrs
Ненова]_{SPKR} []_{ADDR}.
 Nenova.
 ‘The elderly should avoid crowds, Mrs Nenova advises.’

Table 15 presents the most frequent verbs in Bulgarian evoking the frame Telling, while Table 16 shows the typical valence patterns. The MESSAGE and the ADDRESSEE tend to co-occur syntactically, while the TOPIC is expressed more rarely.

2 Language-independent and language-specific properties

Table 15: Syntactic expression of the Telling frame elements in Bulgarian.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>уверявам/уверя</i> (assure)									
MESSAGE			1			24	6		31
ADDRESSEE		31							31
SPEAKER	31								31
<i>съобщавам/съобщя</i> (tell, let know)									
MESSAGE		3				2			5
ADDRESSEE			5						5
SPEAKER	5								5
<i>уведомявам/уведомя</i> (inform, notify)									
MESSAGE					5	5	3		13
ADDRESSEE	1	15							16
TOPIC			4						4
SPEAKER	15				1				16
<i>казвам/кажа</i> (tell)									
MESSAGE		11				15	6		32
ADDRESSEE			32						32
SPEAKER	32								32

Table 16: FrameNet valence patterns of the frame Telling, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *информирам, осведомявам/осведомя* (inform), *казвам, съобщавам/съобщя* (tell, let know), *посъветвам* (advise), *уведомявам/уведомя* (notify), *уверявам/уверя* (assure).

Pattern	#	verbs
[NP.Ext] _{SPEAKER} [Clause] _{MSG} [NP.Obj] _{ADDR}	32	<i>осведомявам/осведомя, уверявам/уверя, уведомявам/уведомя</i>
[NP.Ext] _{SPEAKER} [Clause] _{MSG} [PP] _{ADDR}	15	<i>казвам/кажа, съобщавам/съобщя</i>
[NP.Ext] _{SPEAKER} [NP.Obj] _{MSG} [PP] _{ADDR}	14	<i>казвам/кажа, съобщавам/съобщя</i>
[NP.Ext] _{SPEAKER} [NP.Obj] _{ADDR} [Quote] _{MSG}	9	<i>уверявам/уверя, уведомявам/уведомя</i>
[NP.Ext] _{SPEAKER} [PP] _{ADDR} [Quote] _{MSG}	6	<i>казвам/кажа</i>
[NP.Ext] _{SPEAKER} [NP.Obj] _{ADDR} [PP] _{TOP} [] _{MSG-INI}	5	<i>осведомявам/осведомя, уведомявам/уведомя</i>
[NP.Ext] _{SPEAKER} [Quote] _{MSG} [] _{ADDR-INI}	4	<i>информирам, посъветвам</i>

Svetlozara Leseva & Ivelina Stoyanova

4.5 Frame Judgment_communication

Definition of the frame Judgment_communication: A COMMUNICATOR communicates a judgement of an EVALUEE to an ADDRESSEE. The judgement may be positive (e.g. *praise.v*) or negative (e.g. *criticise.v*). Core frame elements: COMMUNICATOR, EXPRESSOR, REASON, MEDIUM, TOPIC, EVALUEE; Non-core: ADDRESSEE.

The frame Judgment_communication inherits from both the Statement and the Judgment frame (weak inheritance through the *Uses* frame-to-frame relation). Verbs included in this frame concern acts of speech which also convey judgement on a certain topic, the EVALUEE. The frame elaborates on the frame Statement most notably in the interpretation of the MESSAGE as a judgement on a complex state-of-affairs concerning an additional participant, represented by the frame element EVALUEE. The EVALUEE can be a person, an object, an action or any topic (Example 11a, 11b, 11f). The judgement may be positive, e.g. *praise.v*, *commend.v*, *acclaim.v*, or negative, e.g. *criticise.v*, *condemn.v*, *denounce.v*; its value is encoded by the verb. In addition, the frame element REASON denotes the argumentation for the judgement. The ADDRESSEE is a non-core frame element, reflecting the fact that the judgement regarding the EVALUEE may but need not be intended for another participant.

4.5.1 Syntactic realisation of the Judgment_communication frame elements

The frame Judgment_communication specifies the more general frame element COMMUNICATOR rather than inheriting the SPEAKER from the Statement frame. The reason for this is that the frame also includes verbs which represent communication acts that are more general or complex than speech acts, e.g. *belittle.v*, *ridicule.v*.

The COMMUNICATOR is usually realised as the external argument and can be represented by a person, a group or an organisation (Example 11a, 11b).

The EVALUEE is most often expressed in the position of the NP direct object (Example 11a, 11b, 11c), while the REASON can be a prepositional phrase headed by prepositions such as *for*, *of*, *as* (Example 11c, 11d, 11f). Instead of the REASON, a TOPIC can be present (Example 11e).

The ADDRESSEE, whenever overt, is expressed as a prepositional phrase (Example 11b).

The EXPRESSOR is rare with verbs evoking this frame and usually represents a body part or an action performed by the COMMUNICATOR in order to convey the judgment (Example 11g).

2 Language-independent and language-specific properties

- (11) a. [Frank]_{COM} **RIDICULED** [everything]_{EVAL}.
 b. [Jon]_{COM} **BELITTLED** [Madie]_{EVAL} [to her colleagues]_{ADDR}.
 c. [Georgia]_{COM} *has* **ACCUSED** [Russian troops]_{EVAL} [of backing separatists]_{REAS}.
 d. [I]_{COM} *have* **PRAISED** [her]_{EVAL} [for her work]_{REAS}.
 e. [He]_{COM} **CRITICISED** [the president]_{EVAL} [over his decision to go to war]_{TOP}.
 f. [The conservatives]_{COM} **DENOUNCED** [the proposed reforms]_{EVAL} [as an attempt to distract voters]_{REAS}.
 g. [His glance]_{EXR} **DENIGRATED** [her attempt at humour]_{EVAL}.

Table 17 shows some of the frequent verbs of the frame and the realisation of their frame elements.

4.5.2 Judgment_communication valence patterns

The valence patterns characteristic for verbs in the FrameNet frame Judgment_communication are presented in Table 18. The most common ones involve a COMMUNICATOR as the external argument, a direct object NP EVALUÉE, and an either overtly expressed or implicit REASON or much more rarely a TOPIC.

4.5.3 Syntactic realisation of the frame Judgement_communication in Bulgarian

The COMMUNICATOR is expressed as the external NP (Example 12a). The EVALUÉE can be any concrete or abstract entity, quality, property, etc., whose properties are being evaluated, and is usually realised as the direct NP object (Example 12a, 12b) or as a prepositional phrase for a limited number of verbs such as *подузвам се* (mock, ridicule) in (Example 12e).

The REASON is expressed as a prepositional phrase with a range of prepositions such as *за*, *от*, *на* (Example 12c, 12d, 12f), or more rarely as a clause (Example 12b) or a direct quote (Example 12g). In some cases the EVALUÉE and the REASON can be expressed jointly (Example 12f).

The ADDRESSEE is rarely expressed and is realised as a prepositional phrase (Example 12g).

Svetlozara Leseva & Ivelina Stoyanova

Table 17: Syntactic expression of the Judgment_communication frame elements in selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>condemn</i>									
COMMUNICATOR	105		21		11				137
EVALUÉE	32	103	1		2				138
MEDIUM	1		5						6
REASON		4	44		90				138
<i>criticize</i>									
COMMUNICATOR	88		15		47				150
ADDRESSEE			1						1
EVALUÉE	74	71			9			1	155
REASON		4	87		59				150
TOPIC		1	3						4
MEDIUM	5		3						8
<i>praise</i>									
COMMUNICATOR	50		12		18				80
EVALUÉE	27	49						4	80
REASON	3	2	38		34				77
MEDIUM			5						5
<i>ridicule</i>									
COMMUNICATOR	14		16		16				46
EVALUÉE	38	9							47
MEDIUM	1		2						3
REASON		1	13		33				47

Table 18: FrameNet valence patterns of Judgment_communication verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{COM} [NP.Obj] _{EVAL} [PP] _{REAS}	263	<i>accuse, deprecate, denigrate, censure, castigate, condemn, ridicule, commend, belittle, denounce, praise, damn, criticize, execrate, mock</i>
[NP.Ext] _{COM} [NP.Obj] _{EVAL} [] _{REAS-DNI}	138	<i>accuse, deprecate, denigrate, censure, ridicule, commend, castigate, acclaim, belittle, condemn, denounce, praise, damn, criticize</i>
[NP.Ext] _{COM} [NP.Obj] _{EVAL} [] _{REAS-INI}	25	<i>criticize, denigrate, mock, castigate, condemn, denounce</i>
[NP.Ext] _{COM} [NP.Obj] _{EVAL} [PP] _{TOP}	7	<i>slam, charge, criticize</i>

2 Language-independent and language-specific properties

- (12) a. [*Нашето посолство*]_{COM} **ОСЪДИ** [*разрушаването на*
Our embassy condemned destruction-DEF of
храма в Скопие]_{EVAL} [_{REAS-INIT}].
church-DEF in Skopje.
‘Our embassy condemned the destruction of the church in Skopje.’
- b. [_{COM-DNI}] *Не могат да* [*ме*]_{EVAL} **ОБВИНЯВАТ**, [*че съм ги*
Not can.3pl to me accuse that have them
оскърбил]_{REAS}.
offended.
‘They cannot accuse me of offending them.’
- c. [*България*]_{COM} [*ни*]_{EVAL} **ПРОКЛИНА** [*за нещастията си*]_{REAS}.
Bulgaria us condemns for misfortunes-DEF REFL.
‘Bulgaria condemns us for its misfortunes.’
- d. [_{COM-DNI}] **ОБВИНЯВАШЕ** [*ме*]_{EVAL} [*в коравосърдечие*]_{REAS}.
Accused me in cold-heartedness.
‘He/she accused me of cold-heartedness.’
- e. [*Ти*]_{COM} **ПОДИГРАВАШ** ли **СЕ** [*с мен*]_{EVAL}?
You mock QST REFL with me?
‘Are you mocking me?’
- f. [*Мускетарите*]_{COM} *се* **ПОДИГРАВАХА** [*на кривите му*
Musketeers-DEF REFL ridiculed for bow-DEF his
крака]_{EVAL+REAS}.
legs.
‘The musketeers ridiculed him for his bow legs.’
- g. [*Много е наблюдателна*]_{REAS} – **ПОХВАЛИ** [*я*]_{EVAL} [*мой*]_{COM}
–Very is.3sg observant – praised her he
[*на другите*]_{ADDR}.
to others-DEF.
‘– She is very observant – he praised her to the others.’

4.5.4 Judgment_communication valence patterns in Bulgarian

The valence patterns for the Bulgarian verbs in this frame are presented in Table 20. Similarly to English, the most typical ones involve the expression of the COMMUNICATOR and the EVALUEE and possibly the REASON; in the dataset there have not been cases of TOPIC.

Table 19: Syntactic expression of the Judgment_communication frame elements in Bulgarian.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>похвалявам/похваля</i> (praise)									
COMMUNICATOR	16								16
EVALUÉE		15			1				16
REASON			2						2
<i>обвинявам/обвиня</i> (blame)									
COMMUNICATOR	12								12
EVALUÉE		11			1				12
REASON			5	1		2			8
<i>подигравам се/подигряя се</i> (mock, ridicule)									
COMMUNICATOR	15								15
MEDIUM			1						1
EVALUÉE			13		2				15
REASON			2		1				3

4.6 Frame Questioning

Definition of the frame Questioning: A SPEAKER asking an ADDRESSEE a question, which represents the MESSAGE, calling for a reply. Core frame elements: SPEAKER, MESSAGE, ADDRESSEE, TOPIC.

4.6.1 Syntactic realisation of the Questioning frame elements

The semantic specification of the core frame elements is similar to those in the other related frames. As questioning is a purposeful action, the SPEAKER is necessarily a person or an organisation. The SPEAKER is the external argument projected as a subject NP.

The central role of the ADDRESSEE is reflected in the fact that it is frame element that is typically expressed as the direct object NP (except for *inquire.v* and some uses of *ask.v*) where it can be expressed as a prepositional complement headed by *of*.

Except for a small number of occurrences with the same verbs, i.e. *inquire.v* and *ask.v*, where it takes the direct object position, MESSAGE is typically expressed as a direct quote or an embedded question.

2 Language-independent and language-specific properties

Table 20: FrameNet valence patterns of Judgment_{communication} verbs, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *величая*, *възхвалявам/възхваля* (extol), *виня*, *обвинявам/обвиня* (blame), *гавря се* (deride), *заклеймявам/заклеймя* (condemn), *заяждам се/заям се* (criticise), *иронизирам* (ironise), *клеветя* (denigrate), *критикувам* (criticise), *кълна* (damn), *омаловажавам/омаловажа* (belittle), *осъждам/осъдя* (judge), *отричам/отрека* (denounce), *подценявам/подценя* (disparage), *подигравам се/подиграя се*, *присмивам се/присмея се* (mock, ridicule), *порицавам/порицяя* (castigate), *похвалвам/похваля*, *хваля* (commend, praise) *прославям/прославя*, *славя* (laud).

Pattern	#	verbs
[NP.Ext] _{COM} [NP.Obj] _{EVAL} [] _{REAS-DNI/INI}	57	<i>величая</i> , <i>виня</i> , <i>възхвалявам/възхваля</i> , <i>иронизирам</i> , <i>клеветя</i> , <i>критикувам</i> , <i>кълна</i> , <i>обвинявам/обвиня</i> , <i>омаловажавам/омаловажа</i> , <i>осъждам/осъдя</i> , <i>отричам/отрека</i> , <i>подценявам/подценя</i> , <i>порицавам/порицяя</i> , <i>похвалвам/похваля</i> , <i>прославям/прославя</i> , <i>славя</i> , <i>хваля</i>
[NP.Ext] _{COM} [PP] _{EVAL}	28	<i>гавря се</i> , <i>заяждам се/заям се</i> , <i>подигравам се/подиграя се</i> , <i>присмивам се/присмея се</i>
[NP.Ext] _{COM} [NP.Obj] _{EVAL} [PP] _{REAS}	12	<i>заклеймявам/заклеймя</i> , <i>иронизирам</i> , <i>обвинявам/обвиня</i> , <i>подигравам/подиграя</i> , <i>порицавам/порицяя</i> , <i>похвалвам/похваля</i> , <i>проклинам/прокълна</i>

The TOPIC is either expressed as a prepositional complement or remains implied but non-overt syntactically.

The verbs evoking the frame Questioning are divided along two lines:

- (i) whether they tend to express the MESSAGE over the TOPIC or vice versa;
- (ii) whether they tend to leave the Addressee unexpressed if it is understood from the context or not.

With respect to the first criterion, the valence patterns for the verbs in the frame clearly show that the MESSAGE and the TOPIC rarely co-occur. Out of the verbs listed in this frame, *grill.v*, *interrogate.v*, *question.v* and *quiz.v* strongly favour the TOPIC (Example 13a, 13b), with much rarer occurrences of the MESSAGE,

usually in the form of a direct quotation (Example 13c); at least in the annotated corpus the two frame elements do not co-occur with these verbs.

The remaining verbs: *ask.v*, *inquire.v*, *query.v* tend to express the content of the question, i.e. the MESSAGE rather than its subject matter, the TOPIC, but TOPICS do occur. Besides, the two frame elements can co-occur provided that the MESSAGE is not realised by a clause, compare (Example 13d and Example 13e) or a quote. With the verb *inquire.v*, the MESSAGE may be realised not only as a clause or a quote but also (though rarely) as a prepositional complement (Example 13f). In addition, both *inquire.v* and *ask.v* allow the MESSAGE to be expressed as an object NP (Example 13d, 13h). This pattern is typical of *ask.v* and rare for *inquire.v*. In such cases the ADDRESSEE is expressed as an indirect (Example 13h) or a prepositional object, see (Example 13i), which is a rephrase of (Example 13h).

As regards the second distinction, the same verbs that favour TOPICS over MESSAGES – *grill.v*, *interrogate.v*, *question.v* and *quiz.v* – show preference to expressing the ADDRESSEE as an object NP, rather than leaving it implicit (Example 13j). As shown in Table 22, they tend to realise the ADDRESSEE together with the TOPIC (expressed as a PP headed by *about*). When the MESSAGE is expressed, the ADDRESSEE is often left out.

- (13) a. [*Journalists*]_{SPKR} **GRILLED** [*Mr. Major*]_{ADDR} [*about Maastricht*]_{TOP}.
 b. [*She*]_{SPKR} **QUESTIONED** [*him*]_{ADDR} [*about his aspirations*]_{TOP}.
 c. [*I*]_{SPKR} **QUIZZED** [*him*]_{ADDR}: [*“Who are you?”*]_{MSG}.
 d. [*You*]_{SPKR} **ASK** [*many questions*]_{MSG} [*about her*]_{TOP} [_{ADDR-DNI}].
 e. [*The clerk*]_{SPKR} **INQUIRED** [_{ADDR-DNI}] [*if it would be cash*]_{MSG}.
 f. [*He*]_{SPKR} **INQUIRED** [_{ADDR-DNI}] [*as to their whereabouts*]_{MSG}.
 g. [*I*]_{SPKR} *did not* **INQUIRE** [*the reason*]_{MSG} [_{ADDR-DNI}].
 h. [*They*]_{SPKR} **ASKED** [*the newcomer*]_{ADDR} [*his name*]_{MSG}.
 i. [*They*]_{SPKR} **ASKED** [*the name*]_{MSG} [*of the newcomer*]_{ADDR}.
 j. [*They*]_{SPKR} **QUESTIONED** [*the convict*]_{ADDR} [*about the money*]_{TOP}.
 k. [*“Why not?”*]_{MSG} **QUERIED** [*she*]_{SPKR} [_{ADDR-DNI}].

Table 21 shows some of the frequent verbs of the frame and the realisation of their frame elements.

2 Language-independent and language-specific properties

Table 21: Syntactic expression of the Questioning frame elements in selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>inquire</i>									
SPEAKER	37								37
ADDRESSEE			5		32				37
MESSAGE		3	1			5	18		27
TOPIC			10						10
<i>question</i>									
SPEAKER	34		4		9				47
ADDRESSEE	13	29			5				47
MESSAGE							5		5
TOPIC			25		17				42
<i>ask</i>									
SPEAKER	68				8				76
ADDRESSEE	7	27			35				69
MESSAGE	2	7	3		5	26	18		61
TOPIC		4	8			1		1	14

4.6.2 Questioning valence patterns

The valence patterns (Table 22) show the tendency outlined above: the preference for expressing the ADDRESSEE together with the TOPIC or to leave it non-overt when the focus is on the MESSAGE (i.e. it is syntactically expressed).

Table 22: FrameNet valence patterns of Questioning verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [Quote] _{MSG}	55	<i>quiz, inquire, question, query, ask</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [PP] _{TOP}	48	<i>quiz, interrogate, question, ask, grill</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [] _{TOP-DNI/INI}	38	<i>quiz, grill, interrogate, question</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [Clause] _{MSG}	13	<i>ask</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [PP] _{TOP}	12	<i>inquire, ask</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [Clause] _{MSG}	10	<i>inquire, query, ask</i>

4.6.3 Syntactic realisation of Questioning in Bulgarian

Most of the Bulgarian counterparts are derived from the basic Questioning verb *numam* (ask) – *nonumvam*, *zanumvam* (ask), *paznumvam* (ask, question, grill),

Svetlozara Leseva & Ivelina Stoyanova

npenumтам (quizz, query). Typically, either the MESSAGE or the TOPIC is expressed (Example 14a, 14b). The two may co-occur only if the MESSAGE is nominalised, usually by means of any of a small inventory of pronouns such as *нещо* (something, anything), *нищо* (nothing, thing), *това* (this, that) or some other expressions (Example 14c). If the MESSAGE is expressed otherwise, most often as a quote or an embedded clause, the two frame elements cannot co-occur. The TOPIC is expressed as a PP headed by the prepositions *за* or *относно* (about, regarding), while the ADDRESSEE occupies the direct object position – NP.Obj (Examples 14a, 14b, 14c).

The predominant valence patterns in Bulgarian are similar (Table 24), although the data show that the ADDRESSEE co-occurs more frequently with MESSAGE (Example 14d) than in English.

- (14) a. [– *Какво мислиш?*]_{MSG} – **ПОПИТА** [*я*]_{ADDR} [*той*]_{SPKR}.
 – What think.2sg? – asked her he.
 ‘– What do you think? – he asked her.’
- b. [*Тя*]_{SPKR} [*го*]_{ADDR} **РАЗПИТВА** [_{MSG}] [*за личния му живот*]_{TOP}.
 She him inquires about personal-DEF his life.
 ‘She inquires him about his personal life.’
- c. [_{SPKR}] *Ще* [*те*]_{ADDR} **ПОПИТАМ** [*нещо*]_{MSG} [*за Арон*]_{TOP}.
 Will you ask.1sg something about Aaron.
 ‘I will ask you something about Aaron.’
- d. [*Престъпникът*]_{SPKR} **ПОПИТАЛ** [*полицая*]_{ADDR} [*дали може да си купи цигари*]_{MSG}.
 Criminal-DEF asked policeman-DEF whether can to himself buy cigarettes.
 ‘The criminal asked the policeman whether he could buy cigarettes.’

The most frequent verbs and the syntactic realisation of the frame elements of Questioning is shown in Table 23.

4.7 Frame Communication_response

Definition of the frame Communication_response: A SPEAKER communicates a reply or response, a MESSAGE, to some prior communication or action, the TRIGGER. Core frame elements: SPEAKER, MESSAGE, TRIGGER, ADDRESSEE, TOPIC.

2 Language-independent and language-specific properties

Table 23: Syntactic expression of the Questioning frame elements in Bulgarian.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>nonumвам/nonumam</i>									
(ask)									
MESSAGE		1			1	20		7	29
ADDRESSEE		34							34
TOPIC			2						2
SPEAKER	57								57
<i>zanumвам/zanumam</i>									
(ask, question)									
MESSAGE						6		14	20
ADDRESSEE		16							16
TOPIC			3						3
SPEAKER	23								23
<i>numam</i>									
(ask)									
MESSAGE	1	1			2	6		15	25
ADDRESSEE		16			3				19
TOPIC			5						5
SPEAKER	29				1				30

Table 24: FrameNet valence patterns of the frame Questioning, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *numam*, *zanumвам/занумам*, *nonumвам/nonumam* (ask, question), *интересувам се/поинтересувам се* (inquire), *разnumвам/разnumам* (question, grill, interrogate).

Pattern	#	verbs
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [Sinterrog] _{MSG}	25	<i>numam</i> , <i>занитвам/занитам</i> , <i>попитвам/попитам</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [] _{MSG-INI}	25	<i>разпитвам/разпитам</i>
[NP.Ext] _{SPKR} [Quote] _{MSG} [] _{ADDR-INI}	20	<i>numam</i> , <i>занитвам/занитам</i> , <i>попитвам/попитам</i> , <i>поинтересувам се</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [Quote] _{MSG}	14	<i>numam</i> , <i>занитвам/занитам</i> , <i>попитвам/попитам</i>
[NP.Ext] _{SPKR} [Sinterrog] _{MSG} [] _{ADDR}	13	<i>интересувам се</i> , <i>numam</i> , <i>занитвам/занитам</i> , <i>попитвам/попитам</i> , <i>полюбопитствам</i>
[NP.Ext] _{SPKR} [NP.Obj] _{ADDR} [PP] _{TOP}	9	<i>numam</i> , <i>занитвам/занитам</i> , <i>попитвам/попитам</i> , <i>разпитвам/разпитам</i>

Svetlozara Leseva & Ivelina Stoyanova

The `Communication_response` frame inherits from the frame `Communication`. It elaborates on the prototypical frame by introducing a new frame element, the `TRIGGER`, which requires a response, expressed as the `MESSAGE`.

4.7.1 Syntactic realisation of `Communication_response` frame elements

The `SPEAKER` inherits the frame element `COMMUNICATOR` which exhibits the same characteristics and behaviour as in the other frames in the domain, and is realised most often as an external NP.

The `TRIGGER` is the prior communication or action to which a response is given. It can be implicit, or overtly expressed either as an NP object or as a prepositional complement (Examples 15a, 15a).

The `MESSAGE` is not necessarily expressed when the `TRIGGER` is present (Examples 15a, 15b). When the `MESSAGE` is realised, it predominantly takes the form of an embedded clause (Example 15c) or a direct quote (Example 15d).

Although rarely, the `TRIGGER` and the `MESSAGE` may co-occur (Example 15g).

The `ADDRESSEE` is the person to whom the response is directed. When expressed, it occurs as a prepositional phrase introduced by the preposition ‘to’ (Example 15e) or as an indirect object (Example 15f).

The `TOPIC` is possible but rare with verbs from this frame.

- (15) a. [*Sue*]_{COM} **ANSWERED** [*the question*]_{TRIG}.
 b. [*The US*]_{COM} *has not* **RESPONDED** [*to the offer*]_{TRIG}.
 c. [*Blanche*]_{COM} **RESPONDED** [*that the police were talking to everyone*]_{MSG}.
 d. [*‘Does it matter?’*]_{MSG} [*she*]_{COM} **COUNTERED** *defeatedly*.
 e. [*Sue*]_{COM} **RESPONDED** [*to Bob*]_{ADDR} *immediately*.
 f. [*The senator*]_{COM} *took the floor to* **ANSWER** [*critics of the deal*]_{ADDR}.
 g. [*‘Does it matter?’*]_{MSG} **REPLIED** [*she*]_{COM} [*to his question*]_{TRIG}.

Table 25 shows the syntactic realisations of verbs evoking the frame `Communication_response`.

4.7.2 `Communication_response` valence patterns

Table 26 illustrates the valence patterns that characterise the verbs in the frame `Communication_response`. The most frequent pattern have the `MESSAGE` realised as a direct quote, followed by the pattern with an embedded clause or a PP. The

2 Language-independent and language-specific properties

Table 25: Syntactic expression of the Communication_response frame elements for selected FrameNet lexical units.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>answer</i>									
SPEAKER	31		1		2				34
ADDRESSEE		1			30			3	34
MESSAGE		2	4		21	2	5		34
TRIGGER	3	11			18	1			33
<i>reply</i>									
SPEAKER	69								69
ADDRESSEE			4		65				69
MESSAGE			4		21	11	27	6	69
TRIGGER			13		56				69
<i>respond</i>									
SPEAKER	25								25
ADDRESSEE			1		23				24
MESSAGE			3	1	4	4	13		25
TRIGGER			1		24				25
<i>retort</i>									
SPEAKER	46								46
ADDRESSEE			1		45				46
MESSAGE		2	1		2	21	23		49
TRIGGER			1		45				46

TRIGGER is expressed in fewer instances and in such cases the ADDRESSEE and the MESSAGE remain non-overt.

Table 26: FrameNet valence patterns of Communication_response verbs, their frequency in the FrameNet corpus and the verbs they appear with.

Pattern	#	verbs
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [Quote] _{MSG} [] _{TRIG-DNI}	83	<i>answer, rejoin, counter, reply, respond, retort</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [Clause] _{MSG} [] _{TRIG-DNI}	34	<i>answer, rejoin, counter, reply, respond, retort</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [PP] _{MSG} [] _{TRIG-DNI}	14	<i>answer, counter, reply, respond</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [] _{MSG-INI} [PP] _{TRIG}	10	<i>reply</i>
[NP.Ext] _{SPKR} [] _{ADDR-DNI} [] _{MSG-INI} [NP.Obj] _{TRIG}	7	<i>answer</i>

Svetlozara Leseva & Ivelina Stoyanova

4.7.3 Syntactic realisation of **Communication_response** frame in Bulgarian

In Bulgarian the syntactic realisation of the frame is similar to English. The **MESSAGE** most often appears as an embedded clause (Example 16a) or as a direct quote (Example 16b), and in some cases as a direct object (Example 16d) or a prepositional phrase (Example 16c). The **TRIGGER** is realised as a prepositional phrase (Example 16e).

- (16) a. [*Той*]_{SPKR} [*ми*]_{ADDRESSEE} **ОТГОВОРИ**, [*че няма да отиде*]_{MSG}.
 He me answered that not to go.
 ‘He answered me that he won’t go.’
- b. [*Студентът*]_{SPKR} **ОТГОВОРИЛ**: [*– Професоре, забравих!*]_{MSG}
 Student-DEF responded: – Professor, forgot.1sg!
 ‘The student responded: – Professor, I forgot!’
- c. [*Мнозинството*]_{SPKR} **ОТВРЪЩА** [*с надменни приказки*]_{MSG}
 Majority-DEF answers with arrogant words
 [*за своята безалтернативност*]_{TOPIC}.
 about their-REFL lack of prospects].
 ‘The majority answers with arrogant words about their lack of prospects.’
- d. [*Той*]_{SPKR} *не* **ОТГОВОРИ** [*нищо*]_{MSG}.
 He not responded nothing.
 ‘He did not respond anything.’
- e. [*На този въпрос*]_{TRIG} *ще* **ОТГОВОРИ**
 To this question will answer
 [*министър-председателят*]_{SPKR}.
 prime minister-DEF.
 ‘The prime minister will answer this question.’

The most frequent verbs evoking the frame **Communication_response** and the realisation of their frame elements are shown in Table 27. The associated valence patterns are presented in Table 28.

5 Conclusions

In this paper we have discussed the universal features of the conceptual description of verbs which is transferable across languages. We illustrate our analysis

2 Language-independent and language-specific properties

Table 27: Syntactic expression of the Communication_response frame elements in Bulgarian.

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Quote	Other	Total
<i>отвърщам/отвърна</i> (reply)									
TRIGGER			2		4				6
MESSAGE		2			5		23	5	35
ADDRESSEE			14		21				35
MANNER			2					3	5
SPEAKER	35								35
<i>отговарям/отговоря</i> (answer, reply)									
TRIGGER			15		15				30
MESSAGE		3			26		21	16	66
ADDRESSEE			23		43				66
MEDIUM			7						7
MANNER			2					12	14
SPEAKER	66								66

Table 28: FrameNet valence patterns of the frame Communication_response, their frequency in the Bulgarian dataset and the verbs they appear with. English translation equivalents: *контрирам* (counter), *отвърщам/отвърна*, *отговарям/отговоря* (answer, reply, counter, retort).

Pattern	#	verbs
[NP.Ext] _{SPKR} [Quote] _{MSG} [] _{ADDR-DNI}	22	<i>контрирам, отвърщам/отвърна, отговарям/отговоря</i>
[NP.Ext] _{SPKR} [PP] _{ADDR} [Quote] _{MSG}	12	<i>отвърщам/отвърна, отговарям/отговоря</i>
[NP.Ext] _{SPKR} [Clause] _{MSG} [] _{ADDR-DNI}	10	<i>отвърщам/отвърна, отговарям/отговоря</i>
[NP.Ext] _{SPKR} [Clause] _{MSG} [PP] _{ADDR}	9	<i>отвърщам/отвърна, отговарям/отговоря</i>
[NP.Ext] _{SPKR} [PP] _{TRIGGER} [] _{ADDR-DNI} [] _{MSG-INI}	8	<i>отвърщам/отвърна, отговарям/отговоря</i>

Svetlozara Leseva & Ivelina Stoyanova

with examples from the class of verbs of communication with a view to their use in English and Bulgarian.

The universality of the semantic relations of inheritance (from a more generalised to a more specialised entity) underlies the hierarchical organisation of both the FrameNet frames and the WordNet synsets. The configuration of frame elements describing the behaviour of verbs evoking a particular frame are also language-independent, as well as the semantic restrictions determining their selection. Moreover, we have shown that the principles of syntactic realisation of the frame elements as represented by the generalised valence patterns are also valid to a large degree across different languages. For Bulgarian and English we have established substantial correspondence in both the valence patterns and the syntactic categories and grammatical functions by which frame elements are expressed.

Further, we have outlined some basic language-specific properties of the syntactic realisation of semantic frames and their corresponding frame elements. In some cases the two languages give different preference to the overt expression of particular frame elements. For example, the TOPIC is more frequent in English and rarely expressed with Bulgarian communication verbs (e.g., evoking the frames *Statement* and *Communication_manner*). We also observe differences in the syntactic realisation of particular frame elements due to the distinct syntactic properties of the two languages. For example, Bulgarian lacks infinitives and *-ing* clauses, so clausal complements expressing the frame element MESSAGE are finite clauses. Differences at the syntactic level between Bulgarian and English are also found between verbs considered as translation equivalents (belonging to corresponding synsets in Bulgarian and English). For example, with the verb *ridicule* (evoking the frame *Judgment_communication*) the EVALUEE is expressed predominantly as a direct object, while the Bulgarian verb *нодузпavam ce* realises it as an indirect object due to the fact that reflexive verbs do not take a direct object.

The analysis confirms the assumption that a large part of a verb's semantic valency and syntactic behaviour is predictable from its lexical meaning and the semantic class it belongs to. The various semantic classifications of verbs focus on different semantic and/or syntactic properties, relying mostly on theoretical analysis or expert intuition rather than on authentic corpus data. A study based on corpus analysis and statistical observations on the frequency of valence patterns could provide more reliable evidence for the behaviour of verbs, in particular in view of cross-linguistic studies. Moreover, this will confirm the validity of the cross-linguistic analysis and the universality of semantic and syntactic features.

2 Language-independent and language-specific properties

In our work on describing the conceptual and syntactic properties of Bulgarian verbs, we have found the applicability of the conceptual description encoded in the FrameNet frames to be to a great extent language-independent and transferable cross-linguistically, even if in some cases adjustments may be necessary. Given the fact that the alignment between equivalent senses in the wordnets developed for different languages is ensured by means of shared identification numbers with the original Princeton WordNet, the conceptual information from FrameNet is mappable across languages via WordNet¹⁵.

Abbreviations

Addr	ADDRESSEE		instantiation
Auth	AUTHOR	Manr	MANNER
BulEnAC	Bulgarian-English	Med	MEDIUM
	Sentence- and Clause-	Msg	MESSAGE
	Aligned Corpus	N or n	Noun
BulSemCor	Semantically annotated	NP	Noun phrase
	corpus for Bulgarian	PP	Prepositional phrase
CNI	Constructional null	PWN	Princeton WordNet
	instantiation	Reas	REASON
Com	COMMUNICATOR	SemCor	Semantically annotated
Cont	CONTENT		corpus for English
DNI	Definite null	Spkr	SPEAKER
	instantiation	Top	TOPIC
Eval	EVALUEE	Trig	TRIGGER
Exr	EXPRESSOR	V or v	Verb
INI	Indefinite null		

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

¹⁵For a list of existing wordnets in the world, see <http://globalwordnet.org/resources/wordnets-in-the-world/>.

Svetlozara Leseva & Ivelina Stoyanova

References

- Agirre, Eneko & David Martínez. 2002. Integrating selectional preferences in WordNet. *CoRR* cs.CL/0204027. <https://arxiv.org/abs/cs/0204027>.
- Baker, Collin F. 2008. FrameNet, present and future. In Jonathan Webster, Nancy Ide & Alex Chengyu Fang (eds.), *Proceedings of the First International Conference on Global Interoperability for Language Resources*. Hong Kong: City University.
- Baker, Collin F. & Christiane Fellbaum. 2009. WordNet and FrameNet as complementary resources for annotation. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, 125–129. DOI: 10.3115/1698381.1698402.
- Boas, Hans C. 2010. The syntax-lexicon continuum in construction grammar: A case study of English communication verbs. *Belgian Journal of Linguistics* 24. 54–82.
- Burchardt, Aljoscha, Katrin Erk & Anette Frank. 2005. A WordNet detour to FrameNet. In Bernhard Fisseni, Hans-Christian Schmitz, Bernhard Schröder & Petra Wagner (eds.), *Sprachtechnologie, mobile kommunikation und linguistische ressourcen*, vol. 8 (Computer Studies in Language and Speech), 408–421. Frankfurt: Lang.
- Burchardt, Aljoscha & Marco Pennacchiotti. 2008. FATE: A FrameNet-annotated corpus for textual entailment. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/pdf/143_paper.pdf.
- Das, Dipanjan, Desai Chen, André F. T. Martins, Nathan Schneider & Noah A. Smith. 2014. Frame-semantic parsing. *Computational Linguistics* 40(1). 9–56.
- de Lacalle, Maddalen Lopez, Egoitz Laparra & German Rigau. 2014. Predicate matrix: Extending SemLink through WordNet mappings. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 903–909. Reykjavik, Iceland: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/589_Paper.pdf.
- Di Fabio, Andrea, Simone Conia & 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 (EMNLP-IJCNLP)*, 627–637. Hong Kong, China: Association for Computational Linguistics. DOI: 10.18653/v1/D19-1058. <https://aclanthology.org/D19-1058>.

2 Language-independent and language-specific properties

- Fellbaum, Christiane. 1999a. The organization of verbs and verb concepts in a semantic net. In Patrick Saint-Dizier (ed.), *Predicative forms in natural language and in lexical knowledge bases*, vol. 6 (Text, Speech and Language Technology), 93–110. Dordrecht: Springer.
- Fellbaum, Christiane (ed.). 1999b. *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Fillmore, Charles J., Collin F. Baker & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the Conference COLINGACL '98, Montreal, Canada*, 86–90.
- Gilardi, Luca & Collin F. Baker. 2018. Learning to align across languages: Toward multilingual FrameNet. In Tiago T. Torrent (ed.), *Proceedings of the International FrameNet Workshop 2018: Multilingual FrameNets and Constructicons*, 13–22.
- Horak, Ales, Piek Vossen & Adam Rambousek. 2008. The development of a complex-structured lexicon based on WordNet. In Attila Tanács, Dóra Csendes, Veronika Vincze, Christiane Fellbaum & Piek Vossen (eds.), *Proceedings of the 4th International Global WordNet Conference (GWC 2008)*, 200–208.
- Jackendoff, Ray. 1990. *Semantic structures*. Cambridge: MIT Press.
- Johnson, Christopher R., Charles J. Fillmore, Esther J. Wood, Margaret Urban, Miriam R. L. Petruck, Collin F. Baker & Charles J. Fillmore. 2001. *The FrameNet project: Tools for lexicon building*. <https://citeseerx.ist.psu.edu/pdf/0ece390b6f4e6b38c5733248992ff73f846d91aa>.
- Kipper, Karin, Anna Korhonen, Neville Ryant & Martha Palmer. 2008. Language resources and evaluation. *Communications. ACM* 42(1). 21–40.
- Kipper-Schuler, Karin. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania. (Doctoral dissertation).
- Koeva, Svetla. 2010. Bulgarian WordNet – current state, applications and prospects. In *Bulgarian-American dialogues: Proceedings of the Eight Joint Meeting of Bulgarian and North American Scholars, Varna, Bulgaria, June 13-15, 2008 and of the Seven Joint Meeting of Bulgarian and North American Scholars, Columbus, Ohio, USA, October 9-12, 2003*, 120–132. Sofia: Prof. M. Drinov Academic Publishing House.
- Koeva, Svetla. 2021a. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla. 2021b. Towards a typological analysis of complements in Bulgarian. In *Proceedings of the Annual International Conference of the Institute of Bulgarian Language Prof. Lubomir Andreychin (Sofia, 2021)*, vol. 2, 13–27.

Svetlozara Leseva & Ivelina Stoyanova

- Koeva, Svetla, Svetlozara Leseva, Borislav Rizov, Ekaterina Tarpomanova, Tsvetana Dimitrova, Hristina Kukova & Maria Todorova. 2011. Design and development of the Bulgarian sense-annotated corpus. In María Luisa Carrió Pastor & Miguel Ángel Candel Mora (eds.), *Proceedings of the III international congress of corpus linguistics. Information and communications technologies: Present and future in corpus analysis*, 143–150.
- Koeva, Svetla, Svetlozara Leseva & Maria Todorova. 2006. Bulgarian sense tagged corpus. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, 79–86.
- Koeva, Svetla, Borislav Rizov, Ekaterina Tarpomanova, Tsvetana Dimitrova, Rositsa Dekova, Ivelina Stoyanova, Svetlozara Leseva, Hristina Kukova & Angel Genov. 2012. Bulgarian-English sentence- and clause-aligned corpus. In *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*, 51–62. Lisboa: Colibri.
- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Rositsa Dekova, Tsvetana Dimitrova & Ekaterina Tarpomanova. 2012. The Bulgarian national corpus: Theory and practice in corpus design. *Journal of Language Modelling* 0(1). 65–110. DOI: 10.15398/jlm.v0i1.33.
- Landes, Shari, Claudia Leacock & Randee Tengi. 1998. Building semantic concordances. In Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*. Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/7287.003.0015.
- Leseva, Svetlozara & 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*, 229–238.
- Leseva, Svetlozara & Ivelina Stoyanova. 2020. Beyond lexical and semantic resources: Linking WordNet with FrameNet and enhancing synsets with conceptual frames. In Svetla Koeva (ed.), *Towards a semantic network enriched with a variety of semantic relations*, 21–48. Sofia: Prof. Marin Drinov Academic Publishing House of the Bulgarian Academy of Sciences.
- Leseva, Svetlozara, Ivelina Stoyanova & Maria Todorova. 2018. Classifying verbs in WordNet by harnessing semantic resources. In *Proceedings of CLIB 2018, Sofia, Bulgaria*.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Levin, Beth, Grace Song & Beryl T. S. Atkins. 1997. Making sense of corpus data: A case study of verbs of sound. *International Journal of Corpus Linguistics* 2. 23–64.

2 Language-independent and language-specific properties

- Litkowski, Ken. 2014. *The FrameNet frame element taxonomy*. <http://www.clres.com/online-papers/FETaxonomy.pdf>.
- Lopatková, Markéta, Václava Kettnerová, Eduard Bejček, Anna Vernerová & Zdeněk Žabokrtský. 2016. *Valenční slovník českých sloves VALLEX*. Praha: Karolinum. 698.
- McCrae, John P., Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka & Luis Morgado Da Costa. 2021. The GlobalWordNet formats: Updates for 2020. In Piek Vossen & Christiane Fellbaum (eds.), *Proceedings of the 11th Global Wordnet Conference*, 91–99. University of South Africa (UNISA): Global Wordnet Association. <https://aclanthology.org/2021.gwc-1.11>.
- Miller, George A. 1995. WordNet: A lexical database for English. *Commun. ACM* 38(11). 39–41.
- Miller, George A., Martin Chodorow, Shari Landes, Claudia Leacock & Robert G. Thomas. 1994. Using a semantic concordance for sense identification. In *Proceedings of a Workshop on Human Language Technology, Plainsboro, New Jersey, 1993*. <https://aclanthology.org/H93-1061>.
- Miller, George A., Claudia Leacock, Randee Tengi & Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of a Workshop on Human Language Technology, Plainsboro, New Jersey, 1993*. <https://aclanthology.org/H93-1061>.
- Navigli, Roberto & Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In Jan Hajič, Sandra Carberry, Stephen Clark & Joakim Nivre (eds.), *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*, 216–225. Uppsala, Sweden: Association for Computational Linguistics. <https://aclanthology.org/P10-1023>.
- Nitsolova, Ruselina. 2008. Problems of the complex sentences with complements in the Bulgarian language (in Bulgarian). *Yuzhnoslovenski filolog* 64. 261–272.
- Palmer, Martha. 2009. SemLink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*. 9–15.
- Palmer, Martha, Claire Bonial & 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*, 13–17. Association for Computational Linguistics.
- Pedersen, Bolette, Sanni Nimb, Anders Søgaard, Mareike Hartmann & Sussi Olsen. 2018. A Danish FrameNet lexicon and an annotated corpus used for training and evaluating a semantic frame classifier. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA). <https://aclanthology.org/L18-1378>.

Svetlozara Leseva & Ivelina Stoyanova

- Penchev, Yordan. 1998. *Syntax of the modern literary Bulgarian language (in Bulgarian)*. Asenovgrad: Vechernik Publishing House.
- Petruck, Miriam R. 2019. Meaning representation of null instantiated semantic roles in FrameNet. In *Proceedings of the First International Workshop on Designing Meaning Representations*, 121–127. Association for Computational Linguistics.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin. F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.
- Schneider, Nathan, Behrang Mohit, Kemal Oflazer & Noah A. Smith. 2012. Coarse lexical semantic annotation with supersenses: An Arabic case study. In Haizhou Li, Chin-Yew Lin, Miles Osborne, Gary Geunbae Lee & Jong C. Park (eds.), *Proceedings of the 50th annual meeting of the Association for Computational Linguistics*, vol. 2: Short Papers, 253–258. Association for Computational Linguistics. <https://aclanthology.org/P12-2050>.
- Shi, Lei & Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing. CICLing 2005. Lecture notes in computer science*, vol. 3406, 100–111. Berlin, Heidelberg: Springer.
- Tisheva, Yovka. 2000. *Models for the interpretation of the complex sentence in the modern Bulgarian language (in Bulgarian)*. Sofia: Sema RSH.
- Tisheva, Yovka. 2004. Structural models of the interrogative sentences in Bulgarian (in Bulgarian). *Annual of Sofia University* 93-95. 5–67.
- Tonelli, Sara & Daniele Pighin. 2009. New features for FrameNet – WordNet mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL'09), Boulder, USA*.
- Urban, Margaret & Josef Ruppenhofer. 2001. Shouting and screaming: Manner and noise verbs in communication. *Literary and Linguistic Computing* 16. 77–97.
- Urešová, Zdenka, Eva Fučíková, Eva Hajičová & 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*, 10–19.
- Urešová, Zdenka, Eva Fučíková, Eva Hajičová & Jan Hajič. 2020b. Syntactic-semantic classes of context-sensitive synonyms based on a bilingual corpus. In Zygmunt Vetulani, Patrick Paroubek & Marek Kubis (eds.), *Human Language Technology. Challenges for Computer Science and Linguistics. Proceedings of the 7th Language and Technology Conference, LTC 2015, Poznań, Poland*, 242–255. Springer International Publishing.

2 *Language-independent and language-specific properties*

- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Vossen, Piek. 2004. EuroWordnet: A multilingual database of autonomous and language-specific wordnets connected via an inter-lingual index. *International Journal of Lexicography* 17(2). 161–173. DOI: [10.1093/ijl/17.2.161](https://doi.org/10.1093/ijl/17.2.161).
- Wierzbicka, Anna. 1987. *English speech act verbs: A semantic dictionary*. Sydney: Sydney Academic Press.

Chapter 3

The complex conceptual structure of verbs of change

Ivelina Stoyanova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

The study focuses on verbs of change and their description in two large lexical-semantic resources (WordNet and FrameNet) with a view to Bulgarian. The synonym sets (or synsets) encoding verbs of change in WordNet and their assigned semantic frames from FrameNet are studied in order to obtain a hierarchical organisation of causative and non-causative (inchoative) semantic frames aiming at a classification of the verbs with respect to their semantic and syntactic features. The main objective is to uncover the systemic semantic relations in each of the resources and employ them for the purpose of the comprehensive conceptual description of verbs of change. The work can contribute to the enrichment of WordNet with conceptual information and can support the study of semantic specialisation of verb meanings.

1 Introduction

Research in the field of conceptual description of verbs develops in several directions. On the one hand, there are theoretical studies on the classification of predicates and their argument structure, and on the other hand, applied works on the compilation of language resources and computer applications for semantic and syntactic analysis, which in turn contributes to developing methods and applications for language understanding and generation, machine translation, etc.

The observations are focused on the verbs in Princeton WordNet (Fellbaum 1999), as well as in the Bulgarian WordNet (Koeva 2010b, 2021), and the corresponding semantic frames from FrameNet (Fillmore et al. 1998) that describe



Ivelina Stoyanova. 2024. The complex conceptual structure of verbs of change. In Svetla Koeva (ed.), *Universality of semantic frames and language specific Bulgarian data*, 129–170. Berlin: Language Science Press. DOI: ??

Ivelina Stoyanova

them, along with the semantic relations between the frames and their frame elements.

The main objective of this study is to offer a description framework and based on that, to propose a shallow classification within the semantic class of verbs of change using a set of lexical, semantic and syntactic features. The description of verbs of change is based on several key semantic features – type of change (causal or internalised), scalability of the property of change (quantised or non-quantised change), and the frame elements of the frames describing the change. The study relies on previous well-established semantic classifications of verbs, and the contribution is in offering a classification scheme that covers the verbs in WordNet, exploring its relational structure and semantic description. The classification offered is based on FrameNet frames which group together verbs with similar conceptual properties and syntactic behaviour.

The analysis also aims at defining the class of change predicates within established classifications of predicates, in particular the classification of Levin (1993) and other classifications stemming from it, by discussing some key features determining the semantic and syntactic properties of verbs. One relevant aspect of the analysis of verbs of change is the relation of causativity between pairs of senses (synonym sets, or synsets in WordNet's structure) denoting a causative and a non-causative/inchoative (internal) change, and between the frames in FrameNet that describe the relevant semantics. Causative and inchoative verbs and the respective frames they evoke exhibit similarities in the definitions and the correspondence between the core frame elements involved, their semantic type and the restrictions on their realisation. The typology of the complex semantics of causative and inchoative verbs outlines the specific features of the semantic classes of the verbs of change and sheds light on their syntactic expression. The analysis also attempts to point out the types of change and the aspects of the lexical meaning of the verbs that block the inchoative or the causative interpretation.

Another aspect of the analysis focuses on the distinction between quantised and non-quantised change as an inherent feature of the verb's semantic interpretation. Verbs of non-quantised change (such as *die*, *be born*, *fall*, *arrive*, and many verbs of inherently directed motion) are telic, entailing a specific terminal state. On the other hand, verbs of quantised change (such as *bend*, *boil*, *freeze*) can have either a telic or an atelic reading depending on whether the THEME is explicitly presented as incremental. The distinction between quantised and non-quantised change is important for the analysis of the syntactic behaviour of verbs both in terms of their grammatical features and the realisation of their frame elements.

3 *The complex conceptual structure of verbs of change*

The result of the work is a system of verb classes evoking a system of frames, related to each other by means of frame-to-frame relations, that describe the semantics of the verbs of change along with the relevant aspects of the change involved and the specifics of their conceptual and syntactic structure. The set of semantic frames from FrameNet with their frame-to-frame relations provide a classification scheme which is applied for the purpose to classify the verbs of change in WordNet and the Bulgarian WordNet. Although the semantic frames and their description with the relevant frame elements are adopted from FrameNet, they are additionally organised and aligned in order to reflect the structure of the semantic class of verbs of change as represented in WordNet with respect to their lexical and semantic properties. Further, the syntactic valence patterns associated with these verbs are discussed briefly.

The structure of the paper is as follows. Section 2 shows an overview of relevant theoretical studies focused on verb semantics and their conceptual description. Section 3 briefly presents the resources employed in the study – WordNet (Princeton WordNet and Bulgarian WordNet) and FrameNet, as well as the sources of usage examples – the SemCor and BulSemCor semantically annotated corpora. Section 4 focuses on some of the key properties relevant for the description and the classification of the verbs of change – telicity (Section 4.1), causativity (Section 4.2), and the hierarchy of frame elements in their semantic frames (Section 4.3). Section 5 discusses an approach towards the semantic classification of the verbs of change with a view to their semantic properties and syntactic realisation. Finally, some conclusions are drawn in Section 6.

2 Related works

The objective of the literature review is two-fold: on the one hand, to outline the place of the class of verbs of change within known classifications of verb predicates, and on the other hand, to summarise the discussion on the set of lexical, semantic and syntactic properties that determine their realisation.

The classification of predicates and the semantic relations between predicates and their arguments have been studied within various theoretical approaches, based in general on their syntactic properties and behavior (Levin 1993, Pinker 1989, among others), the thematic structure (Van Valin & LaPolla 1997) or the semantics of frames (Fillmore 1982). In particular, verbal argument realisation behavior within and across languages has been studied actively in the last two decades (Rappaport Hovav & Levin 2010, Boas 2010, Beavers & Koontz-Garboden 2012, Levin 2015, Dux 2020, among others).

Ivelina Stoyanova

Although semantic roles and their hierarchies have been researched and applied extensively for the purposes of lexical semantic representation (most prominently, Fillmore 1968, Jackendoff 1972, among others), the role-based approaches for representing verb syntactic behaviour have shown many shortcomings discussed in a number of studies (Dowty 1991: 553–559, Levin & Rappaport Hovav 2005: 38–44, Fillmore 2012: 711, Dux 2020: 28–29).

This gave rise to event-structural approaches to studying the syntactic behaviour of verbs based on the situations they describe, motivated by both the verb's lexical meaning and its realisation in terms of grammatical features and valence combinations.

With respect to analysing the situations described by the verbs, as a point of departure, the present study adopts Vendler's aspectual classes of activities, achievements, accomplishments and states (Vendler 1957, 1967). Vendler based this taxonomy on the analysis of English verbs but its validity has been confirmed across many different languages.

Dowty (1979: 37–132) relates word meaning to Vendler's aspectual classes by means of lexical decompositions which aim to reflect the logical structures of sentences. However, the classes defined by decompositions do not translate uniformly into the classes defined in aspectual terms (Rappaport Hovav 2008: 16–20). The decompositions show that some verbs can be classified in more than one aspectual class depending on the use, for example some verbs can be both activities and accomplishments (e.g. *walk/walk to the store*) and both states and achievements (e.g. the ambiguity of many mental state verbs like *recognize*, *understand*, *know*). As a consequence, the aspectual classes are considered at the VP level rather than at the lexical level, which means that the aspectual properties are expressed in a complex lexical, morpho-syntactic and valence-related way by the verb and its arguments.

These approaches to representing the structure of the situations described by the verbs are taken into account in the present study in order to outline the place of the class of verbs of change in comparison to other verb classes, with respect to their lexical, semantic and grammatical properties determining their behaviour. We also consider the logical structure of the predicates under observation in connection with the main participants in the described situation as represented by the configuration of frame elements in the semantic frame evoked by the verb.

Fillmore (1970) focuses on verbs of change and their semantic and syntactic behaviour by analysing extensively the use of verbs like *hit* and *break*. He discusses three uses of the verb *break* (see Example 1) and only two use cases of the verb *hit* (see Example 2). While senses in examples 1b and 1c relate to 2a and 2b,

3 The complex conceptual structure of verbs of change

respectively, the verb *hit* does not have an intransitive use like the verb *break* in Example 1a.

- (1) a. *The stick broke.*
 b. *John broke the stick (with a rock).*
 c. *The rock broke the stick.*
 d. * *The rock broke the stick with a hammer.*
- (2) a. *John hit the tree (with a rock).*
 b. *A rock hit the tree.*
 c. * *A rock hit the tree with a stick.*

Further, Fillmore (1970: 131–132) states that while *break* is a change of state predicate, the verb *hit* does not necessarily involve change, and classifies it as a surface contact predicate (see Example 3).

- (3) a. *I hit the window with a hammer; nothing happened to the window.*
 b. * *I broke the window with a hammer; nothing happened to the window.*

The two classes demonstrate different syntactic behaviour, in particular with respect to derived adjectives – passive construction alternation and the possessed body part as object or as a place alternation, which is valid only for the verb *hit*. In summary, Fillmore (1970: 135–137) outlines the properties of change of state predicates in contrast to surface contact predicates, or more generally, other classes of predicates that although implying change (as a consequence) do not necessarily involve change as part of their lexical meaning.

Levin (1993) offers a comprehensive classification of English verbs based on their syntactic properties and their participation in particular argument alternations. Levin continues the line of analysis regarding *break* and *hit* by considering four classes of verbs and offering a number of tests to distinguish pure change of state predicates:

- (a) Break Verbs (*break, crack, rip, shatter, snap*);
- (b) Cut Verbs (*cut, hack, saw, scratch, slash*);
- (c) Touch Verbs (*pat, stroke, tickle, touch*); and
- (d) Hit Verbs (*bash, hit, kick, pound, tap, whack*).

Ivelina Stoyanova

One test offered by Levin (1993) is of (zero-)derivative nouns, which in the case of *break* and *cut* refer to the result of the action, while the zero-derived nominals from *hit* and *touch* do not allow this interpretation and refer only to the action itself. Thus follows the conclusion that *cut* and *break* are both verbs of causing a change of state since the nominalisation shows a result (terminal state) included in its semantics. And on the other hand, since *hit* or *touch* are not change of state verbs (although they might be associated with a change as a consequence), these verbs are not found in the causative/inchoative alternation.

Rappaport Hovav (2008: 16–20) outlines the aspectually relevant properties encoded in the meanings of verbs and the grammatical reflexes of these properties. The author looks at the relationship between the aspectual property of telicity and the notion of scale which in the event structure can be interpreted as a measure, incremental theme, quantity criterion, etc. The conclusion drawn by Rappaport Hovav (2008) is that verbs lexically encoding a scalar change cannot be classified once and for all either as activities, accomplishments or achievements.

Rappaport Hovav & Levin (2005: 278–279) argue that argument expression is not determined purely by the aspectual properties of the sentence but rather is also motivated by certain lexical features of the verb. In particular, the authors analyse verbs of change to show that certain alternations can only be explained by the lexicalised property¹ of the verbs, namely change of state, and whether it can be incremental or instantaneous.

The study of argument structure of verbs and the properties of their arguments suggests that argument structures might be derivable to a large extent from the meaning of words and by combining similar verbs into classes with similar syntactic behaviour (Levin 1993: 1–3, Pinker 1989: 4–7, Chomsky 1986: 11–12) which necessitates the representations of linguistically relevant aspects of word meaning and the formulation of principles to deduce syntactic properties from word meaning. To this end, here we employ semantic frames which are assigned to synsets, allowing us to study both semantically related words (via WordNet relations) and their corresponding conceptual descriptions (from FrameNet frames and the relations between them).

¹Here we assume the following definition of the term *to lexicalise*: ‘to represent (a set of semantic features) by a lexical item’; in particular, a lexicalised property, as used by Rappaport Hovav & Levin (2005: 273), is a semantic property that has become an integral part of the verb’s lexical meaning as opposed to being expressed by morphological means or by the valency configuration of the verb. For example, the semantic property ‘temperature’ as the property of change is lexicalised in verbs such as *warm* or *freeze*.

3 The complex conceptual structure of verbs of change

With respect to the classification of verbs of change Dowty (1979: 66–78) considers several semantic classes and places them within the two aspectual classes of achievements and accomplishments.

Achievements (may be coextensive with inchoative): Change of physical state (for absolute states) – intransitives (*melt, freeze, die, be born, molt, ignite, explode, collapse*) or two-place phrasal (*turn into N, turn to N, become Adj*); Change of physical state (degree state) – intransitive (*darken, warm, cool, sink, improve*) or phrasal (*become Adj*); Aspectual complement verbs – infinitive complement (*begin, start, cease*), gerundive complement (*stop, resume, begin, start*) or event nominal as subject (*end, stop, resume, start, begin*); Possessive (*acquire, receive, get, lose*); Change of state of consciousness (*awaken, fall asleep*).

Accomplishments: Locatives – transitive verbs involving enclosure (*hide, cover, box, uncover, create*), two-place phrasals derived from activity verbs with locative result state (*drive, carry, push*), or two-place phrasals not derived from activity verbs (*put, place, set NP into NP*), transitive with extent (*carry, push, drive NP a mile*); Intransitives that are not locatives (*shape up, grow up*); Transitive verbs of creation (*accusativus effectivus*) – derived from activities (*draw, knit, dig*) or not derived from activities (*make, build, create, construct, erect*); Transitive verbs of destruction (*destroy, obliterate, raze, melt, erase, eat*); Transitive change of state (*kill, petrify, marry NP to NP, cook, paint*).

Levin (1993) offers a more detailed classification of English verbs which lies in the base of the present study. The classification relies on the verbs' general semantics and the diatheses in which they participate. Within the class of so called "pure" change of state verbs, Levin distinguishes between several semantic subclasses: change of physical integrity (break verbs, 45.1²); verbs to change shape without breaking the physical integrity (bend verbs, 45.2); verbs for heat treatment (cooking verbs, 45.3); verbs of change on a scale (calibratable change of state, 45.6), e.g. *increase, decrease*; verbs of entity-specific change of state (45.5), which combine predicates denoting a change characteristic of certain entities, e.g. *bloom, rust, erode*; other change-of-state verbs (45.4), united by their similar syntactic behavior, including conversion and verbs derived from adjectives or nouns, e.g. *clean, narrow, blunt, soften, flatten, decentralize, caramelize*.

²Here and below, when referring to Levin's verb classes, when appropriate, the relevant number of the class will be presented as per the classification of Levin (1993).

Ivelina Stoyanova

Some of the other categories, although implying a change, exhibit a more complex semantic structure, thus they require separate detailed analysis. Additional attention is needed to handle verbs of creation and transformation (26), destroy verbs (44), verbs of killing (42), verbs of ingesting (39) which involve an agent and a patient that undergoes a change of state – comes to existence or ceases to exist.

In addition, within the verbs of change class we also include several more groups of verbs. Firstly, we take into account verbs denoting externally inflicted physical change (verbs of cutting, 21), change of (body) position (roll verbs, 51.3.1) and change of location (verbs of putting, 9; verbs of removing, 10; verbs of sending and carrying, 11), as well as verbs denoting change in the psychological or emotional state (amuse type psych-verbs, 31.1) and verbs of change of possession (give verbs, 13.1, contribute verbs, 13.2, verbs of obtaining, 13.5, verbs of exchange, 13.6) which are always transitive and require an agent. Secondly, we consider verbs that are inherently intransitive and do not allow a causative counterpart such as verbs of appearance, disappearance and occurrence (48).

In a refined and enriched form, with integrated additional classes, the classification of Levin (1993) becomes the basis of the organisation of verbs in VerbNet (Kipper-Schuler 2005), a lexical-semantic resource presenting a shallow hierarchy of verb classes in English, the semantic roles describing the argument structure of the predicates of each class, the selectional restrictions of the arguments, their syntactic realisation, the diatheses in which the verbs participate, etc.

Levin (2010), following research by Kennedy & Levin (2008), Krifka (1998), Rapaport Hovav (2008), Beavers (2008), proposed a unified approach to the determination of telicity by considering three verb classes: incremental theme verbs with respect to the extent of the argument (e.g., volume, area, height, etc.); true change-of-state verbs with an argument exhibiting a gradable property of the argument; and inherently directed motion verbs for which the path of motion is a measurable feature. Further, to integrate the classes together with respect to the treatment of telicity, Levin (2010: 1–2) introduces the feature *SCALE* and considers it as an integral part of the verb semantics and conceptual structure. Three types of scales are then distinguished: (a) *extent scales* characterising verbs with incremental theme; (b) *property scales* most often found with change of state verbs; and (c) *path/spacious scales* most often found with inherently directed motion verbs.

Property scales can be fully lexicalised in verbs: it may be a binary scale (transitions from one state to another, e.g. *break*, *crack*, *die*), but can also involve multiple stages or degrees (e.g. *cool*, *age*, *deepen*); a change of state verb may specify all components of such a scale (e.g., *redde*n) or only some of them (e.g., *darken*).

3 The complex conceptual structure of verbs of change

Extent scales are not lexicalised in verbs but concern the THEME in their conceptual structure. Path scales denote change of place through motion and may be lexicalised in some motion verbs (e.g., *ascend*, *descend*).

Levin (2010: 2) points out that verbs of scalar change with a bound have inherent telic interpretation; non-scalar verbs (e.g., *walk*, *run*, etc. – verbs for habitual activities) can also achieve telic interpretation when used in particular context. An interesting observation is that scalar verbs rarely lexicalise manner while non-scalar do.

Van Valin (2005) also considers verbs denoting change of state to belong to the classes of achievements and accomplishments due to the fact that change implies an inherent terminal point, therefore it entails telicity, and only these two classes are telic as compared to states and activities which are atelic. Van Valin (2005: 33), following previous research (Vendler 1967, Dowty 1979), states the four key semantic features which determine the scope of the classes: [\pm static], [\pm dynamic], [\pm telic] and [\pm punctual]. Punctuality distinguishes achievements (which are punctual) from accomplishments (which are non-punctual). The logical structure of the predicates denoting change (Van Valin 2005: 42–49) are as follows (where BECOME and INGR are operators and the ingressive operator INGR denotes the implied resulting state in the change of state verb):

Achievement: INGR **predicate'** (x) or (x, y), or INGR **do'** (x, [**predicate'** (x) or (x, y)]).

Accomplishment: BECOME **predicate'** (x) or (x, y), or BECOME **do'** (x, [**predicate'** (x) or (x, y)]).

Active accomplishment: **do'** (x, [**predicate'**₁ (x, (y))]) & INGR **predicate'**₂ (z, x) or (y).

Causative: α CAUSE β , where α , β are logical structures of any type.

Active accomplishments are composed of an activity and a terminal point with result state, thus they are more accurately characterised as 'active achievements' rather than activities.

In summary, we adopt Vendler's classification of activities, achievements, accomplishments and states with the relevant features and tests to distinguish between them. Verbs of change within this classification fall into the categories of achievements and accomplishments as they express a transition from one state to another, so there is an inherent end point.

Ivelina Stoyanova

The literature review shows that in the last decades much effort has been invested into the semantic classification of verbs. The overview of classifications points to the main features underlying the syntactic behaviour of verbs of change. Also, the review leads to the conclusion that the syntactic realisation of the verb is not determined fully by its lexical meaning; however, there are clearly features which are derivable solely by its semantics (the meaning and the belonging to a particular semantic class).

The well-known classifications presented here are taken into account in order to offer a comprehensive description of the most frequent verb subclasses within the class of verbs of change as represented in Princeton WordNet. The classification of Levin (1993) as one of the most extensive with respect to verbs of change is (partially) aligned with the classification proposed in the current study based on FrameNet frames (Section 5).

The review also outlines the key lexical, semantic and grammatical features which are taken into account in the analysis: telicity, quantised change, causativity and inchoativity, frame elements and their configurations in frames evoked by the verbs.

The review of relevant studies on verbs of change primarily focused on English verbs, their properties and classification, helps to outline the scope of the study and to establish the background for the study of Bulgarian verbs by applying cross-linguistic analysis and transfer of information employing the cross-linguistic potential of WordNet and FrameNet.

3 Resources

The description of verb semantics and the grouping of verbs into semantically and syntactically homogeneous classes takes different directions depending on the adopted theoretical approach and the degree of detail of the description in the different resources. The focus is on the representation of verbs of change in the hierarchical structure of Princeton WordNet and the Bulgarian WordNet and the semantic frames that describe them, along with the semantic relations between the frames. The study of the systematic semantic relations in each of the resources, as well as the characteristics determining the relationship between their basic units (synsets and semantic frames) and the relations between them, supports the enrichment of synsets in WordNet with conceptual information. Moreover, this will support the study of the degree of semantic specialisation of verb meanings and the granularity of conceptual description (using frames).

3 The complex conceptual structure of verbs of change

The present study is based on information from two main semantic resources – Princeton WordNet (WordNet) and the Bulgarian WordNet (BulNet), and FrameNet. The semantic features discussed are relevant at synset level and is largely applicable to both English and Bulgarian, that is why we consider the two word-nets as one collective resource.

3.1 Verbs of change in WordNet

Given the diversity in the semantics of change of state verbs, the most comprehensive information about the set of these verbs is provided in the lexical-semantic network WordNet (Miller 1995, Fellbaum 1999), which represents the lexicon in the form of a network of synonym sets (synsets) interconnected by semantic, lexical and other relations. One of the main relations building the structure of WordNet is hypernymy (and its opposite relation – hyponymy), by which the vocabulary of a given semantic field is organised into a tree, the beginning (root) of which is the most general or abstract concept of the corresponding field.

WordNet, as well as its Bulgarian counterpart BulNet (Koeva 2007, 2021), is the main resource used in the study. The semantic description of verb predicates in WordNet also includes their classification into general semantic classes based on assigned semantic primitives (Miller & Fellbaum 2007), e.g. verbs of motion, verbs of emotion, verbs of communication, verbs of change, etc. Verbs of change in WordNet are a large and diverse class. In general, they include:

- (a) Causative verbs of change covering the tree of synsets rooted at eng-30-00126264-v: {*change*:1; *alter*:1; *modify*:3}, ‘cause to change; make different; cause a transformation’ (regardless of the semantic primitive).
- (b) Non-causative verbs of change covering the tree of synsets rooted at eng-30-00109660-v: {*change*:2}, ‘undergo a change; become different in essence; losing one’s or its original nature’ (regardless of the semantic primitive).
- (c) Verbs lying outside of the two WordNet trees but assigned the semantic primitive verb.change, together with their hyponyms.

Around half of the verbs falling under (a), (b) and (c) are labelled with the semantic primitive verb.change (Leseva & Stoyanova 2021) while others have distinct semantics expressed by a different semantic primitive. More specific details on semantic classes of verbs in WordNet will be discussed in Section 5.

3.2 Frames related to change in FrameNet

FrameNet (Fillmore et al. 1998) is a system of semantic frames with their frame elements. Frames are schematic descriptions of the conceptual structure of situations through actors, circumstances, and other conceptual roles presented as frame elements. The frames are organised in a system by means of a number of relations – hierarchical (Inheritance, Use, Subframe, etc.) and other types of relations (for example, Causation).

There have been efforts to construct a FrameNet-based resource for Bulgarian. Koeva (2010a) discusses the properties of the resource BulFrameNet – a corpus-based lexicon giving an exhaustive account of the semantic and syntactic combinatory properties of Bulgarian verbs. Koeva & Doychev (2022) presents BulFrame – a web-based system for the extensive description of verbs using semantic frames offering a unified theoretical model for the formal presentation of frames and frame elements. The framework expands on FrameNet semantic frames by defining the sets of nouns that can be combined with a given verb. This is achieved by an ontological representation of noun semantic classes.

WordNet and FrameNet are automatically mapped (Leseva & Stoyanova 2020), with synsets in WordNet being assigned semantic frames from FrameNet. Over 5000 of the frames assigned to synsets have been manually validated (Leseva & Stoyanova 2020). Verb literals in a synset have a common meaning, which implies a common semantic frame. This makes assigned frames valid regardless of the language and allows the transfer of frame information from one language to another. This is in line with previous observations that the argument structure depends to a large degree on the lexical meaning of words (Levin 1993: 1–3), which suggests that verbs with similar meaning would exhibit the same or similar valency patterns.

Lexical units in FrameNet, in particular verbs, are grouped in semantic frames based on common semantics, formalised through a common set of participants and circumstances (frame elements) and the relations between them (Fillmore 1982, 1985, Fillmore et al. 2003, Fillmore & Baker 2009, Ruppenhofer et al. 2016) with valence patterns inductively derived from corpus evidence. The underlying principle of organisation is semantic similarity in terms of the situation described by the verb, which may lead to grouping together verbs with different syntactic behaviour in terms of alternations as shown by Baker & Ruppenhofer (2002) in their comparative analysis between Levin's classes and FrameNet frames.

A number of parallels can be drawn between the classes proposed by Levin (1993) and the frames in FrameNet. Thus, for example, verbs of change of physical integrity (45.1) largely correspond to the *Cause_to_* fragment and *Breaking_*

3 The complex conceptual structure of verbs of change

apart frames, verbs of change of shape (45.2) correspond to the Reshaping frame, cooking verbs (45.3) – with the frames Apply_heat and Cooking_creation, verbs of calibratable change (45.6) – with the frame Cause_change_of_position_on_a_scale and Change_position_on_a_scale, etc. Many of the verbs of entity-specific change (45.5) and the remaining verbs of change of state (45.4), which do not fall into the listed categories and which constitute the majority of verbs of change, correspond to a number of other frames. Thus, for example, many predicates formed from adjectives, depending on the changing feature which they lexicalise, correspond to specific frames: *dry* (causative and inchoative) – Cause_to_be_dry and Become_dry; *cool, chill, warm, heat* (causative and inchoative) – Cause_temperature_change and Change_of_temperature; *blacken, redden, darken* – Cause_change_of_colour and Change_of_colour, etc. Large groups of transitive verbs belong to frames that inherit the frame Transitive_action, i.e. represent more specific situations that have the same or a more specific set of frame elements.

Frames that cover specific verb subclasses within the class of verbs of change will be discussed in Section 5, together with their frame elements and examples for their use in English and Bulgarian.

3.3 Usage examples

Usage examples in Section 5 illustrating the use of verbs of change and their behaviour are mostly drawn from the SemCor and BulSemCor – both annotated with WordNet senses.

SemCor (Miller et al. 1993, 1994, Landes et al. 1998) is manually annotated corpus developed by the Princeton WordNet team. Open-class single words and multiword expressions are assigned unique WordNet senses. The corpus contains a total of 226,040 sense annotations.

BulSemCor (Koeva et al. 2006, 2011, Koeva 2012) follows the methodology of SemCor and aims to ensure good coverage of varied general lexis. In addition to open-class words, in BulSemCor closed-class words (preposition, conjunctions, particles) are also annotated. The size of BulSemCor is close to 100,000 annotated units.

The use of SemCor and BulSemCor is motivated by the fact that verbs are annotated with word senses, so they are uniquely referred to a particular WordNet synset which allows the straightforward extraction of comparable examples in English and Bulgarian.

4 Properties determining the syntactic realisation of verbs of change

4.1 Telicity and quantised change

Rappaport Hovav (2008) discusses at length the relationship between the aspectual property of telicity and the notion of scale. Before that, Dowty (1979: 88–90) talks about degree achievements and Hay et al. (1999) consider the classes of verbs of quantised and non-quantised change which exhibit different syntactic behaviour based on the scalar structure of the base adjective (representing the property of change). With verbs of quantised change the progressive does not entail the perfect (which means that performing the action does not entail the result), whereas with verbs of non-quantised change it does (e.g., *The soldier is dying* \Rightarrow *the soldier has died*; *The child is growing* \Rightarrow *the child has grown*). The verbs of non-quantised change are telic, entailing a specific terminal state. In this class fall verbs like die, be born, fall, arrive, and many verbs of inherently directed motion (Levin 1993).

For verbs of change of state, the entity associated with the change is considered as an incremental theme (Dowty 1991, Rappaport Hovav & Levin 2005), which limits their argument projection options. When the verb describes a specified degree of change on the scale, it is telic, and when it describes an unspecified degree of change, it is atelic. The objects of both traditional incremental theme verbs have a scale for measuring the event's progress and determines the telicity of its sentence. If the object is quantised, as is the change on the given scale, the sentence is telic. If the object is not quantised, the scale lacks a specified endpoint, the change on this scale is unspecified, and the sentence is atelic. This comes to show that although argument expression is determined by the verb's meaning, it is also influenced by specific aspectual properties at lexical or syntactic level.

Telic verbs are incompatible with some adverbials such as *completely* and *not completely*, given that they have a specific final state as one of their entailments (*The soldier has died #completely/#but not completely*). Hay et al. (1999: 129) state that for verbs of non-quantised change telicity is implied when the adjectival base of a deadjectival verb (e.g., *empty*, *straighten* etc.) includes a gradable property in its semantics. Such verbs are telic (e.g., *He is emptying the bathtub* \Rightarrow *He has emptied the bathtub*) but are compatible with *completely/not completely* (*He has emptied the bathtub completely/but not completely*).

The telicity of a sentence with a change of state verb depends on the nature of the scalar property lexicalised in the verb and the scale along which this property changes, for example verbs entailing temperature change are associated with a

3 The complex conceptual structure of verbs of change

temperature scale, and a sentence with such verb is telic if the change in temperature is specified and atelic otherwise (*The water warmed to room temperature* vs. *The water warms slowly*).

Telicity can also arise based on a conventional (not lexicalised but assumed) property of the undergoer (usually the frame element THEME). Verbs like *read*, *write*, and verbs of consumption allow their direct object to be interpreted as an incremental theme. When the quantised object is expressed, they may appear either as telic or atelic (see Example 4). The same can be observed also with some verbs of change (Examples 4e, 4f).

- (4) a. *He READ for hours.* (atelic)
- b. *He READ the book for hours.* (atelic)
- c. *He READ the book.* (telic)
- d. *He READ the book in a day.* (telic)
- e. *He BUILT the house for years.* (atelic)
- f. *He BUILT the house in five years.* (telic)

The syntactic behaviour of verbs of change of location also can be analysed in terms of the incremental theme. Some verbs allow for different interpretations and assuming different incremental property/entity (e.g., locative alternation verbs) as their direct object (Example 5).

- (5) a. *I SPRAYED the whole can of paint onto the wall.*
- b. *I SPRAYED the whole wall with paint.*

In Bulgarian and other Slavic languages the verb aspect is a lexical category and the perfective and imperfective verbs are considered different words with different lexical meanings derived as a result of word formation. The properties of the verb aspect have been studied extensively (Andreychin 1944, Ivanchev 1971, Nitsolova 2008, Kutsarov 2007, Koeva 2011, Charalozova 2021).

Example 6 show use cases of the verbs *готвя* ‘cook’ (imperf.) and *сготвя* ‘cook’ (perf.), the latter derived from the former using prefixation. Relevant to the study of the semantic properties of verbs of change is the fact that perfective verbs are always telic (Examples 6d, 6e) and the limitations in their interpretation stemming from that (e.g., Example 6f). Ambiguity at the lexical level between the telic and atelic interpretation of verbs can only occur with imperfective verbs (Example 6a, 6b vs. 6c).

Ivelina Stoyanova

- (6) a. Той **ГОТВИ** часове наред.
He cooks hours in sequence.
'He cooks for hours on end.' (*готвя* 'cook', imperf.; atelic)
- b. Той **ГОТВИ** ястието часове наред.
He cooks dish-DEF hours in sequence.
'He cooks the dish for hours.' (*готвя* 'cook', imperf.; atelic)
- c. Той **ГОТВИ** това ястие за един час.
He cooks this dish for one hour.
'It takes him one hour to cook this dish.' (*готвя* 'cook', imperf.; telic)
- d. Той **СГОТВИ** ястието.
He cooked dish-DEF.
'He cooked the dish.' (*сготвя* 'cook a complete dish', perf.; always telic)
- e. Той **СГОТВИ** ястието за един час.
He cooked dish-DEF for one hour.
'He cooked the dish in an hour.' (*сготвя* 'cook a complete dish', perf.; always telic)
- f. *Той **СГОТВИ** ястието часове наред.
He cooked dish-DEF hours in sequence.
*'He (completely) cooked the dish for hours.' (*сготвя* 'cook a complete dish', perf.; always telic)

4.2 Causative and inchoative verbs of change

An inchoative/causative verb pair is a pair of verbs which express the same change of state and differ only in that the causative verb meaning includes an agent or a cause participant who causes the situation, whereas the inchoative verb meaning excludes a causing agent and presents the situation as occurring spontaneously (Example 7).

- (7) a. (inchoative) *The stick broke.*
b. (causative) *The girl broke the stick.*

Van Valin (2005: 45) presents the following logical structure of the causative verb: ' α CAUSE β , where α , β are logical structures of any type'. In particular, verbs of change are marked by the operators BECOME and INGR which show either that the entry changes state or that there is a resulting terminal state in the verb's semantics (Example 8).

3 The complex conceptual structure of verbs of change

(8) Inchoatives

- a. *The window shattered.*
INGR shattered (window)
- b. *The balloon popped.*
INGR popped (balloon)
- c. *The snow melted.*
BECOME melted (snow)
- d. *Mary learned French.*
BECOME know (Mary, French)

Causatives

- a. *The dog scared the boy.*
[do (dog, \emptyset)] CAUSE [INGR feel (boy, [afraid])] ³
- b. *Max melted the ice.*
[do (Max, \emptyset)] CAUSE [BECOME melted (ice)]
- c. *The cat popped the balloon.*
[do (cat, \emptyset)] CAUSE [INGR popped (balloon)]

In his study on deep lexical semantics linking lexical meaning to underlying abstract core theories via lexical decompositions, Hobbs (2008, 2014), following on the work of Fillmore et al. (1988), considers causal complex (as opposed to cause only) as a composite structure which includes all the states and events that have to happen or hold in order for the effect to occur. In view of this, Hobbs (2008: 190) considers FrameNet to reflect these principles, describing frames as “axiomatic characterizations of abstract situations”, and presents causation in terms of a complex structure involving multiple participants and conditions (Hobbs 2008: 184–188). This approach allows for more thorough analysis of causativity as well as the relation between the properties of the situations described by the inchoative and the causative verbs in the pair.

Levin & Rappaport Hovav (1995: 90–98) argue that with respect to causativity, the verb’s behaviour is determined by its lexical meaning and depends on whether the verb lexicalises an internally caused event or an externally caused event. An externally caused event implies an external cause with immediate control over the event, such as with verbs for change of state, which are usually transitive but also allow intransitive use where the cause is not explicit (e.g., *He*

³The original logical structure in Van Valin (2005) is [do (dog, \emptyset)] CAUSE [feel (boy, [afraid])] and the operator INGR here is added for conformity to denote that this is the final state.

Ivelina Stoyanova

opened the door – *The door opened*). In contrast, an internally caused event occurs due to some inherent properties of the entity participating in the situation, without an external cause, and thus these verbs are typically intransitive and do not enter the causativity alternation.

For English verbs causativity related to verbs of change is analysed with a view to the causative/inchoative alternation. Levin (1993: 10, 30) argues that this alternation is sensitive and only applies to pure change of state verbs, roughly covering verbs of change of state and change of position/location. Verbs that imply a change of state only indirectly in their meaning, for example as a consequence of the situation described by the verb, are not found in the causative/inchoative alternation. These include some verbs that are only used transitively (in a causative meaning), such as verbs of change of possession (*loan, rent, lend, refund, donate, transfer*), verbs of contact by impact (*hit, bang, beat, kick*), verbs of cutting (*cut, chop, slice, shed*), etc., as well as some verbs that are only used intransitively (in an inchoative meaning), such as verbs of appearance, disappearance and occurrence (*appear, disappear, occur, arise, emerge, erupt, happen, expire*) and verbs of entity-specific change of state (*blister, bloom, blossom, decay, deteriorate, erode, ferment, flower, germinate, rust, sprout, swell*). Psych-verbs (verbs of psychological state change) rarely participate in the causative/inchoative alternation in English (*amuse, madden, puzzle, sadden, sicken, worry*), but this is a typical alternation in other languages such as French, Italian, and Russian (Levin 1993: 30) as well as Bulgarian (see Example 9).

- (9) a. (inchoative)
Детето СЕ ЗАРАДВА (на подаръка).
 Kid-DEF became happy (at gift-DEF).
 ‘The kid became happy (because of the gift).’
- b. (causative)
Майката ЗАРАДВА детето (с подарък).
 Mother-DEF made happy kid-DEF (with gift).
 ‘The mum made the kid happy (with a gift).’
- c. (causative)
Подаръкът ЗАРАДВА детето.
 Gift-DEF made happy kid.
 ‘The gift made the kid happy.’

As stated by Mel’cuk (1967: 352–353) and Haspelmath (1993: 89), causative verbs, on purely semantic grounds, are considered to be derived from (also seen

3 The complex conceptual structure of verbs of change

in the logical structures in Example 8 above). Semantically, *A melts* is a simpler structure than its causative *B causes A to melt*. However, it has been pointed out that in Russian (and the same is valid for Bulgarian), considering derivation in terms of form, there are predominantly examples of the reverse derivation where the inchoative is formally derived from the causative (Haspelmath 1993: 89, and Example 10). Haspelmath (1993: 103–106) argues that the likelihood of spontaneous vs. caused events is the main factor determining the direction of derivation in inchoative/causative verb pairs. While the direction of formal derivation and the investigation of the meaning formation is beyond the scope of the current analysis, we discuss the causative/inchoative pairs of verbs and the correspondence of their meanings in view of the semantic and derivational relation they exhibit. In some cases, with a view to Bulgarian, the morphological means to derive counterparts in the causative/inchoative pair are also relevant in the analysis of the lexical meaning of verbs of change.

- (10) a. (causative) RU: *расплавить* /rasplavit’/ ‘melt (tr.)’
 BG: *топя* /topya/ ‘melt (tr.)’
 b. (inchoative) RU: *расплавиться* /rasplavit’sja/ ‘melt (intr.)’
 BG: *топя се* /topya se/ ‘melt (intr.)’

The inchoative member of an inchoative/causative verb pair is semantically similar to the passive of the causative but the difference is in that there is neither an explicit or implicit agent or cause as the situation occurs spontaneously.

Some causative verbs exhibit agent-oriented meaning components and in that case they don’t have an inchoative counterpart since the inchoative member implies the absence of an agent. Such component might be the instrument or means of performing an action – e.g., the verb *cut* implies a sharp instrument even if not explicitly stated in the sentence, so it does not enter the causative/inchoative alternation, as opposed to the verb *tear* (Example 11).

- (11) a. *I cut the sheet (with scissors).*
 b. **The sheet cuts.*
 c. *I tore my dress (on the fence).*
 d. *My dress tore (on the fence).*

Similar verbs that imply agent-oriented components are verbs like *wash*, *execute*, *tie*, in contrast to verbs such as *clean*, *kill* (*kill/die* counterparts), *untie*. In Bulgarian such verbs although resembling the form of verbs with the reflexive

Ivelina Stoyanova

particle *se* in Examples 9 and 10, when they are perfective or derived from perfective, have only a passive interpretation (Example 12c, 12d). In the case when the verbs are imperfective, alongside the passive interpretation, they can also have a stative interpretation (Example 12b – describing the property of the verb to be cut).

- (12) a. **НАРЯЗВАМ** листа на ленти (с ножица).
Cut.1sg sheet-DEF into strips (with scissors).
'I cut the sheet into strips (with scissors).'
- b. Листът **СЕ РЕЖЕ** (лесно).
Sheet-DEF cuts (easily).
'The sheet cuts (easily).' (stative interpretation of the imprperf. verb *режа*)
- c. ?Листът **СЕ НАРЯЗВА** (лесно).
?Sheet-DEF cuts (easily).
? 'The sheet cuts (easily).' (stative interpretation of the imprperf. verb *нарязвам* which is derived from the perf. *нарежа* is blocked; the construction is perceived as passive)
- d. Листът **СЕ НАРЯЗВА** на ленти (с ножица).
Sheet-DEF is cut into strips (with scissors).
'The sheet is cut into strips (with scissors).' (passive interpretation)

The observations on morphological vs. lexical ambiguity are aimed at demonstrating the scope of the complex nature of the class of verbs of change and the influence of their semantic and morphosyntactic properties in determining their behaviour and interpretation in context. In particular, we would like to distinguish cases of passive with reflexive article *se* from cases of the use of independent inchoative lexical units. Moreover, further research is needed into the derivational means in Bulgarian involved in filling the gaps in possible causative/inchoative pairs within the class of verbs of change. This line of analysis would be beneficial for enriching WordNet and FrameNet both in terms of lexical coverage and in terms of relational structure.

4.3 Structure of the situation described by verbs of change

The analysis of the situations described by the verbs of change adopted here relies on Frame Semantics (Fillmore 1982, 1985, Fillmore & Baker 2009). With respect to causativity and the analysis of the situation, to some extent the generalised semantic roles of Role and Reference Grammar (Van Valin 1993, Van Valin

3 The complex conceptual structure of verbs of change

& LaPolla 1997), and their hierarchical organisation, are also relevant. However, here we focus on the frames evoked by verbs of change and consider the more detailed, fine-grained and verb class specific participants in the situation, as represented by the frame elements within each frame. The most characteristic core frame elements with regards to the verbs of change and their description will be discussed. Moreover, the syntactic realisation of the frame elements and the valence configurations they enter is also considered since it is relevant for the classification of the verbs of change and their use in the inchoative and the causative scenario.

Pure change of state verbs such as *break*, *tear*, *bend*, in both their transitive and intransitive uses, express a change of state (plus a notion of cause when transitive). Very often, in its inchoative variant, a verb of change of state has one core frame element, denoting the entity undergoing the change of state, which is realised as the external NP. As already discussed above, the verb includes in its semantics the notion of the property that changes. This property then serves as a restriction on the type of the entity that can take this position (Example 13).

- (13) a. Physical entity (flexible).
*The stick bends, The girl bends, *The parliament bends.*
- b. Physical entity (body, substance).
*The water froze, The ground froze, *The religion froze.*
- c. Physical entity (growing).
*The child grows, The city grows, *The rocks grow.*
- d. Physical entity (fluids).
*The water drained, The blood drained, *The rocks drained.*
- e. Sentient entity.
Момичето се развесели.
 Girl-DEF became happy.
 ‘The girl became happy.’
Селото се развесели.
 Village-DEF became happy.
 ‘The village became happy.’
**Дървото се развесели.*
 Tree-DEF became happy.
 **‘The tree became happy.’

The causative variant of the verb requires one more participant, namely the AGENT. In general, the AGENT can be expressed by a sentient entity or by a non-sentient CAUSE that evokes the change; they appear as frame elements in the

Ivelina Stoyanova

respective causative semantic frame in FrameNet and are usually realised as the external NP in the sentence (Example 14). The meaning of some verbs inherently involves agent-specific components, for example an INSTRUMENT or MEANS. As already mentioned above, such verbs require the existence of an AGENT that uses this INSTRUMENT/MEANS to bring about a change of state in the PATIENT (hence such verbs do not have an inchoative counterpart). In the causative realisation, either the AGENT, the CAUSE or the INSTRUMENT/MEANS can take the position of the subject (Example 15).

- (14) A pair of corresponding inchoative and causative frames (Breaking_apart and Cause_to_fragment).

Frame: Breaking_apart

WHOLE **PIECES**

[The river]_{WHO} **SPLITS** [into two streams]_{PIE}.

Frame: Cause_to_fragment

WHOLE_PATIENT **PIECES** **AGENT** **CAUSE**

[She]_{AGE} **SMASHED** [the plate]_{WHOPAT} [into little pieces]_{PIE}.

- (15) a. [The man]_{AGE} **CUT** the bread with a knife.
 b. [The knife]_{INS} **CUT** the bread.
 c. With this interview [the president]_{AGE} **ANNOUNCED** his resignation.
 d. [The TV channel]_{MEDIUM} **ANNOUNCED** the president's resignation.
 e. [The wind]_{CSE} **BROKE** the branch.
 f. [I]_{AGE} **DISTRACTED** him with my questions.
 g. [My questions]_{MNS} **DISTRACTED** him.

The changing quantised property can also be concretised in terms of its degree on the scale, as seen in Section 4.1 (Example 16). It is interesting to analyse the adverbials expressing the position of the value on the scale up to the terminal point (reaching a terminal value or state).

- (16) a. **ИЗПРАЗНИХ** ваната наполовина.
 Emptied.1sg bath-DEF halfway.
 'I emptied the bathtub halfway.'
 b. **ЗАТОПЛИХ** водата до 40 градуса.
 Warmed.1sg water-DEF to 40 degrees.
 'I warmed the water up to 40 degrees.'

3 The complex conceptual structure of verbs of change

- c. *Децата СТИГНАХА почти до парка.*
 Kids-DEF reached almost to park-DEF.
 ‘The kids went almost as far as the park.’
- d. *Той РАЗСМЯ публиката до сълзи.*
 He made laugh audience-DEF until tears.
 ‘He made the audience laugh to tears.’

Other frame elements that appear in the frames evoked by verbs of change include:

- GOAL: realised in the logical configuration INGR/BECOME be-at/in/on (x, y) where y is the GOAL (Example 17).
- SOURCE: appearing in a possible configuration INGR/BECOME NOT have (x, y) or TERMINATE be-at/in/on (x, y₀) & INGR/BECOME be-at/in/on (x, y) (Example 18).
- RECIPIENT: may be defined as the possessor argument in a configuration such as INGR/BECOME have (x, y) where the RECIPIENT is y (Example 19).

- (17) *The river level **RISES** [up to 10 m]_{GOAL} in the spring.*
*We **LIFT** the load [to the 10th floor]_{GOAL}.*
- (18) *The tea **SPILLED** [from the cup]_{SRC}.*
*Sam **POURED** hot water [from the teapot]_{SRC}.*
- (19) *She **LENT** her bicycle [to Sam]_{REC}.*
*We **CONTRIBUTED** our paycheck [to the foundation]_{REC}.*

By assigning frames to the synsets in WordNet, we aim at defining semantic classes of verbs based on similar lexical semantics, but more importantly, that evoke the same or similar (related) frames which exhibit similar configurations of frame elements. Further, annotated corpus examples as presented in Section 5 provide material for the comparative analysis of verbs within the same frame that exhibit similar or different valence patterns and syntactic realisation of their frame elements. Moreover, these analyses can be extended to a cross-language level in an attempt to analyse the semantic and syntactic properties determining the verb class behaviour for various languages.

The set of frame elements characterising the evoked frames also participate in a (shallow) hierarchical structure determined by the inheritance relations between the frames. For example, the frame element PATIENT in the frame Transitive_action is inherited in its inherited frame Cause_to_become_dry as the

Ivelina Stoyanova

DRYEE frame element which is more specific, either a surface or an entire entity, which is able to retain water inside and/or out and consecutively, to become dry.

The inheritance and correspondence between the more general and the more specific frame elements is demonstrated in Section 5.

5 Towards a classification of verbs of change with respect to their semantic properties and conceptual structure

The number and type of frame elements and their corresponding semantic restrictions are determined to a large degree by the lexical meaning of the verb, so it is theoretically founded to consider groups of semantically related verbs (verbs with a common hypernym) and to base the semantic classification of verbs of change on WordNet.

The class of verbs of change includes groups of predicates such as: (i) change in the degree of an inherent quality or property along a scale (Section 5.1); (ii) change in the integrity of an object (Section 5.2); (iii) change of state verbs involving the (momentous) transition into a new state, e.g. a change in the mode or form of existence (Section 5.3); (iv) change involving creation or transformation (Section 5.4); (v) changes in the conditions of an entity as a result of an outside influence (such as various manners of treating, adjusting, etc.), including ones involving the movement of something (putting, removing, etc.). The last group falls outside the scope of the present study.

Each subclass is characterised by a specific property of change which is either scalable (the property can be interpreted as a measure, incremental theme, quantity criterion, etc.) or momentous (the property describes transitioning into a new state). The property is lexicalised; in most cases the final (resulting) value is also lexicalised (e.g., *freeze*, *dry*, *vanish*) and thus have no syntactic realisation in the sentence.

Verbs with more general meaning can realise specific frame elements denoting the change (frame elements that are not incorporated in the verb's meaning), e.g. INITIAL_CATEGORY/FINAL_CATEGORY in the frame Cause_change, PRIOR_STATE/POST_STATE in the frame Progression, etc.

The two hypernym trees in WordNet containing the largest number of verbs of change are eng-30-00126264-v: {*change*:2} and eng-30-00109660-v: {*change*:1}, with causative and non-causative meanings, respectively. These two trees are relatively symmetrical, as many of the hyponyms in one tree have counterparts in the other established by the relation 'causes'.

3 The complex conceptual structure of verbs of change

There are also some causative/inchoative synset pairs for which no ‘causes’ relation has been defined. Moreover, there are also cases where the causative and inchoative meaning of a verb is encoded in the same synset (e.g., eng-30-00280532-v: {*blacken*; *melanize*; *melanise*; *nigrify*; *black*} ‘make or become black’). These are cases that ideally will be treated separately as they should be assigned different frames.

Correspondence in the organisation of the two WordNet trees is also reflected in the frames assigned to the synsets in them. Two interframe relations denoting causation and inchoativity are encoded in FrameNet: **Causative of** and **Inchoative of**. However, these relations are not consistently reflected in FrameNet either. Additional work was done to correlate the pairs of synsets and the pairs of frames via the relation of causation based on similarities in definitions, matching of semantic classes, and correspondence between the frame elements (Leseva & Stoyanova 2021).

Among the hyponyms within the two trees of causative and inchoative verbs of change there is a range of semantic classes, the most frequent of which are: verb.contact (verbs for /physical/ contact), verb.possession (verbs for possession), verb.body (verbs of bodily processes, dressing, etc.), verb.social (verbs, representing social relations and interaction), verb.motion (verbs for motion), etc. They show a narrowing and specialisation of the meaning and can serve as a basis for forming semantic subclasses within the class of verbs of change. Most of these classes fall into the description of the Aktionsart classes pertaining to the verbs of change and their logical structure (Van Valin 2005: 42–49) and can also be aligned to Levin’s classes (Levin 1993: 240–247).

The beginning of the structure of frames in FrameNet, describing a significant part of the causative verbs of change, is the abstract frame *Transitive_action*, which refers to situations in which an ‘AGENT or CAUSER affects a PATIENT’. The frames that inherit it represent instantiations of it, the most general among which is the frame *Cause_change*, which is assigned to the root of the causative tree, the synonym set eng-30-00126264-v: {*change*:2}, as well as to other verbs whose semantics is adequately described by the frame or no other, more specific frame has been found when applying the automatic frame assignment procedure. Its direct inchoative counterpart is the frame *Undergo_change* assigned to the synset eng-30-00109660-v: {*change*:1}. In the course of the manual verification, where possible, new, more specific frames describing the semantics of a group of synsets have been assigned and some new frames have also been defined to cover larger sets of synsets (Leseva & Stoyanova 2019, 2021).

This section presents the linking of: (1) the general semantic classes of verbs of change; (2) their corresponding (sub)classes in Levin’s classification (Levin

1993) which is more often not a direct match but a many-to-many mapping; (3) a shallow hierarchy of FrameNet frames evoked by verbs of these semantic classes with their corresponding frame elements (only some characteristic frames are discussed); (4) inchoative – causative correspondences between frames when such correspondence exists; (5) a typical WordNet synset that often presents a (sub)tree of synsets (its hyponyms) evoking the given frame. Further, the examples demonstrate comparable realisations of some typical valence patterns in Bulgarian and English.

However, there have been studies pointing out the limitations of FrameNet especially with respect to the comprehensive coverage of verb semantics and its granularity. Rosca (2013) discusses entity-specific change of state verbs (45.5 in Levin’s classification): (i) verbs which describe an increase in size (e.g. *bloom, blossom, flower, germinate, sprout, swell*); (ii) verbs that describe a negative, destructive change affecting the integrity of an entity (e.g. *burn, corrode, decay, deteriorate, erode, rust, rot*); (iii) the verb *ferment* not related to a measurable increase/decrease in the values characterising the THEME. These verbs do not take part in a causative – inchoative opposition.

Some large semantic classes, although considered in general to the class of verbs representing change, are not covered here as they require a separate detailed analysis. Such is the case with the semantic class of motion verbs in WordNet which comprises a large group of verbs referring to change of spatial location of an object to a GOAL and/or from a given SOURCE, as well as the verbs denoting physical contact between two objects which results in a change of their properties, thus can be referred to the class of verbs of change. These verbs are traditionally classified as Verbs of Putting and Verbs of Removing (Levin 1993: 111–131), to which the frames Placing and Removing are assigned. Closely related to them are also the frames Filling and Emptying, where the action is directed to filling the location with the objects or emptying it of them, respectively. Another class that falls beyond the scope of the present study are verbs of emotional and psychological change and verbs of change of possession.

The theoretical analysis on the verb subclasses within the class of verbs of change is supplemented with observations on 30 FrameNet frames related to verbs of change which are assigned to over 100 WordNet synsets. In addition, the presentation of verb classes and the discussion on the syntactic realisations of verbs of change and their frame elements rely on 3482 sentences with a total of 9048 annotations in English extracted from FrameNet examples and SemCor, and 415 sentences with 866 annotated frame element realisations in Bulgarian extracted from BulSemCor and manually annotated.

3 The complex conceptual structure of verbs of change

To facilitate visualisation of the corresponding elements of the frame description, in the tables below presenting the frames covering the subclasses of verbs of change, the corresponding frame elements are presented in the same colour. The sign “>” in front of a frame name shows that it inherits from the frame directly above it in the table.

5.1 Verbs of change in the physical parameters of the object

A group of verbs lexicalise change in a physical or other property of the THEME which is a physical object or abstract entity: density or consistency, physical phase; temperature; volume, size or extent; color; change in appearance or characteristics, etc. (Table 1).

A large part of the verbs in this group are derived from adjectives expressing the relevant properties underlying the definition of the corresponding frames. Some of these frames are not included in the original frame system of FrameNet, but were defined in the course of subsequent work on FrameNet frames and their assignment to the synsets in BulNet (Leseva & Stoyanova 2021).

Examples 20a–20b show corresponding realisations of common valence patterns for the inchoative and the causative verbs, respectively. As seen from example 20a, although the frames concern physical characteristics of objects, they can receive metaphorical use as well (increasing in size of non-living entity, e.g. shadows, example 20a) or be applied to abstract entities (expand activity, example 20b). The causative frame Cause_expansion requires an additional frame element, AGENT or CAUSE.

- (20) a. Expansion: [NP.Ext]_{ITEM}
 [Дългите сенки]_{ITEM} **РАСТАТ** по пясъка и доближават
 [Long-DEF shadows]_{ITEM} grow on sand-DEF and approach
 към нас. (BulSemCor)
 to us.
 ‘The long shadows grow on the sand and approach us.’
- b. Cause_expansion: [NP.Ext]_{AGE} [NP.Obj]_{ITEM}
 [Софийският университет]_{AGE} постоянно **РАЗШИРЯВА**
 [Sofia-DEF University]_{AGE} constantly expands
 [участието си в културния живот]_{ITEM}. (BulSemCor)
 [participation-DEF in cultural-DEF life]_{ITEM}.
 ‘Sofia University constantly expands its participation in cultural life.’

Table 1: Frames for verbs of change in the physical parameters.

Inchoative	Causative
Frames denoting change of the physical parameters of the object Bend Verbs (45.2), some of Verbs of Entity-Specific Change of State (45.5)	
Go_into_shape THEME eng-30-00374135-v: {freeze:6}, 'change to ice'	Reshaping PATIENT CONFIGURATION DEFORMER CAUSE eng-30-00142191-v: {shape:2; form:2}, 'give shape or form to'
Change_posture PROTAGONIST eng-30-01983771-v: {change posture:1}, 'undergo a change in bodily posture'	
Change_of_phase PATIENT eng-30-00374135-v: {freeze:6}, 'change to ice'	Cause_change_of_phase PATIENT AGENT CAUSE eng-30-00375865-v: {freeze:7}, 'cause to freeze'
Expansion ITEM eng-30-00157844-v: {enlarge:2}, 'become larger or bigger'	Cause_expansion ITEM AGENT CAUSE eng-30-00154778-v: {enlarge:1}, 'make larger'
Becoming_dry ITEM eng-30-00219403-v: {dry:2; dry out:2}, 'become dry or drier'	Cause_to_be_dry DRYEE AGENT CAUSE eng-30-00218475-v: {dry:1; dry out:1} 'remove the moisture, make dry'
	Cause_to_be_wet PATIENT LIQUID AGENT CAUSE eng-30-00214951-v: {wet:2}, 'cause to become wet'
Corroding PATIENT eng-30-00273963-v: {corrode:1; rust:2}, 'become destroyed by water, air, or a corrosive such as an acid'	

3 The complex conceptual structure of verbs of change

5.2 Verbs denoting change in integrity

Some of the verbs are related to a (reversible or irreversible) change in the physical integrity of an object (Table 2). This group includes the class of verbs of breaking. The reason to consider them separately, is the fact that they involve mostly a momentous change in state rather than change on a scale like the verbs describing change of physical parameters above.

The pair of examples 21a–21b illustrate the use of verbs of this class with their core frame elements and typical valence patterns. The frame *Cause_to_fragment* inherits from the frame *Transitive_action*, and also (weakly) inherits from the frame *Destroying* (through the *Uses* relation).

Table 2: Frames for verbs of change in (physical) integrity

Inchoative	Causative
Frames denoting (reversible or irreversible) change in integrity Break Verbs (45.1), Destroy Verbs (44)	
	Destroying PATIENT DESTROYER CAUSE eng-30-01619929-v: {destroy:4; destruct:1}, ‘do away with, cause the destruction or undoing of’
Breaking_apart WHOLE PIECES eng-30-00334186-v: {break:12; separate:3; split up:1; fall apart:1; come apart:1}, ‘become separated into pieces or fragments’	> Cause_to_fragment WHOLE_PATIENT AGENT PIECES CAUSE eng-30-00334996-v: {break:13} ‘destroy the integrity of; usually by force; cause to separate into pieces or fragments’,

- (21) a. Breaking_apart: [NP.Ext]_{WHO} [PP]_{PIE}
[Моливът]_{WHO} се бееше СЧУПИЛ [на две]_{PIE}. (BulSemCor)
Pencil-DEF REFL has broken into two.
‘The pencil has broken into two.’
b. Cause_to_fragment: [NP.Ext]_{AGE} [NP.Obj]_{WHO PAT} [PP]_{PIE-INI}

Ivelina Stoyanova

[Тоѹ]_{AGE} **ИЗПОТРОШИ** [всички прозорци]_{WHO PAT} []_{PIE-INI}.
 He broke all windows.
 (BulSemCor)
 ‘He broke all windows.’

5.3 Verbs denoting transition into a new state

The verbs in this class are predominantly inchoative and include transition into a new state or phase in the development according to some known model concerning the development or the functioning of the ENTITY (Table 3), or the existence of the ENTITY (Table 4).

Examples 22a and 22b show the use of the frame *Transition_to_a_state* with the ENTITY realised as the external element. The new state can be expressed by an adjective, a noun phrase or an adverbial expression. The frame is applicable to both momentous transition and gradual (scalable) change, e.g. *become sensitive – become more sensitive* (Example 22b).

- (22) a. *Transition_to_a_state*: [NP.Ext]_{ENT}
Властта не искала [примерът им]_{ENT} да СТАНЕ
 Authorities-DEF did not want example-DEF their to become
 [заразителен]_{FINQ}. (BulSemCor)
 contagious.
 ‘The authorities did not want their example to become popular.’
- b. []_{ENT-DNI} **СТАВАШ** [по-чувствителен на външно влияние]_{FINQ}.
 Become.2sg more sensitive to external influence.
 ‘You become more sensitive to outside influence.’

5.4 Verbs of creation and transformation

The verbs described by the listed frames involve various types of manipulation or modification that lead to the creation of the ENTITY (Table 5): creating objects out of materials; building an object out of components; heat treatment and cooking, or its transformation (Table 6).

Predominantly the verbs of creation in this group imply agentivity, and thus have no inchoative counterparts. Verbs denoting transformation have both an inchoative and causative counterpart. Here transformation is assumed to lead to a categorically new ENTITY (starting as one category and ending in a new category of object) rather than it entering a new state or phase in its development as verbs in Section 5.3.

3 The complex conceptual structure of verbs of change

Table 3: Frames for verbs of transitioning into a new state

Inchoative	Causative
Frames denoting transition into a new state according to a known model Verbs of Entity-Specific Change of State (45.5)	
Progression	
ENTITY	
PRIOR_STATE	
POST_STATE	
eng-30-00094460-v: {grow:1; develop:1; produce:1; get:4; acquire:1}, ‘come to have or undergo a change of (physical features and attributes)’	
Transition_to_a_state	
ENTITY	
FINAL_CATEGORY or	
FINAL_QUALITY or	
FINAL_SITUATION	
eng-30-00149583-v: {become:1; go:1; get:7}, ‘enter or assume a certain state or condition’	

The semantic frame *Cooking_creation* while inheriting from *Intentionally_create*, inherits also (weakly) from *Apply_heat*, thus these frames exhibit more complex frame-to-frame relations. This is why the causative – inchoative pair of frames *Apply_heat* – *Absorb_heat* is also included in Table 5.

Examples 23a and 23b show the inheritance between the frames *Intentionally_create* and *Cooking_creation* with the inheritance between the frame elements, which become more concrete and specific, e.g. while the frame *Intentionally_create* allows for both abstract and concrete *CREATED_ENTITY*, the frame *Cooking_creation* implies that the entity is food or similar edible product of human activity (cooking).

- (23) a. *Intentionally_create*: [NP.Ext]_{CREATOR} [NP.Obj]_{CRENT}
[Президентът]_{CREATOR} СЪСТАВЯ [служебно]
President-DEF forms caretaker

Table 4: Frames for verbs of appearing and disappearing

Inchoative	Causative
Frames denoting coming into existence and ceasing to exist Verbs of Appearance, Disappearance, and Occurrence (48)	
Coming_to_be	
ENTITY	
eng-30-00425071-v: {appear:2; come along:2}, ‘come into being or existence, or appear on the scene’	
> Being_born	
CHILD	
Giving_birth	
CHILD	
MOTHER	
FATHER	
PARENTS	
Ceasing_to_be	
ENTITY	
eng-30-02156546-v: {vanish:5; disappear:4; go away:4}, ‘become invisible or unnoticeable’	
> Death	
PROTAGONIST	
Killing	
VICTIM	
KILLER	
CAUSE	
INSTRUMENT	
MEANS	

3 The complex conceptual structure of verbs of change

Table 5: Frames for verbs of creation

Inchoative	Causative
Frames denoting creation Verbs of Creation and Transformation (26), Cooking Verbs (45.3)	
	Creating CREATED_ENTITY CREATOR eng-30-01654628-v: {construct:4; build:6; make:27}, ‘make by combining materials and parts’
	> Intentionally_create CREATED_ENTITY CREATOR
	>> Building CREATED_ENTITY AGENT COMPONENTS
	>> Manufacturing PRODUCT PRODUCER FACTORY eng-30-01621555-v: {produce:2; make:22; create:2}, ‘create or manufacture a man-made product’
	>> Cooking_creation PRODUCED_FOOD COOK eng-30-01664172-v: {cook:3; fix:15; ready:4; make:28; prepare:5}, ‘prepare for eating by applying heat’
Absorb_heat ENTITY CONTAINER HEAT_SOURCE eng-30-00375021-v: {boil:3}, ‘come to the boiling point and change from a liquid to vapor’	Apply_heat FOOD COOK CONTAINER HEAT_INSTRUMENT eng-30-00328128-v: {boil:1}, ‘immerse or be immersed in a boiling liquid, often for cooking purposes’

Table 6: Frames for verbs of transformation

Inchoative	Causative
Frames denoting transformation	
Verbs of Creation and Transformation (26)	
Undergo_transformation	Cause_change
ENTITY	ENTITY
INITIAL_CATEGORY	INITIAL_CATEGORY
FINAL_CATEGORY	FINAL_CATEGORY
	INITIAL_VALUE
	FINAL_VALUE
	AGENT

правителство]_{CR_{ENT}}. (BulSemCor)
government.

‘The president forms a caretaker government.’

- b. Cooking_creation: [NP.Ext]_{COOK} [NP.Obj]_{FOOD}
Вкъщи [*тя*]_{COOK} *си* **НАПРАВИ** [*кафе*]_{FOOD}.
At home she herself made coffee.
‘At home she made herself a coffee.’

5.5 Discussion

Observations on the data show some discrepancies in the information represented in different resources. The lack of conformity can be due to various factors:

- (a) Some frames describe verbs in which the change is necessarily caused by an external agent (AGENT or CAUSE) and cannot take place spontaneously (e.g., frame Creating), thus the frame is only causative and does not enter a causative/inchoative frame pair.
- (b) Some gaps are due to systemic factors and the imbalance between causative and inchoative verbs where certain verb has no counterpart (blocked by the verb’s meaning, e.g. the inchoative frame Coroding has no causative correspondence).

3 The complex conceptual structure of verbs of change

- (c) The absence of certain frames in the frame structure might be due to the incompleteness in the frame system, e.g. frames such as *Becoming_wet*, *Becoming_sharp*, etc. would fulfill gaps in the inchoative as correspondences of existing frames *Cause_to_be_wet* and *Cause_to_be_sharp*. However, in some cases there is only a limited number of lexical units and introducing a new frame might be impractical.

In some cases, introducing new frames in order to ensure ballanced representation in resources might be helpful for a given language with a well-developed class of specific verbs, while not necessary for another language where this class of verbs is not present, e.g., in the Princeton WordNet there is a large set of verbs for cooking based on the MANNER, such as *simmer*, *braise*, *coddle*, or HEATING_INSTRUMENT involved, such as *griddle*, *barbecue*, *microwave*, etc. (direct or indirect hyponyms of the synset eng-30-00322847-v: {cook:1}, ‘transform and make suitable for consumption by heating’), these are without lexicalisation in Bulgarian and a detailed set of frames profiling either the MANNER or the HEATING_INSTRUMENT, may not be needed.

The analysis of the available asymmetries on the one hand sheds light onto the set of semantic features that can give rise to a comprehensive classification of verbs of change, and on the other hand gives evidence for the formulation of new frames with regard to the Bulgarian language.

A set of principles can be derived for the consistent semantic description of verbs of change through semantic frames: (i) parallelisation of corresponding areas of the lexical hierarchy (such as causative and inchoative verbs of change) in WordNet and linking them with appropriate relations where necessary; (ii) detection of inconsistencies and gaps in the hierarchical structure in each of the two resources such as frames which are not defined but whose existence is predicted by the general structure of FrameNet (e.g., causative or inchoative correlates of existing frames); frames that adequately reflect the level of specialisation and concretisation of meanings; frames describing parts of the vocabulary not yet covered, etc.

The significant alignment between Levin’s syntactically oriented classification, the hierarchical organisation of the verbs of change in the lexical-semantic network WordNet, and the corresponding system of frames in FrameNet points out to the relevance of the identified key semantic features. Further, observations on Bulgarian data, as well as the examples presented in Section 5, show that valence patterns from English are at least to some degree applicable to Bulgarian, and possibly to other languages. The valence patterns considered here cover the

Ivelina Stoyanova

most typical (with high frequency in the annotated examples) realisation of the configurations of core frame elements for each frame. The particular syntactic patterns in the realisation of frame elements exhibit language-specific features and require procedures for validation.

6 Conclusion

The argument expression possibilities of verbs of change appear to be determined by their lexicalised property – the property subject to change, together with a set of semantic features (such as the type of change – causative or inchoative, quantised or non-quantised, as well as the restrictions on the participants in the situation, etc.). Although argument expression cannot be handled by purely aspectual non-lexical theories of argument projection, verb aspect still plays a key role in the realisation of arguments and the alternations that the verb enters.

The current study outlines some of the main specific features of the verbs denoting change and does not aim at completeness and extensive coverage of all semantic classes. As is evident from the data, the class of verbs of change covers a wide range of semantically diverse verbs. More detailed analysis is required to uncover the specific features of certain subclasses within the class and be able to fully describe their syntactic realisation and alternations.

A contribution of the present study is the proposed classification of verbs of change in WordNet, which goes beyond classifying lexical units (single verbs) to classifying synsets (sets of verbs with the same lexical meaning). This is in line with the assumption that lexical meaning and semantic features determine the syntactic behaviour of verbs. On the other hand, differences in the realisation of verbs from the same WordNet synset lead to the conclusion that additional valence description adds on the semantics of the verb in order to define its syntactic behaviour. Thus, verb-specific level of representation of syntactic patterns is needed in order to obtain a more comprehensive description of the verb classes.

Employing semantic frames in order to present the semantic and conceptual description of synsets in WordNet facilitates the parallel study of both semantically related words (via WordNet relations) and their corresponding conceptual descriptions (from FrameNet frames). Frame-to-frame relations provide means for aligning lexical-semantic relations to conceptual relations and studying the main features that influence the syntactic realisation of frame elements across different frames.

Research in this direction can also contribute to the enrichment of both WordNet and FrameNet and improving their structure to accommodate the complex semantic structure of verbs denoting change.

3 The complex conceptual structure of verbs of change

Further, as already discussed, WordNet also facilitates cross-linguistic analyses and transfer of information from one language to another and can be used as means to expand conceptual and semantic resources for less-resourced languages such as Bulgarian.

Abbreviations

Age	AGENT	Mns	MEANS
CNI	Constructional null instantiation	N or n	Noun
CrEnt	CREATED_ENTITY	NP	Noun phrase
Cse	CAUSE	Pat	PATIENT
DNI	Definite null instantiation	Pie	PIECES
Ent	ENTITY	PP	Prepositional phrase
Finq	FINAL_QUALITY	PWN	Princeton WordNet
Food	PRODUCED_FOOD	Rec	RECIPIENT
INI	Indefinite null instantiation	Src	SOURCE
Ins	INSTRUMENT	V or v	Verb
		Who	WHOLE
		WhoPat	WHOLE_PATIENT

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

References

Andreychin, Lyubomir. 1944. *Basic Bulgarian grammar (in Bulgarian)*. Sofia: Hemus Publishing House.

Baker, Collin F. & Josef Ruppenhofer. 2002. FrameNet’s frames vs. Levin’s verb classes. In Julie Larson & Mary Paster (eds.), *Proceedings of the 28th annual meeting of the Berkeley Linguistics Society*. Berkeley, CA: BLS. DOI: 10.3765/bls.v28i1.3816.

Beavers, John. 2008. Scalar complexity and the structure of events. In Johannes Dölling, Tatjana Heyde-Zybatow & Martin Schäfer (eds.), *Event structures in linguistic form and interpretation*, 245–265. Berlin: Mouton de Gruyter, Berlin.

Ivelina Stoyanova

- Beavers, John & Andrew Koontz-Garboden. 2012. Manner and result in the roots of verbal meaning. *Linguistic Inquiry* 43. 331–369.
- Boas, Hans C. 2010. The syntax-lexicon continuum in construction grammar: A case study of English communication verbs. *Belgian Journal of Linguistics* 24. 54–82.
- Charalozova, Katya. 2021. The category of verb aspect and methodological guidelines for teaching it at Bulgarian schools abroad (in Bulgarian). *Bulgarian Language* 68. 13–21.
- Chomsky, Noam. 1986. *Knowledge of language*. New York: Praeger.
- Dowty, David. 1979. The semantics of aspectual classes of verbs in English. In *Word meaning and montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*, 37–132. Dordrecht: D. Reidel. DOI: 10.1007/978-94-009-9473-7_2.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.
- Dux, Ryan. 2020. *Frame-constructional verb classes: Change and theft verbs in English and German* (Constructional Approaches to Language 28). Amsterdam/Philadelphia: John Benjamins.
- Fellbaum, Christiane (ed.). 1999. *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Fillmore, Charles J. 1968. The case for case. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 1–88. New York: Holt, Rinehart & Winston.
- Fillmore, Charles J. 1970. The grammar of hitting and breaking. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English transformational grammar*, 120–133. Waltham, MA: Ginn.
- Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Publishing.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni Di Semantica* 6. 222–254.
- Fillmore, Charles J. 2012. Encounters with language. *Computational Linguistics* 38. 701–718. DOI: DOI:10.1162/COLI_a_00129.
- Fillmore, Charles J. & Collin F. Baker. 2009. A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 313–340. Oxford: Oxford University Press.
- Fillmore, Charles J., Collin F. Baker & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the Conference COLINGACL '98, Montreal, Canada*, 86–90.

3 The complex conceptual structure of verbs of change

- Fillmore, Charles J., Christopher R. Johnson & Miriam R. Petruck. 2003. Background to framenet. *International Journal of Lexicography* 16(3). 235–250. DOI: 10.1093/ijl/16.3.235.
- Fillmore, Charles J., Paul Kay & Catherine O'Connor. 1988. Regularity and idiomatcity in grammatical constructions: The case of let alone. *Language* 64. 501–538.
- Haspelmath, Martin. 1993. More on the typology of inchoative/causative verb alternations. In Bernard Comrie & Maria Polinsky (eds.), *Causatives and transitivity* (Studies in Language Companion, Series 23), 87–120. Amsterdam: John Benjamins.
- Hay, Jennifer, Christopher Kennedy & Beth Levin. 1999. Scalar structure underlies telicity in degree achievements. In Tanya Matthews & Devon Strolovitch (eds.), *Proceedings of the 9th semantics and linguistic theory (SALT 9)*, Ithaca, NY, 127–144. CLC Publications: Cornell University.
- Hobbs, Jerry R. 2008. Deep lexical semantics. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics*.
- Hobbs, Jerry R. 2014. Case, constructions, FrameNet, and the deep lexicon. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929–2014)*, 10–12.
- Ivanchev, Svetomir. 1971. *Problems of the aspect in Slavic languages (in Bulgarian)*. Sofia: Bulgarian Academy of Science.
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Kennedy, Christopher & Beth Levin. 2008. Measure of change: The adjectival core of verbs of variable telicity. In Louise McNally & Christopher Kennedy (eds.), *Adjectives and adverbs in semantics and discourse*, 156–182. Oxford: Oxford University Press.
- Kipper-Schuler, Karin. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania. (Doctoral dissertation).
- Koeva, Svetla. 2007. BulNet (lexical-semantic network of the Bulgarian language) – part of the world lexical-semantic network (in Bulgarian). *Bulgarian Language* 1. 19–32.
- Koeva, Svetla. 2010a. *Bulgarian FrameNet (in Bulgarian)*. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2010b. Bulgarian WordNet – Current State, Applications and Prospects. In *Bulgarian-American Dialogues*, 120–132. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Koeva, Svetla. 2011. Verb aspect, alternations and quantification. *Cognitive Studies / Études cognitives* 11. 125–134.

Ivelina Stoyanova

- Koeva, Svetla. 2012. Bulgarian sense-annotated corpus – between the tradition and novelty. *Cognitive Studies / Études cognitives* 12. 181–198.
- Koeva, Svetla. 2021. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla & Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the fifth international conference computational linguistics in bulgaria*, 203–214. <https://aclanthology.org/2022.clib-1.23>.
- Koeva, Svetla, Svetlozara Leseva, Borislav Rizov, Ekaterina Tarpomanova, Tsvetana Dimitrova, Hristina Kukova & Maria Todorova. 2011. Design and development of the Bulgarian sense-annotated corpus. In María Luisa Carrió Pastor & Miguel Ángel Candel Mora (eds.), *Proceedings of the III international congress of corpus linguistics. Information and communications technologies: Present and future in corpus analysis*, 143–150.
- Koeva, Svetla, Svetlozara Leseva & Maria Todorova. 2006. Bulgarian sense tagged corpus. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, 79–86.
- Krifka, Manfred. 1998. The origins of telicity. In Susan Rothstein (ed.), *Events and grammar*, 197–235. Dordrecht: Kluwer.
- Kutsarov, Ivan. 2007. *Theoretical grammar of Bulgarian. Morphology (in Bulgarian)*. Plovdiv: Publishing house of Plovdiv university.
- Landes, Shari, Claudia Leacock & Randee Tengi. 1998. Building semantic concordances. In Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*. Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/7287.003.0015.
- Leseva, Svetlozara & Ivelina Stoyanova. 2019. Structural approach to enhancing WordNet with conceptual frame semantics. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 629–637. Varna: INCOMA.
- Leseva, Svetlozara & Ivelina Stoyanova. 2020. Beyond lexical and semantic resources: Linking WordNet with FrameNet and enhancing synsets with conceptual frames. In Svetla Koeva (ed.), *Towards a semantic network enriched with a variety of semantic relations*, 21–48. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Leseva, Svetlozara & Ivelina Stoyanova. 2021. Semantichno opisanie na glagoli za promyana i yerarhichna organizatsiya na kontseptualnite freymove. In vol. 2.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Levin, Beth. 2010. *Lexicalized scales and verbs of scalar change*. <https://web.stanford.edu/~bclewin/cls10change.pdf>.

3 The complex conceptual structure of verbs of change

- Levin, Beth. 2015. Verb classes within and across languages. In Bernard Comrie & Andrej Malchukov (eds.), *Valency classes in the world's languages*, 1627–1670. Berlin: De Gruyter.
- Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge: MIT Press. 350.
- Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization*. Cambridge: Cambridge University Press.
- Mel'cuk, Igor. 1967. K ponjatiu slovoobrazovanija. *Izvestiya Akademii Nauk SSSR. Serija literatury i jazyka* 26. 352–362.
- Miller, George A. 1995. WordNet: A lexical database for English. *Commun. ACM* 38(11). 39–41.
- Miller, George A., Martin Chodorow, Shari Landes, Claudia Leacock & Robert G. Thomas. 1994. Using a semantic concordance for sense identification. In *Proceedings of a Workshop on Human Language Technology, Plainsboro, New Jersey, 1993*. <https://aclanthology.org/H93-1061>.
- Miller, George A. & Christiane Fellbaum. 2007. WordNet then and now. *Language Resources and Evaluation* 41. 209–214.
- Miller, George A., Claudia Leacock, Randee Teng & Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of a Workshop on Human Language Technology, Plainsboro, New Jersey, 1993*. <https://aclanthology.org/H93-1061>.
- Nitsolova, Ruselina. 2008. *Bulgarian grammar. Morphology (In bulgarian)*. Sofia: Publishing house of Sofia University.
- Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge: MIT Press.
- Rappaport Hovav, Malka. 2008. Lexicalized meaning and the internal temporal structure of events. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*, 13–42. Amsterdam: John Benjamins.
- Rappaport Hovav, Malka & Beth Levin. 2005. Change of state verbs: Implications for theories for argument projection. In Nomi Erteschik-Shir & Tova Rapoport (eds.), *The syntax of aspect*, 274–286. Oxford, UK: Oxford University Press.
- Rappaport Hovav, Malka & Beth Levin. 2010. Reflections on manner/result complementarity. In Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds.), *Syntax, lexical semantics, and event structure*, 21–38. Oxford: Oxford University Press.
- Rosca, Andreea. 2013. Framenet and its limitations. The case of entity-specific change-of-state verbs. *Journal of English and American Studies* 47. 13–30.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin. F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.

Ivelina Stoyanova

- Van Valin, Robert D. 1993. Role and reference grammar. *Work Papers of the Summer Institute of Linguistics, University of North Dakota* 37. 65–75. DOI: [10.31356/silwp.vol37.05](https://doi.org/10.31356/silwp.vol37.05).
- Van Valin, Robert D. 2005. *Exploring the syntax–semantics interface*. Cambridge: Cambridge University Press.
- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Vendler, Zeno. 1957. Verbs and times. *Philosophical Review* 66(2). 143–160.
- Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

Chapter 4

The conceptualisation of the route: Non-directed and directed motion verbs in Bulgarian and English

Svetlozara Leseva 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

This chapter offers an analysis of non-directed and directed motion from a frame semantics perspective through studying the semantic description and syntactic realisation of the frame elements of several semantic frames in FrameNet. The study is focused on the conceptualisation and syntactic expression of the elements of the route along which motion occurs: GOAL (the final part of the route), SOURCE (the initial part of the route) and PATH (the middle part of the route) in English and Bulgarian by studying the syntactic expression of the corresponding frame elements in FrameNet. The research questions explored in the chapter deal with the prominent aspects in the semantics of the verbs evoking a particular semantic frame, the syntactic expression of the relevant frame elements, syntactic explicitness and implicitness. The empirical evidence provided by the FrameNet corpus is compared with a sample of annotated Bulgarian examples. The observations made throughout the chapter are brought in the perspective of linguistic hypotheses put forward in the literature: in particular, the goal-over-source hypothesis and the proposal that motion verbs tend to co-occur with expressions that align with the part of the trajectory of motion that is most prominent in their semantics.

1 Introduction

This chapter deals with the semantic and syntactic description of motion verbs in Bulgarian (as compared with English) with respect to: their semantics as de-



Svetlozara Leseva. 2024. The conceptualisation of the route: Non-directed and directed motion verbs in Bulgarian and English. In Svetla Koeva (ed.), *Universality of semantic frames and language specific Bulgarian data*, 171–232. Berlin: Language Science Press. DOI: ??

scribed in terms of semantic frames; the conceptualisation of parts of the trajectory of motion and the corresponding frame elements; the syntactic realisation of the major frame elements as reflected in corpora.

The study is based on the description of verbs in FrameNet (Fillmore et al. 1998) as lexical units evoking particular frames, defined themselves as schematic descriptions of situations in terms of the configurations of participants and props that constitute their meaning. The syntactic description will be focused on the main patterns of syntactic expression of the most essential frame elements in and across the selected motion frames. The proposed description aims at capturing the semantic and syntactic properties of Bulgarian verbs of motion against a more universal background.

The analysed verbs are selected from the Bulgarian WordNet (Koeva 2021a), which were then associated with FrameNet frames (Leseva & Stoyanova 2020a) and further aligned with verbs in other resources, where possible (Leseva & Stoyanova 2022a). As a result, the verb synsets in the Bulgarian WordNet are mapped to FrameNet frames (one frame per synset), making it possible for observations to be made on the basis of the semantic description available for English verbs. The FrameNet model has been widely adopted for building similar descriptions of the lexis of a number of typologically diverse languages – German (Burchardt et al. 2006), Dutch (Vossen et al. 2018), Danish (Pedersen et al. 2018), Swedish (Borin et al. 2010), Latvian (Gruzitis et al. 2018), French (Candito et al. 2014), Spanish (Subirats 2009), Brazilian Portuguese (Torrent et al. 2018), Chinese (You & Liu 2005), Japanese (Ohara et al. 2004), Korean (Hahm et al. 2014), among others (for a more comprehensive description of the existing framenets and the Multilingual FrameNet annotation initiative¹, cf. Gilardi & Baker (2018)). FrameNet's theoretical framework has been adopted for Bulgarian and extended into a model accounting for language-specific features, including verb aspect, semantic and syntactic diatheses and syntactic alternations. The concept was implemented in the development of the Bulgarian FrameNet (Koeva 2008, 2010, 2021b).

This chapter will specifically address: (i) those aspects in the semantics of the verbs evoking the studied frames that are cast as any of the frame elements describing the motion of an entity along a trajectory; (ii) the syntactic expression of the relevant frame elements and the conditions predetermining their syntactic explicitness or implicitness. The empirical evidence provided by the examples in the FrameNet corpus will be studied against a sample of annotated Bulgarian examples, thus testing the cross-lingual validity of the theoretical and practical observations and drawing parallels or distinctions, where appropriate. The observations made throughout the chapter will be analysed in the perspective of

¹www.globalframenet.org/

4 *The conceptualisation of the route: Non-directed and directed motion*

linguistic hypotheses that have been put forward in the literature: in particular: the goal-over-source hypothesis and the proposal that motion verbs tend to co-occur with expressions that align with the most prominent aspect of the trajectory of motion encoded in their semantics.

2 Motion verbs

The semantic representation of motion verbs has been the focus of multitude of studies. One of the major distinctions in the verb lexis is the one between manner and result, which are usually viewed as complementary notions, i.e. verbs lexicalise either one or the other (Levin 2015). In the domain of motion this differentiation criterion takes the form of a distinction between the expression (and possibly the conflation) of manner and path. This distinction has been extensively studied by Talmy (1985, 1991, 2000), who offered a typology of the so-called verb-framed and satellite-framed languages, characterising languages in terms of the lexicalisation patterns of motion events: in satellite-framed languages verbs usually encode manner, while the path of the movement is encoded outside the verb (base) by satellites such as adverbial particles (but also prepositions and prefixes); in verb-framed languages the path is expressed by the verbs and manner is either omitted or realised by means of an adverbial expression. The discovery of finer typological distinction across languages with respect to motion expressions has led to the refinement of the original Talmian typology in the works of a number of authors (Aske 1989, Slobin 1996, Papafragou et al. 2002, Ibarretxe-Antuñano 2004, Slobin 2004, Filipović 2007, Beavers et al. 2010, Croft et al. 2010), among others. The interest in the elements that make up the trajectory, or path, of the motion (Jackendoff 1983: 162, Talmy 1985: 57, Lakoff 1987: 275) has been reflected in numerous studies on the lexical encoding and syntactic expression of these elements in co-occurrence with the verb (Rohde 2001, Rakhilina 2004, Stefanowitsch & Rohde 2004, Cristobal 2010, Kopecka 2010), among others. A related line of research has been the study of the bias with respect to the expression of one path element over another in and across languages (Ikegami 1987, Dirven & VerSpoor 1998, Stefanowitsch & Rohde 2004, Wälchli & Zúñiga 2006, Verkerk 2017), among others.

The distinction between manner and path of motion and the expression and profiling of different sections of the path, have been the prime focus of many other studies. For instance, Viberg (2015) proposes a verb typology with respect to the expression of the endpoint of motion in Swedish in comparison with English, German, French and Finnish. In her study Kopecka (2010) explores lexicalisation patterns of manner of motion verbs in Polish, while Łozinska (2018) delves

into the expression of path and manner in Polish and Russian in contrastive terms. Taremaa (2017, 2021) has explored motion verbs in Estonian, focusing on the expression of source, goal, path, location and direction with both manner of motion verbs and source- and goal-profiling verbs.

Various authors have previously adopted the FrameNet approach in the analysis of motion verbs. Viberg (2008) proposes a study of Swedish verbs of motion in a vehicle; the verbs have been analysed in a cross-linguistic perspective and with respect to their encoding in FrameNet. Cristobal (2010) provides a detailed analysis of Arriving verbs in English and Spanish. Imani & Naeini (2020) study a selection of manner of motion verbs in English and contrast them with their counterparts in Persian.

A number of studies in these lines of research have been dedicated to Bulgarian motion verbs. Tchizmarova (2015) analyses several verbs with respect to the way they divide the space of linear motion, including the co-occurrence with directional phrases. Lindsey (2011) and Speed (2015) explore the preference for and distribution of manner and path verbs in Bulgarian in contrast with other Balkan and Slavic (Balkan and non-Balkan) languages and come to the conclusion that, as suggested for Modern Greek, Bulgarian does not conform to one of the two Talmian typological patterns of conflating motion. In her work Pantcheva (2007a,b) centres on prefixation involving directional prefixes in Bulgarian and how this process affects event structure and syntactic structure, as part of a cross-linguistic study on directional expressions (Pantcheva 2011).

A small number of FrameNet-based studies dealing with Bulgarian motion verbs have also been published, usually focusing on a small selection of predicates and their description in FrameNet, possibly supported by corpus data. For instance, Dekova & Nestorova (2010) offer an analysis of Bulgarian verbs of non-directed motion, while Nestorova (2009) discusses several transitive verbs involving the relocation of masses of people (*populate* verbs).

This chapter's contribution lies in delivering an analysis of a selection of a non-directed and directed motion verbs in Bulgarian as compared with their English counterparts implemented through the adoption of the descriptive devices developed within the Berkeley FrameNet project and applying them to Bulgarian. The proposed methodology provides a solid foundation for cross-linguistic study of the semantic and syntactic properties of verbs.

4 *The conceptualisation of the route: Non-directed and directed motion*

3 The organisation of FrameNet

3.1 Semantic frames and frame elements

FrameNet (Fillmore et al. 1998, Baker 2008) is a lexical resource which couches lexical and conceptual knowledge in the theory of Frame Semantics (Johnson et al. 2001, Fillmore 2003, Ruppenhofer et al. 2016). A semantic frame is a “script-like structure of inferences, linked by linguistic convention to the meanings of linguistic units – in our case, lexical items. Each frame identifies a set of frame elements (FEs) – participants and props in the frame. A frame semantic description of a lexical item identifies the frames which underlie a given meaning and specifies the ways in which FEs, and constellations of FEs, are realized in structures headed by the word” (Johnson et al. 2001: 9). 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 semantic frame: core, peripheral or extra-thematic (Ruppenhofer et al. 2016: 19–20). A core FE is “one that instantiates a conceptually necessary component of a frame, while making the frame unique and different from other frames” (Ruppenhofer et al. 2016: 23). 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, either of an actual event or state of the same type (e.g. the FE ITERATION), or by evoking a larger frame within which the reported state of affairs is embedded. A frame in FrameNet is linked to the meanings of a set of linguistic items, called lexical units (LUs). Each LU is thus a pairing of a word and a meaning whose conceptual semantics is represented by the frame (so that the LU is said to evoke the relevant frame). Below, reference will be made mainly to core FEs as the ones that are most essential to the description of the different frames.

The observations presented below are based on the data in the Berkeley FrameNet requested in 2017. For the sake of consistency, in the course of this work the data have been checked against the online version of the resource.²

3.2 Frame-to-frame relations

FrameNet frames are organised in a network by means of a number of hierarchical and non-hierarchical frame-to-frame relations (Ruppenhofer et al. 2016:

²The official Berkeley FrameNet has migrated to: <http://berkeleyfn.frameenetbr.ufjf.br/>. The online searchable database is available for browsing at <https://framenet2.icsi.berkeley.edu/>.

81–84). Four of them are hierarchical relations that bear relevance to the internal organisation of a particular semantic domain of the lexis and will be discussed below. **Inheritance** is a relation between a more general (parent) frame and a more specific (child) frame where “each semantic fact about the parent must correspond to an equally specific or more specific fact about the child” (Ruppenhofer et al. 2016: 81–82), i.e. there should be a strict correspondence between entities, frame elements, frame relations and semantic characteristics in the parent and the child frame (Petruck 2015). Examples of this relation in the context of the studied domain are represented by the frames *Self_motion*, *Fluidic_motion*, etc. (see Fig. 1), which share the main configuration of frame elements defined for the 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*). **Using**, also defined as weak **Inheritance** (Petruck 2015), is a relation between a parent frame and a child frame in which only some of the FEs in the parent have a corresponding entity in the child, and if such exist, they are more specific (Petruck & de Melo 2012). In the domain of motion such a relation exists between *Motion* and its child *Operate_vehicle*, among others. Like *Motion*, the more specific frame describes the translational motion of a **THEME** from a **SOURCE** to a **GOAL** along a **PATH**, but it elaborates on it by introducing further frame elements: an **AGENT**, who controls the movement, and a **CARRIER**, which is the actual object carrying the **THEME**. **Perspective** is a relation where a more abstract situation viewed as neutral may be specified by means of perspectivised semantic frames that represent “different possible points-of-view on the neutral frame” (Ruppenhofer et al. 2016: 82). For instance, the frames *Operate_vehicle* and *Ride_vehicle* perspectivise different facets of the idea of moving by means of a vehicle described in *Using_vehicle* according to the involvement of a person, who is being transported, as either the driver/operator or as a passenger. **Subframe** captures the relationship between a complex frame referring to “sequences of states and transitions, each of which can itself be separately described as a frame” and the frames denoting these states or transitions (Ruppenhofer et al. 2016: 83–84). For example, the frames *Arriving* and *Departing* are defined as subframes of *Traversing*, as they describe the initial and the final stage of the translational movement that results in a moving entity’s change of location.

A comprehensive description of all the frame-to-frame relations with more examples is provided in Ruppenhofer et al. (2016).

4 English and Bulgarian data employed in the study

4.1 FrameNet and WordNet as a source for the inventory of motion verbs

The inventory of English verbs and their semantic and syntactic description used in the study is directly derived from the description of the lexical units in the studied semantic frames in the Berkeley FrameNet, as well as the lattices summarising the valence patterns attested in the FrameNet corpus, including the particular syntactic realisation of the FEs in terms of their syntactic category and syntactic function. The corpus is also used as a source for the examples illustrating the realisation of the English verbs³.

The semantic frames are adopted from the Berkeley FrameNet without changes, but where relevant, comments regarding the set of frame elements are made. The Bulgarian verbs are studied independently but in comparison with their English counterparts, taking as a point of departure the relevant motion frames and the valence patterns described in the Berkeley FrameNet. This approach has been adopted to facilitate the description of the motion verbs in the Bulgarian FrameNet whose original concept was laid out in Koeva (2008, 2010) and further elaborated in Koeva (2021b), as well as in Chapter 1, this volume. The Bulgarian FrameNet is implemented within BulFrameNet (Koeva & Doychev 2022), a dedicated web-based system allowing the comprehensive description of the semantic and syntactic properties of verbs. The study of the valence patterns of the motion verbs and the syntactic expression of their semantic participants presented below was undertaken specifically as part of the work on the Bulgarian FrameNet.

The set of Bulgarian motion verbs discussed in the chapter is extracted from the Bulgarian WordNet, a lexical-semantic net modelled on the Princeton WordNet (PWN). PWN (Miller 1995, Fellbaum 1999) is a large lexical database for English that represents comprehensively conceptual and lexical knowledge in the form of a network whose nodes denote cognitive synonyms (synsets) connected through a number of conceptual-semantic and lexical relations such as hypernymy, meronymy, antonymy, etc. The synsets in the Bulgarian Wordnet have been developed by translation and adaptation of the PWN counterparts, and the corresponding synsets in the two wordnets are related to each other through unique interlingual identifiers (which also provide links to the respective synsets in all other wordnet that support them). Thus, the lexical and conceptual knowledge is aligned cross-linguistically. In the course of its creation the Bulgarian WordNet has been expanded so as to cover all the synsets included

³For brevity some of the examples throughout the paper will be adapted.

in PWN (117,659 in total) by means of automatic translation followed by manual editing and enrichment (currently 85,954 synsets have been manually validated). The Bulgarian WordNet includes language-specific lexicalisations (synsets with no correspondence in PWN) as well as synsets describing closed-class words: prepositions, conjunctions, pronouns, particles, interjections; as a result it has amounted to 121,282 synsets altogether. It thus provides substantial coverage of the language's lexis, including verbs (forming a total of 14,103 synsets). In addition, BulNet has developed a number of language-specific characteristics, including notation of verb aspect. The current state of the Bulgarian WordNet is detailed in Koeva (2021a)⁴.

WordNet and FrameNet were aligned automatically using several previous mappings coupled with additional procedures for expansion and validation. In particular, the following were employed: (i) direct mappings provided within FrameNet (Baker & Fellbaum 2009), eXtendedWordFrameNet (Laparra & Rigau 2010) and MapNet (Tonelli & Pighin 2009), supplemented with (ii) indirect mapping through VerbNet (Palmer 2009, Palmer et al. 2014). This resulted in 4,306 unique WordNet synsets to FrameNet frame mappings, a coverage of 30.5% of the verb synsets (Leseva & Stoyanova 2020b: 110). A number of procedures inspired by ideas proposed in de Lacalle et al. (2014) and especially in Burchardt et al. (2005) were implemented towards the improvement and extension of the mapping coverage. These procedures, described in Leseva et al. (2018) and further refined in Leseva & Stoyanova (2019, 2020a), are specifically based on exploring the structural properties of the two resources, such as: (i) the assumption that as verbs in a synset denote the same or very similar meaning, they are likely to evoke the same semantic frame; (ii) the hierarchical relational structure of the two resources based on the notion of inheritance from a more general to a more specific synset or frame. As a result, in general, more specific concepts should be associated with the frame of their hypernym(s) or with more specific frames elaborating on (and possibly inheriting from) this frame, although various divergences occur in practice. Part of the other relations among frames and among synsets were also cast as validation procedures. The main mapping mechanism involved: (i) manual assignment of semantic frames to root verb synset to ensure greater accuracy at the highest hierarchical level and to reduce error propagation down the tree; (ii) automatic assignment of the hypernym's frame onto hyponyms which were not previously mapped; and (iii) verification and improvement of the assignments by applying the validation procedures. In this way, the coverage of the automatic mapping has been gradually increased to 94%

⁴The Bulgarian WordNet may be browsed at: dcl.bas.bg/bulnet/.

4 *The conceptualisation of the route: Non-directed and directed motion*

(Leseva & Stoyanova 2020b: 115 – 116). Due to various peculiarities of the structure of WordNet or lack of appropriate frames in FrameNet (as part of the lexis has not yet been described by frames), the automatic assignment has been undergoing manual validation, so far covering almost 50% of the mapping (over 6,000 synsets).

The FrameNet-to-WordNet alignment together with the alignment between the Princeton WordNet and the Bulgarian WordNet has enabled the association of Bulgarian verbs with a FrameNet semantic description. This possibility is founded on the assumption that although the construal of the semantic description of situations across languages (as well as across resources) often differs, the major semantic aspects represent shared conceptual knowledge. Such an assumption underlies the development of both wordnets and framenets for other languages besides English, as well as the Global FrameNet initiative and Shared Annotation Task (cf. Section 1). The genealogical and typological similarities between English and Bulgarian have also made it possible to base the syntactic description of the Bulgarian verbs of motion on the one provided for their English counterparts in the Berkeley FrameNet. Similar ideas have been pursued by other authors who have adopted a FrameNet-oriented approach to the semantic and syntactic analysis for languages other than English (cf. Section 2). The analysis below has been specifically informed by previous work on Bulgarian change (Stoyanova & Leseva 2021) and stative verbs (Leseva & Stoyanova 2022b).

The English and the Bulgarian verbs included in the analysis are members of synsets that have been assigned one of several FrameNet frames belonging to the motion domain. In order to be selected, they had to meet the following requirements: (i) pertain to the general lexis; (ii) have a corresponding LU in FrameNet with a sufficient number of annotated sentences (20+). This means that synsets such as {walk:1} (‘use one’s feet to advance; advance by steps’) and {run:34} (‘move fast by using one’s feet, with one foot off the ground at any given time’), which have as correspondences the LUs *walk.v* and *run.v* in the *Self_motion* frame are included in the analysis, while ones such as {lollop:1} (‘walk clumsily and with a bounce’) and {hare:1} (‘run quickly, like a hare’) are not. These requirements have been adopted for the following reasons: general-lexis verbs are more likely to be represented in BulSemCor (see Section 4.2), hence more Bulgarian examples would be available for them; the representation in the FrameNet corpus provides the pool of examples for English.

4.2 Annotated examples

The statistics and analysis for English are based on the annotated sentences available for the respective verbs in the Berkeley FrameNet.

The examples of the usage of the selected Bulgarian verbs are extracted from BulSemCor (Koeva et al. 2006, 2011) – a 100,000-word corpus designed according to the overall methodology of the English SemCor (Miller et al. 1993, 1994, Landes et al. 1998), further adapted by using criteria for ensuring an appropriate coverage of contemporary Bulgarian general lexis. As BulSemCor is manually annotated with wordnet senses, it provides disambiguated examples for the studied verbs. Where the number of examples is not sufficient, they have 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 have been selected manually so as to correspond to the studied senses.

The Bulgarian examples extracted from the different corpora have been annotated so that the sentence components realising syntactically the core frame elements related to motion are 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 the selected semantic frames: Motion – 149; Self_motion – 262; Arriving – 182; Departing – 178; Traversing – 122.

5 The domain of Motion

5.1 Organising semantic domains

As suggested by Johnson et al. (2001: 16), the lexicon pertaining to a semantic domain is hierarchically organised in a number of semantic frames of different abstraction and specialisation related through the frame-to-frame relations that capture semantic generalisations existing across frames. Thus, as pointed out in the work cited, for many semantic domains, there is one general frame that describes the common aspects of the more specific frames. It may be posited that at the conceptual level all (or most) frames in a domain share this basic structure consisting of a configuration of FEs that defines the distinctive meaning of the domain. The mechanisms that organise such a part of the lexis involve various changes in this prototypical structure that reflect the various ways in which specialisation within the domain occurs: (i) not all frames allow the overt expression

4 *The conceptualisation of the route: Non-directed and directed motion*

of all FEs – some of them may be blocked from overt expression, although they are conceptually necessary and implied in the meaning of the lexical units; (ii) more specific frames may exclude some FEs or demote them to non-core status; for instance, in the GOAL-profiling Arriving and the SOURCE-profiling Departing frames the FEs describing the remaining parts of the route are regarded as non-core; (iii) more specific frames may further narrow down the semantic properties of one or more of the FEs as compared with their counterparts in the more general frame (e.g. impose stricter selectional restrictions on the expressions realising the FEs): e.g. the moving entity is defined as the FE FLUID in the Fluidic_motion frame, and as MASS_THEME in the Mass_motion frame which both inherit from the Motion frame (Fig. 1); (iv) more specific frames may include other FEs besides the ones describing the general frame, may change perspective, incorporate or profile a certain FE. An example of a semantic elaboration resulting in the introduction of a new FE is the specification of the vehicle which “holds or conveys” the traveller (the FE MODE_OF_TRANSPORTATION) in the Travel frame (Fig. 1).

The observations below are based on the theoretical and practical motivations described in Johnson et al. (2001), Petruck & de Melo (2012), Petruck (2015), Ruppenhofer et al. (2016) and the definitions, comments and frame-to-frame relations in FrameNet.

5.2 General organisation of the domain of motion

The lexis denoting movement is most broadly divided between translational and non-translational (or self-contained) motion with respect to some background or location. Based on the definitions in FrameNet⁵, in the first case a moving entity typically starts at some location, moves through space along a trajectory and ends up in another location. Non-translational motion⁶ involves the movement of an entity or parts of it with respect to some fixed location or landmark, without undergoing motion in space or without a significant alteration of configuration or shape. Translational motion is most broadly described by the non-lexicalised frame Motion_scenario, which is further perspectivised by several frames, two of which, Motion and Traversing, form the core of the translational motion domain. Non-translational movement is described by the Moving_in_place frame and its causative counterpart Cause_to_move_in_place, which are evoked by LUs such as *rock*, *shake*, *twirl*, e.g. *The earth shook* vs. *He shook the remote control*. In what follows below, the focus will be on translational motion.

⁵<http://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Motion>

⁶https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Moving_in_place

Svetlozara Leseva

Another major division in the domain of motion is between (i) self-induced motion that a moving entity undergoes on its own – under its own physical power, due to some internal cause, physical forces, features of the relief, etc., and (ii) caused motion that is brought about by an external participant that may be an animate, volitional AGENT or a non-animate CAUSE. The prototypical semantic frame that organises the lexis of this type of translational motion is Cause_motion, which is indirectly integrated in the Motion_scenario through its causative relation to Motion (i.e. Cause_motion Is_causative_of Motion). The frames related to Cause_motion include Bringing (e.g. *bring, carry, transport*), Placing (e.g. *bottle, load, pocket*), Filling (e.g. *fill, flood*), Removing (e.g. *extract, remove*), Emptying (e.g. *empty, purge*), as well as some frames (e.g. Cause_fluidic_motion) that have counterparts in the non-causative domain (Fluidic_motion).

As suggested in the description of the Motion frame, a complex area in the vocabulary of motion is the depiction of the relation of VEHICLES to the moving entity and other participants. In the cases where the moving entity cannot be expressed, the LUs denoting the motion of vehicles are treated as evoking the Self_motion frame and the vehicles are annotated as SELF_MOVERS. When the VEHICLE is profiled as being operated by a DRIVER, the relevant LUs evoke the frame Operate_vehicle; the DRIVER may be construed very generally as the one controlling the vehicle, as attested by verbs such as *bicycle, canoe, paddle, skate*, along with verbs involving special qualifications or skills such as *drive, fly, sail, taxi*. The situation where the moving entities are passengers that are transported by means of the VEHICLE which is not under their control, is described by the Ride_vehicle frame (*bus, hitchhike, ride, sail*).

Another type of elaboration in the motion domain described in the definition of the frame or reflected in the frame-to-frame relational structure refers to properties of the manner of motion, which basically stem from prominent features of the moving entity. One such feature is the requirement for the moving entity to be a living being whose body moves on its own, using its own energy, as in Self_motion (e.g. *jog, limp, run, walk*), and semantic frames inheriting from it such as Cotheme (*accompany, lead, track*) and Travel (*journey, tour, voyage*). Further salient distinctions based on the types of entities involved in the motion and the specific manner of motion typical of them is reflected in the definition of several frames such as: Fluidic_motion (e.g. *cascade, ooze, stream*), describing the motion of liquids; Mass_motion (e.g. *crowd, swarm, throng*), which refers to the motion of a collective of individuals (a MASS_THEME) as one entity; Motion_noise (e.g. *buzz, roar, whirl*), specified according to the type of noise the moving entity produces; Light_movement (e.g. *gleam, shine, twinkle*), describing the emission and movement of light, etc.

4 The conceptualisation of the route: Non-directed and directed motion

Another facet of motion has to do with the feature of directionality, which is lexically encoded in some LUs (e.g. *descend*, *fall*, *rise* in the `Motion_directional` frame). Directed motion is also described in semantic frames that profile parts of the path along which an entity moves, such as its initial (Departing) or final stage (Arriving).

In the remaining part of the chapter the analysis will be focused specifically on non-directed and directed motion verbs as represented by semantic frames such as `Motion` and `Self_motion` on the one hand, and `Traversing`, `Arriving` and `Departing` on the other, drawing parallels between the semantics and syntactic expression of the relevant frame elements.

The overall organisation of the domain of self-induced translational motion is partially represented in Figure 1.

5.3 Motion

5.3.1 Semantic description of the `Motion` frame

The `Motion` frame describes the changing of spatial location of a `THEME` understood in the classical sense of Gruber (Gruber 1965: 27–31, Jackendoff 1972: 29) as an entity that moves. More precisely, the LUs in this frame involve the translational motion of entities that are either not (capable of) moving under their own power or are underspecified for this feature. Therefore, many of the definitions of the motion frames for which this property is relevant note that the `THEME` is frequently a living being moving on its own but need not be. Generalising over FrameNet examples such as the following, one can infer that the motion may be induced by various factors: (i) an outside force: [*The black dust*]_{THEM} *began BLOWING OFF* [*the tailings lake*]_{SRC}; (ii) the `THEME`'s own momentum: *It fell on the floor and* [_{THEM}] *ROLLED* [*towards Uncle Mick's feet*]_{GOAL}; (iii) some internal process: [*Tears*]_{THEM} *ROLLED* [*down my cheeks*]_{PATH}, etc., but it is represented with respect to the involvement of the `THEME` in it, regardless of the cause that has brought it about. Volitional or self-directed motion is elaborated in some of the frames inheriting from `Motion`, `Self_motion` and its descendants in particular. The remaining core FE's of the `Motion` frame describe various elements or properties of the path⁷ that the moving entity moves along.

⁷When used with a capital letter, `PATH` would mean the frame element; in small letters, `path` would be used in its accepted meaning in the literature, i.e. the medial part of the route traversed by a **Figure** (Fillmore 1971: 26, Lakoff 1987: 275, among others). The term “route” will be used instead of “path” to refer to the line of movement that comprises all the three elements: `SOURCE`, `PATH` and `GOAL`.

Svetlozara Leseva

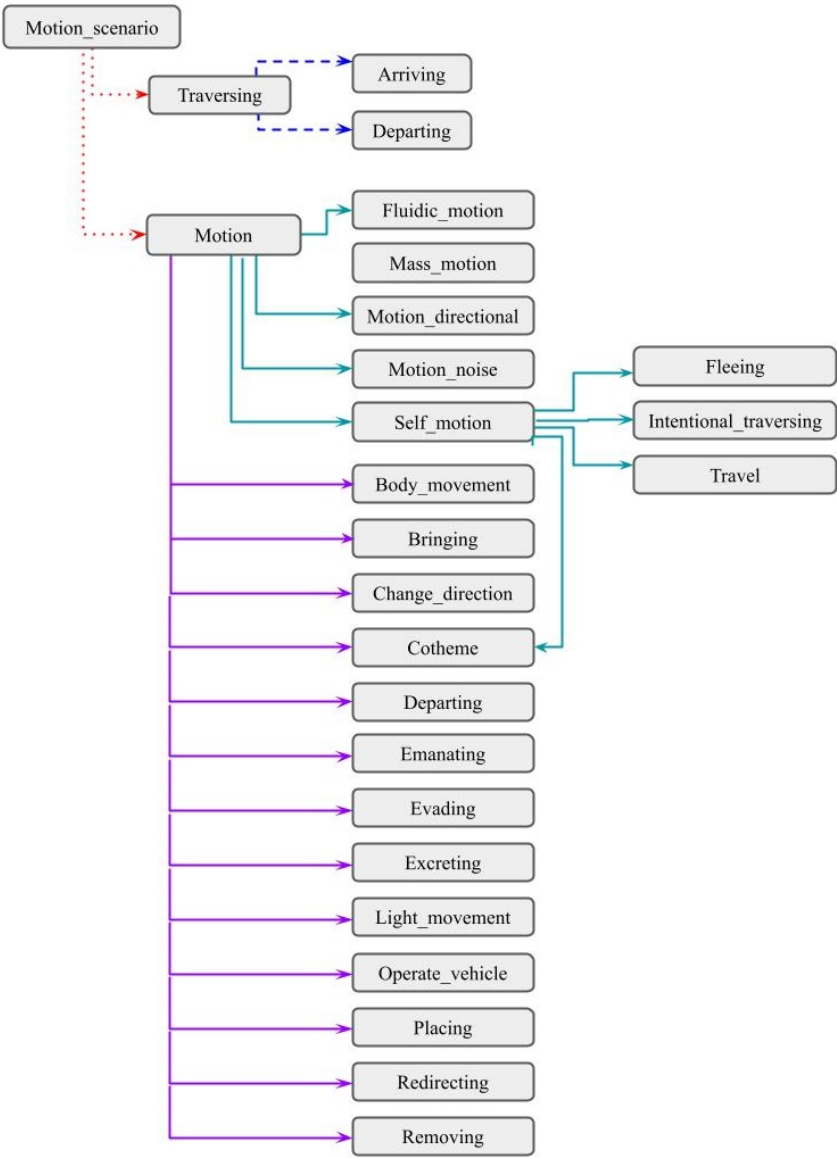


Figure 1: The Motion hierarchy: green lines denote the relation of **Inheritance**; red lines show the relation **Perspectivises**; blue lines correspond to the **Subframe of** relation; magenta lines denote the relation **Using**.

4 *The conceptualisation of the route: Non-directed and directed motion*

Definition of the Motion frame: 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.⁸

Core FEs in the Motion frame: THEME, SOURCE, GOAL, PATH, AREA, DIRECTION, DISTANCE.

The THEME⁹ is defined as “the entity that changes location”. An important feature of this FE noted in the FrameNet description is that it need not be a SELF_MOVER, that is, it need not be capable of moving on its own and by its own power or will. Its semantic specification includes animate beings, physical objects, substances, etc.

The SOURCE is “the location the THEME occupies initially before its change of location”. It may refer to geological and other natural formations, geographical points, celestial bodies; physical objects, including man-made structures, such as buildings, constructions, facilities, other objects occupying space, etc.

The GOAL is “the location the THEME ends up in” as a result of the motion. It has the same semantic specification as the SOURCE.

The PATH refers to “(a part of) the ground over which the THEME travels or to a landmark by which the THEME travels”. Its semantic specification encompasses locations, including geological and other natural formations, geographical points, celestial bodies; physical objects, including man-made structures and other objects occupying space that may be construed as having extent along which the motion takes place; extents of various media, such as water, air, etc.

The AREA identifies the setting where “the THEME’s movement take place without a specified PATH”. A notable consequence of the lack of a single linear trajectory is that the AREA cannot co-occur with SOURCE, GOAL and PATH, i.e. it is defined in an ‘excludes’ relation to each of them as well as to the FEs DISTANCE and DIRECTION which provide additional details referring to the translational motion in space. Like PATH, the semantic specification of AREA refers to locations, physical objects, other objects occupying space, various media, such as water, air, etc., which, however, may be construed as comprising some spatial expanse in, over or around which the motion takes place in an irregular fashion.

⁸The frame definitions are taken verbatim from FrameNet: https://framenet.icsi.berkeley.edu/framenet_search.

⁹The definitions of the FEs are taken from the description of the Motion frame, with further elaboration on their semantic specification informed by the annotated examples studied in the paper.

Svetlozara Leseva

The **DIRECTION** indicates “motion along a line from the deitic center towards a reference point (which may be implicit) that is neither the **GOAL** of the posture change nor a landmark along the way of the moving part of the body. Often **Direction** is defined with reference to the canonical orientation of the Protagonist, or the orientation imposed by an implicit observer”. The semantic specification of this FE includes directions, such as compass points (north, east, south, west), body relative directions (left, right, back, front, backward, forward, up, down), coordinates, etc.

The **DISTANCE** encompasses expressions that characterise “the extent of the motion” covered by the **THEME**. Its semantic specification includes distance denotations expressed either in various systems of measurement or as relative distances (farther, closer), etc.

The basic configuration of the core FEs of the **Motion** frame and the interaction among them determines the overall semantic specification of the prototypical notion of motion, which is subject to various modifications and elaborations in the more specific motion frames.

The syntactic expression of the semantic configuration of the **Motion** frame will be discussed in terms of: (i) the (typical) syntactic projections of each core FE, in particular its syntactic (phrasal) category and grammatical relation; (ii) the most frequent valence patterns, i.e. the various frequent combinations in which the core FEs co-occur in the annotated FrameNet corpus. For English, both types of data are extracted from the summaries provided for each LU in FrameNet; the Bulgarian counterparts are analysed in comparison with the descriptions available for English and tested against the corpus of annotated examples created for Bulgarian (see Section 4.2 above).

5.3.2 Verbs evoking the **Motion** frame

The **Motion** frame is evoked by a couple of basic verbs of inherently directed motion (Levin 1993), such as *come* and *go*, as well as by verbs of non-directed motion. Within the second class, some predicates, such as *move* and *travel*, describe the general idea of moving through space, while others, for instance *blow*, *drift*, *float*, *circle*, *roll*, denote various types of manner of motion; part of these verbs, e.g. *meander*, *spiral*, *weave*, *wind*, *zigzag*, involve complex trajectories.

5.3.3 Syntactic realisation of the frame elements in the **Motion** frame

Table 1 below illustrates the syntactic realisation of several verbs evoking the **Motion** frame, chosen according to the following criteria: (i) having a sufficient

4 *The conceptualisation of the route: Non-directed and directed motion*

number of attestations in the FrameNet corpus, thus allowing for more reliable observations; (ii) illustrating distinct syntactic patterns with respect to the expression of the FEs denoting the different parts or features of the route of movement.

The THEME is typically projected in the subject position and the remaining core FEs are expressed primarily as prepositional (PP) or, more rarely, as adverbial phrases (AdvP). Some of the verbs also allow object NPs, especially as a realisation of PATH: [*She*]_{THEM} *CIRCLES* [*the taxi*]_{PATH}. In addition, the core FEs expressing elements or aspects of the route, may be conceptually present but left syntactically non-overt if they are known or retrievable or implied from the broader context or if the grammatical construction requires them to be left non-explicit (null instantiations), cf. (Ruppenhofer et al. 2016: 28–30).

Several preliminary observations are relevant at this point. Usually, the route is conceived as a tripartite spatial extent consisting of an initial part, SOURCE, a medial part, PATH¹⁰, and a final part, GOAL (Jackendoff 1983: 162, Talmy 1985: 57, Lakoff 1987: 275). As mentioned in Section 2, studies on the co-occurrence of directional phrases with motion verbs have shown that verbs tend to express the element of the route that is most prominent in their semantics. It has been convincingly demonstrated for many languages that GOAL-oriented verbs tend to co-occur with GOAL phrases (Rakhilina 2004: 22–24, Stefanowitsch & Rohde 2004: 255–257, Taremaa 2017: 174–178), among others, and SOURCE-oriented verbs co-occur with SOURCE phrases (Rakhilina 2004: 22–23, Stefanowitsch & Rohde 2004: 255–257, Taremaa 2017: 160–164), see also the analyses in Cristobal (2010), Kopecka (2010), among others.

In addition, it has been posited that there is a marked cross-lingual asymmetry with respect to the expression of the SOURCE and the GOAL, known as the goal-over-source principle. This proposition suggests that GOALS are expressed more frequently, using more fine-grained linguistic devices than SOURCES (Ikegami 1987, Wälchli & Zúñiga 2006, Verkerk 2017, among others). One of the explanations for this peculiarity offered in the literature is that GOALS bear higher information value with respect to the complete conceptualisation of motion (Stefanowitsch & Rohde 2004: 249); for an extensive overview of the discussion, see Verkerk (2017). However, as noted above, the preference for one type of phrase over another depends on the semantics of the verb, specifically, whether the verb conceptualises the motion in terms of a route or not (Stefanowitsch & Rohde 2004).

¹⁰Various names have been applied to this part; here I adopt the name of the relevant FrameNet frame element.

Unlike SOURCE- and GOAL-phrases, PATH-phrases do not express directionality (see also Pantcheva 2011: 31). Considering that many of the verbs in the Motion frame describe non-directed motion, the PATH should be the most prominent phase of motion inherent in their semantics and hence will be favoured for syntactic expression, unless some other aspect of motion turns out to be more prominent. Respectively, we should expect that the inherently directed-motion verbs – *go* and *come* – favour GOAL-phrases. Table 1 confirms these expectations, which are further corroborated by the co-occurrence patterns in Table 2: the ones involving PATHs are the most frequent among the top ranking patterns (Column 1) and have the greatest number of occurrences (Column 2) across the greatest number of verbs (Column 3).

Looking at individual verbs, a couple of trends may be noted with respect to the prevalence of expression of the route-related FEs (i.e. all the core FEs, excluding the THEME) (Table 1).

PATH: Several verbs show marked preference for expressing syntactically PATHs over any other route-related FEs. These include *move* as well as a number of manner of motion verbs: *weave*, *circle*, *glide*, *meander*, *wind*, *zigzag*. Among them *glide* denotes qualitative features of the movement (effortlessness) and the contact with the surface along which the motion takes place. *Weave*, *wind*, *meander* and *zigzag*¹¹ describe complex vacillating or snake-like movement along a more or less linear route or general direction, while *circle* refers to a circular trajectory. In all these cases the PATH – including its form, landmarks, etc. – is the default spatial dimension according to which the movement is characterised.

The second most prominent aspect of motion with the verbs in the Motion frame is the end-point of the route, the GOAL. It is usually less frequent than PATH, to the exception of the inherently directed motion verb *go* (cf. also Stefanowitsch & Rohde (2004: 253–254), for which the prevalence of GOAL- over PATH-phrases is roughly 4:1.

¹¹There are other verbs, e.g. *undulate* and *spiral*, that possibly behave in a like manner, but the number of occurrences is too small to make a judgment.

4 The conceptualisation of the route: Non-directed and directed motion

Table 1: Syntactic expression of the Motion FEs in FrameNet

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>move</i>								
THEME	67							67
AREA			2					2
SOURCE			5	2				7
PATH		3	30	5	21			59
GOAL			5		1			6
DIRECTION				1				1
DISTANCE							1	1
<i>go</i>								
THEME	64				1			65
AREA		1	1					2
SOURCE			1	1	3			5
PATH			8	2	1			11
GOAL	1	2	30	8	6	1		48
DIRECTION			1	5			1	7
DISTANCE			1	1	1			3
<i>drift</i>								
THEME	39							39
AREA			2					2
SOURCE			9					9
PATH			15	4	4			23
GOAL			8	1			1	10
DISTANCE				1			1	2
<i>float</i>								
THEME	43							43
AREA			13	1				14
SOURCE			4	1				5
PATH			13	3	2			18
GOAL			8					8
DISTANCE				1				1
<i>roll</i>								
THEME	31							31
AREA			1					1
SOURCE			1	1				2
PATH		3	17	2				22
GOAL			9	1		1		11

Svetlozara Leseva

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>slide</i>								
THEME	26							26
AREA			3					3
SOURCE			1				1	2
PATH			9					9
GOAL			10					10
DIRECTION			1	5				6
<i>swerve</i>								
THEME	27							27
AREA			2		4			6
SOURCE			4					4
PATH			10					10
GOAL			2					2
DIRECTION			5	2				7
<i>weave</i>								
THEME	27							27
AREA			3					3
SOURCE			1					1
PATH		1	20		1			22
GOAL			1					1
DIRECTION			4					4

The other verbs tend to express either the PATH or another aspect of the route, usually with prevalence of the former.

PATH or GOAL: The preference of either PATH- or GOAL-phrases (but not any other type of phrase) is typical of the verb *roll*: the PATH-expressions outnumber the GOAL-expressions by two to one.

PATH, GOAL or SOURCE: *Blow* and *drift* exhibit preference for either PATH- or GOAL-phrases but also tend to express the SOURCE more often than most of the remaining verbs evoking the MOTION frame. However, with *drift* PATH is the predominant type of expression, with GOALS and SOURCES being twice as few, while with *blow* the three parts of the route are represented equally.

The three other motion FEs – DIRECTION, DISTANCE and AREA – are not usually discussed separately in the literature. By virtue of its definition, AREA occurs in competition with the elements of the route. The rationale is that it describes motion encompassing an expanse that is not construed in terms of a discreet trajectory. AREAS are not equally represented across manner of motion verbs

4 The conceptualisation of the route: Non-directed and directed motion

as a whole, but are typical for some of them. DIRECTION and DISTANCE are represented by just a few examples across different verbs and have much poorer inventories. They still do need to be considered as separate FEs, as (i) some verbs incorporate them (e.g. *descend*, *rise* incorporate DIRECTION), and (ii) they may be independently expressed syntactically (Example 1).

- (1) [*The storm*]_{THM} was MOVING [*north*]_{DIR} [*along the coast*]_{PATH}.

PATH, DIRECTION or AREA: This pattern is represented by the verb *swerve*, which describes motion involving a complex route characterised by an abrupt change in the direction from an imaginary straight line or course. Respectively, it tends to co-occur with PATH-expressions as well as with ones denoting the newly assumed DIRECTION. As this kind of motion may encompass a broader spatial region, the FE AREA is also more frequently expressed than with other verbs.

PATH, AREA or GOAL: The verb *float* denotes a manner of motion which is brought about by the movement of a fluid. As this type of motion tends to encompass an expanse of the medium where it takes place, AREA-expressions are much more typical than with the rest of the verbs evoking the frame – almost on a par with PATHS and more than GOALS.

PATH, GOAL or DIRECTION: This pattern is exemplified by the verb *slide*. While it is expected to co-occur with PATH (like *glide*), the verb also shows a tendency to express directionality either by means of GOAL-phrases, which in the data are represented on a par with PATHS, or by means of the FE DIRECTION.

5.3.4 FrameNet valence patterns

Table 2 sums up the most frequent valence patterns represented among verbs evoking the Motion frame, understood as combinations of FEs which co-occur syntactically, including null instantiations.¹²

The patterns corroborate the prominence of the PATH FE expressed predominantly as a prepositional phrase, followed by indefinite null instantiations, noun phrases and adverbial phrases. The second most frequent pattern involves the GOAL, followed by AREA- and SOURCE-phrases. It is also notable that the simultaneous expression of two route-related FEs is much rarer.

¹²Non-core FEs are not considered in the analysis.

Table 2: FrameNet valence patterns of Motion verbs

Pattern	#	Verbs
[NP.Ext] _{THM} [PP] _{PATH}	134	<i>move, meander, go, roll, snake, float, undulate, zigzag, coast, fly, slide, swerve, glide, blow, circle, weave, drift, wind</i>
[NP.Ext] _{THM} [PP] _{GOAL}	60	<i>move, fly, slide, meander, go, roll, soar, swerve, come, blow, float, drift</i>
[NP.Ext] _{THM} [] _{PATH-INI}	40	<i>coast, move, go, soar, glide, blow, float, circle, weave, drift, zigzag</i>
[NP.Ext] _{THM} [PP] _{AREA}	29	<i>move, fly, slide, go, roll, snake, swerve, blow, float, weave, drift</i>
[NP.Ext] _{THM} [PP] _{SRC}	28	<i>move, slide, snake, swerve, come, glide, blow, float, drift, zigzag</i>
[NP.Ext] _{THM} [NP.Obj] _{PATH}	14	<i>coast, move, roll, soar, circle, weave</i>
[NP.Ext] _{THM} [PP] _{DIR}	11	<i>swerve, come, weave</i>
[NP.Ext] _{THM} [AVP] _{PATH}	11	<i>move, soar, glide, blow, float, drift</i>
[NP.Ext] _{THM} [AVP] _{GOAL}	9	<i>go, roll, blow</i>

5.3.5 Syntactic realisation of Motion verbs in Bulgarian

The list of Bulgarian verbs evoking the Motion frame includes the Bulgarian counterparts of the verbs considered above. In particular, it features (i) a couple of verbs of directed motion which belong to the central part of the motion lexis – *идвам* (come), *отивам* (go) and their perfective aspect counterparts¹³ (on the deictic aspects of these verbs, cf. Nitsolova 1984, Trifonova 1982, Stanisheva 1985, among others), and (ii) a number of non-directed motion verbs, predominantly ones describing various manners of motion, such as *вия се* (wind, weave), *духам* (blow), *летя* (fly), *лъкатуша* (meander), *нося се* (drift, float), *плувам, плавам* (float), *кръжа, обикалям* (circle), *търкалям се* (roll), etc., as well as ones denoting the general idea of moving through space, such as *движи се* (move, locomote) and *пътувам* (travel).

A selection of corpus examples have been collected for several of them (verbs having correspondences among the English predicates represented in Table 1), and annotated with the core FEs: *вия се* (wind, weave), *движи се* (move), *нося*

¹³For brevity only the imperfective members of aspectual pairs will be listed in the text. The annotated examples include verbs of both aspects, where such exist.

4 The conceptualisation of the route: Non-directed and directed motion

се (drift, float), *отивам/отида* (go), *търкалям се* (roll). Although on a smaller scale, the results, shown in Table 3, are consistent with the observations on the FrameNet corpus. In particular, *отивам* shows a very strong preference for GOAL-phrases similarly to *go* (Example 2), while the rest of the verbs (Examples 3–6) favour PATHs, with different proportions of other FEs, in particular AREAS for *нося се* (float, drift) and GOALS for *търкалям се* (roll).

- (2) [Те]_{THM} **ОТИВАТ** *право* [*в печатницата*]_{GOAL}.
They go-PRS.3PL straight to printer's-DEF.
'They are going straight to the printer's.'
- (3) [Кучетата]_{THM} **СЕ ДВИЖАТ** [*по мекия сняг*]_{PATH}.
Dogs-DEF move-PRS.3PL across soft-DEF snow.
'The dogs are moving across the soft snow.'
- (4) [Топката]_{THM} **СЕ ТЪРКАЛЯ** [*по тревата*]_{PATH}.
Ball-DEF roll-PRS.3PL across grass-DEF.
'The ball is rolling across the grass.'
- (5) [Колата]_{THM} **СЕ ВИЕШЕ** [*по завоите*]_{PATH}.
Car-DEF wind-PST.3SG along turns-DEF.
'The car was winding along the turns of the road.'
- (6) [Туфа водорасли]_{THM} **СЕ НОСИ** [*във водата*]_{AREA}.
Clump-INDF seaweed float-PRS.3SG on water-DEF.
'A clump of seaweed is floating on the water.'

Table 3: Syntactic expression of the Motion FEs in Bulgarian

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>вия се</i> (wind, weave)								
THEME	26							26
AREA			5					5
SOURCE			4					4
PATH			10					10
GOAL			3					3
DIRECTION				1				1

Svetlozara Leseva

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>нося се</i> (float, drift)								
THEME	32							32
AREA			9					9
SOURCE			5	1				6
PATH			8	1				9
GOAL			6					6
DIRECTION			2	1				3
<i>движа се</i> (move)								
THEME	31							31
AREA			1					1
SOURCE			1					1
PATH			18	1				19
GOAL			2					2
DIRECTION			3					3
<i>търкалям се</i> (roll)								
THEME	30							30
SOURCE			2					2
PATH			19					19
GOAL			7					7
DIRECTION				2				2
<i>отивам/отида</i> (go)								
THEME	28							28
SOURCE			1					1
PATH			2					2
GOAL			15	4	1			20
DIRECTION				3				3
DISTANCE				2				2

5.3.6 Valence patterns in the Bulgarian dataset

The valence patterns in the Bulgarian dataset, represented in Table 4, show similar results to the ones in the FrameNet corpus: in particular, a prevalence of patterns exhibiting PP PATHS, followed by a more modest representation of GOALS and AREAS. Among the several top valence patterns, combinations of SOURCES and GOALS are also found.

4 The conceptualisation of the route: Non-directed and directed motion

Table 4: FrameNet valence patterns of Motion verbs in Bulgarian

Pattern	#	Verbs
[NP.Ext] _{THM} [PP] _{PATH}	55	<i>вия се, въртя се, движа се, нося се, търкалям се, търкулна се, отивам/отида</i>
[NP.Ext] _{THM} [PP] _{GOAL}	25	<i>вия се, движа се, нося се, търкалям се, търкулна се, отивам/отида</i>
[NP.Ext] _{THM} [PP] _{AREA}	15	<i>вия се, движа се, нося се</i>
[NP.Ext] _{THM} [] _{PATH-INI}	10	<i>вия се, движа се, нося се, отивам/отида, търкалям се, търкулна се</i>
[NP.Ext] _{THM} [PP] _{SRC}	6	<i>вия се, нося се</i>
[NP.Ext] _{THM} [AVP] _{DIR}	5	<i>нося се, търкалям се, отивам/отида</i>
[NP.Ext] _{THM} [PP] _{GOAL} [PP] _{SRC}	5	<i>движи се, нося се, търкалям се, отивам/отида</i>
[NP.Ext] _{THM} [PP] _{DIR}	4	<i>движи се, нося се</i>
[NP.Ext] _{THM} [AVP] _{GOAL}	4	<i>отивам/отида</i>

5.4 Self_motion

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

5.4.1 Semantic description of the Self_Motion frame

Frame definition: 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 and the one that primarily motivates the differentiation between Motion and Self_motion is the capability of the SELF_MOVER to change location by exercising their own will and power by the coordinated movement of their bodies¹⁴, which is not a necessity with the Motion THEME. By metaphorical extension, SELF_MOVERS may be self-directed entities such as vehicles. The remaining core FEs in this frame are the ones defining the elements

¹⁴https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Self_motion

Svetlozara Leseva

and aspects of the route of movement¹⁵.

Core FEs in the *Self_motion* frame: *SELF_MOVER*, *SOURCE*, *GOAL*, *PATH*, *AREA*, *DIRECTION*, *DISTANCE*. *SELF_MOVER* is the entity (living being or a vehicle) that changes location under its own power and direction. Its semantic specification includes **animate beings** and **vehicles**. The remaining core FEs have the same semantic specification as their counterparts in the *Motion* frame from which they are inherited.

5.4.2 Verbs evoking the *Self_motion* frame

Unlike its parent frame, *Self_motion* prototypically describes individuals capable of applying their own will and bodies to perform the motion. The verbs thus encode various aspects of motion impossible for inanimate beings. These involve modes of motion: (i) characteristic of differently organisms, e.g. *fly*, *swim*, *crawl*, *slither*, *walk*, etc.; (ii) requiring different configuration of the body: *slouch*, *shoulder*; (iii) (lack of) purposefulness: *roam*, *saunter*, *wander*; (iv) intent: *prowl*, *hike*, *hitchhike*; (v) different kinds of steps, speed, weight or force applied: *mince*, *scurry*, *shuffle*, *plod*, *trample*, *run*, *jog*, *hop*, etc.

5.4.3 Syntactic realisation of the frame elements in the *Self_motion* frame

The expression of the core FEs according to syntactic categories and syntactic function is similar to those of the corresponding FEs in the *Motion* frame. The *SELF_MOVER* is realised as the external argument; the remaining core FEs are typically realised as prepositional or adverbial phrases.

Table 5 illustrates the syntactic expression of the core FEs for several English verbs with the highest number of attestations in the FrameNet corpus. The verbs evoking these semantic frames further extend the observations made for the *Motion* frame with respect to the tendency for the various verbs to co-occur with motion expressions. Overall, the *PATH* is the prevalent FE to be expressed, followed by *GOALS*, *AREAS* and *SOURCES* in descending order (see the valence patterns in Table 6).

PATH: Several verbs exhibit a strong preference for *PATHS* over any other core FE: *amble*, *drive*, *make*, *plod*.

PATH OR GOAL: Verbs that show preference to co-occur with either of these FEs can be further distinguished into two patterns.

¹⁵*DISTANCE* is not defined as a core FE, but will be treated on a par with its equivalent in the mother frame.

4 The conceptualisation of the route: Non-directed and directed motion

The first one is PATH > GOAL: In this case, the examples with PATH show prevalence, amounting to around or even more than half of the examples, and GOALS usually account for a quarter to a third, rarely more, see *hop* in Table 5. This pattern is further illustrated by *hurry, jog, limp, lumber, lunge, lurch, proceed, skip, stagger, stroll, stumble, swagger, totter, trot, trumble, trek*.

PATH = GOAL: With the second pattern, there is no marked preference for one FE over the other, as exemplified by *walk* (Table 5). Other verbs which pattern in a similar way are: *barge, clamber, dash, head, hasten, pad, romp, sidle, toddle, wade*.

PATH, GOAL or SOURCE: This pattern is distinguished from the second subgroup of the previous one empirically on the basis of the greater ratio of SOURCES against the overall number of examples for each of the verbs. The verbs in this group tend to co-occur with expressions denoting any of the three parts of the route more consistently than the remaining verbs evoking the Self_motion frame. As already shown in the Motion frame, the frequency of each of these FEs is not equal across verbs. In this group one finds that PATHS account for half to up to two-thirds of the examples, GOALS – for a quarter to a third of the examples, SOURCES – usually for a fifth to a quarter of the examples, as illustrated by *crawl* (Table 5), *creep, dart, march, saunter, scamper, scramble, shuffle, spring, sprint, stride, trudge*. Another variation is represented by the verbs *lope, leap, jump*, where PATHS account for half or more of the examples, and GOALS and SOURCES are on a par, about one third of the instances.

PATHS = GOALS or SOURCE: This pattern shows no marked difference between PATHS and GOALS with a weaker preference for SOURCES: *climb* (Table 5), *rush, scuttle*.

GOAL: A couple of verbs, such as *file* and *pounce* show marked preference for GOAL-expressions over all other motion-related FEs.

GOAL or PATH: These verbs tend to co-occur with both GOALS and PATHS with a prevalence of the former (about a half of the examples) to the latter (around a third of the examples): *steal, run* (Table 5).

GOAL, PATH or SOURCE: This combination is exemplified by verbs such as *troop, sneak* (Table 5), *stalk*. The GOALS amount to half or more of the instances, while the PATHS and SOURCES are fewer – around one-third and a quarter, respectively for *troop* and equally distributed for *sneak* and *stalk*.

Svetlozara Leseva

The FE AREA usually alternates with expressions denoting one or another element or aspect of the route of a moving entity and as a whole accounts for much fewer cases than PATHS and GOALS in the frame. For some verbs, however, it is either the preferred motion expression or is much frequent than with most verbs. This characteristic is typical of verbs that describe motion that encompasses or spreads over a larger region or expanse.

AREA: The verbs *traipse* and *skulk* show a much more marked preference for AREAS than for other motion-related FEs: half of the instances for *traipse*, two-thirds for *skulk*.

AREA or PATH: Other verbs tend to co-occur with either AREAS or PATHS with a prevalence of the former (half or more of the examples) to the latter (around one-third of the examples): *prance*, *prowl*, *roam* (Table 5).

PATH or AREA: The opposite is observed with *strut* and *flit* where PATHS are preferred (between half and two-thirds of the examples) to AREAS (a quarter of the examples).

PATH, AREA or GOAL: This pattern shows prevalence of PATHS (with half or more of the instances), a substantial number (a quarter to one-third) of AREAS and a smaller number (one-sixth to one-fifth of the examples) of GOALS: *dance*, *pace*, *swim* (Table 5), *tread*, *tramp*. In the case of *fly* the number of AREAS and GOALS is equal.

PATH, GOAL, AREA or SOURCE: This pattern shows prevalence of PATHS (around half of the examples), with various distributions (between one-fifth and one-third) of the other three FEs: *scurry*, *slither*, *waddle*, *wander* (Table 5).

A couple of verbs, such as *flounce* and *storm*, show preference to SOURCES over other motion-related FEs.

4 The conceptualisation of the route: Non-directed and directed motion

Table 5: Syntactic expression of the Self_motion FEs in FrameNet

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>climb</i>								
SELF_MOVER	115							115
AREA		1	2		2			5
SOURCE			21					21
PATH		1	46	3	4			54
GOAL			59	1				60
<i>crawl</i>								
SELF_MOVER	140							140
AREA			18		4			22
SOURCE			23	6				29
PATH			58	7	9		1	75
GOAL			31	4		1		36
<i>hop</i>								
SELF_MOVER	103							103
AREA			14		1			15
SOURCE			9	2				11
PATH			50	8	2		2	62
GOAL			28	4				32
<i>hurry</i>								
SELF_MOVER	74							74
SOURCE			10	2				12
PATH			41	7	2			50
GOAL			28	8				36
<i>roam</i>								
SELF_MOVER	66							66
AREA		13	26	1	5			45
SOURCE			1					1
PATH			13	2	3			18
GOAL			2	2				4
<i>run</i>								
SELF_MOVER	64							64
AREA			1	3	2			6
SOURCE			3	1	2			6
PATH			16		3			19
GOAL			16	3	9			28
DIRECTION			4	2	2			8

Svetlozara Leseva

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>sneak</i>								
SELF_MOVER	68							68
AREA			2					2
SOURCE			17	4				21
PATH			20					20
GOAL		1	37	6				44
<i>swim</i>								
SELF_MOVER	259				1			260
AREA	1	2	54	3	1			61
SOURCE			19	14				33
PATH		5	95	25	42			167
GOAL		1	45	3	1	1		51
DIRECTION				1				1
<i>walk</i>								
SELF_MOVER	102							102
AREA			9	2	4		1	16
SOURCE			17		1			18
PATH		2	36		6		1	45
GOAL			29	2	1	3	7	42
DIRECTION			3	3	1			7
<i>wander</i>								
SELF_MOVER	81							81
AREA			17		3			20
SOURCE			12	5				17
PATH			33	4	4			41
GOAL			27	2				29

5.4.4 FrameNet valence patterns

The valence patterns exhibited in the Motion frame are confirmed on a larger scale by Self_motion, in particular the prevalence of PATH-expressions over GOALS, AREAS and SOURCES in descending order. It is worth noting that the number of the second most frequent pattern as compared with the most frequent one is higher than for Motion verbs (66% and 45%, respectively) i.e. GOAL expressions are found more frequently as compared with PATH expressions with Self_motion verbs. In addition, the most frequent patterns involving two motion-related FEs are GOAL + PATH and GOAL + SOURCE representing about 19% and 11% of the number of the most frequent pattern; this ratio is much greater than for Motion, where the pattern GOAL + PATH amounts to 6% of the most frequent one.

4 The conceptualisation of the route: Non-directed and directed motion

An interesting hypothesis to test on this amount of data would be whether this observation ties with animacy and/or agentivity.

5.4.5 Verbs evoking the Self_motion frame in Bulgarian

Many of the Bulgarian verbs that evoke the Self_motion frame are manner of motion simplex verbs. This aligns with the fact that, overall, SELF_MOTION describes the idea of movement without profiling any of the route-related aspects of motion. Other verbs, such as *втурвам се* (rush), *отправлям се* (head, make), *спускам се* (dash, dart), *налитам*, *хвърлям се* (barge), etc., involve not manner but rather directed motion or the initial phase of motion. Such verbs usually come in aspectual pairs.

5.4.6 Syntactic realisation of Self_motion verbs in Bulgarian

Table 7 shows the results for several frequent Bulgarian verbs with correspondence in Table 5. For the English data GOALS are found in competition with PATHS and other motion-related FEs, either in fewer numbers, but still well-represented across many verbs, or in greater numbers than the other FEs co-occurring with the respective predicates. The most notable difference found in the Bulgarian sample is the lower frequency of GOALS as compared with the data in the Frame-Net corpus.

The verbs *бродя* (roam, wander), *вървя* (walk), *катеря се* (climb), *плувам* (swim), *пълзя* (crawl), *тичам* (run) (Examples 7–12) all show a lower occurrence of GOALS, whereas in English *climb* and *walk* co-occur equally with both FEs and *run* shows preference for GOALS over PATHS.

- (7) [Ноя]_{SMOV} **ТИЧАШЕ** [към него]_{GOAL}.
Noah run-PST.3SG towards him.
'Noah was running towards him.'
- (8) [Той]_{SMOV} **ТИЧАШЕ** [по дългия коридор]_{PATH}.
He run-PST.3SG down long-DEF hall.
'He was running down the long hall.'
- (9) [Той]_{SMOV} **СЕ КАТЕРИ** [по хълма]_{PATH}.
He climb-PRS.3SG up hill-DEF.
'He is climbing up the hill.'
- (10) [Те]_{SMOV} **БРОДЕХА** [по коридорите]_{PATH}.
They wander-PST.3PL along corridors-DEF.
'They wandered along the corridors.'

Table 6: FrameNet valence patterns of Self_Motion verbs

Pattern	#	Verbs
[NP.Ext] _{SMov} [PP] _{PATH}	1576	<i>stumble, mince, lurch, frolic, stride, climb, tramp, scurry, trip, stalk, rip, burrow, strut, roam, dance, prow, jump</i>
[NP.Ext] _{SMov} [PP] _{GOAL}	1035	<i>stumble, mince, lurch, stride, climb, tramp, scurry, trip, stalk, rip, burrow, strut, roam, press</i>
[NP.Ext] _{SMov} [PP] _{AREA}	599	<i>stumble, hobble, mince, lurch, lope, frolic, stroll, bustle, stride, swagger, crawl, scramble, climb, spring, tramp</i>
[NP.Ext] _{SMov} [PP] _{SRC}	415	<i>stumble, hobble, mince, lurch, sprint, lope, run, lunge, stroll, bustle, stride, slosh, swagger, crawl, scramble</i>
[NP.Ext] _{SMov} [] _{PATH-INI}	375	<i>stumble, hobble, lurch, sprint, run, lunge, stroll, stride, swagger, crawl, scramble, climb, spring, tramp, pad</i>
[NP.Ext] _{SMov} [PP] _{GOAL} [PP] _{PATH}	297	<i>stumble, hobble, mince, lurch, sprint, lope, stroll, stride, swagger, crawl, scramble, climb, spring, tramp, pad</i>
[NP.Ext] _{SMov} [AVP] _{PATH}	187	<i>stumble, lurch, frolic, lunge, stroll, bustle, stride, swagger, crawl, climb, spring, tread, trek, scurry, trip, trundle, strut</i>
[NP.Ext] _{SMov} [PP] _{GOAL} [PP] _{SRC}	175	<i>stumble, hobble, lurch, sprint, lope, lunge, stroll, bustle, stride, crawl, scramble, climb, spring, pad, trek, edge</i>
[NP.Ext] _{SMov} [AVP] _{GOAL}	149	<i>stumble, hobble, lurch, sprint, lope, run, lunge, stroll, bustle, crawl, scramble, climb, spring, tramp, pad, edge</i>
[NP.Ext] _{SMov} [AVP] _{SRC}	113	<i>stumble, hobble, sprint, lope, stroll, bustle, stride, swagger, crawl, scramble, tramp, pad, scurry, romp, stalk, sneak</i>
[NP.Ext] _{SMov} [PP] _{PATH} [PP] _{PATH}	107	<i>stumble, hobble, lurch, sprint, lope, lunge, stroll, bustle, stride, crawl, scramble, climb, spring, tramp, pad, trek</i>
[NP.Ext] _{SMov} [PP] _{PATH} [PP] _{SRC}	96	<i>jog, prance, clamber, trudge, lurch, scamper, sprint, hop, lope, troop, stagger, stroll, bustle, stride, swagger, crawl, climb</i>

4 The conceptualisation of the route: Non-directed and directed motion

- (11) [*Назгулите*]_{SMOV} **БРОДЯТ** [*по земята*]_{AREA}.
 Nazgul-DEF roam-PRS.3PL across earth-DEF.
 ‘The Nazgul roam the earth.’
- (12) [*Той*]_{SMOV} **ПЛУВАШЕ** [*из бурното море*]_{AREA}.
 He swim-PST.3SG across stormy-DEF sea.
 ‘He was swimming in the stormy seas.’

This is at least partly predictable: while English manner of motion verbs express directionality by means of GOAL- or SOURCE-phrases or particles with a similar meaning, the corresponding Slavic (Bulgarian) simplex verbs may also derive new verbs with a directional meaning through prefixation (Beavers et al. 2010, Lindsey 2011, Pantcheva 2007a,b, 2011, Speed 2015), among many others. While simplex verbs can realise directionality by means of route-related phrases, the derived prefixed verbs profile the relevant aspect of the route and encode it in their lexical structure; the two types of verbs may be used interchangeably in certain contexts, but not in others. As a cursory illustration of this point, consider the verb in (Example 13a), whose directional meaning cannot be expressed by the simplex verb it is derived from; hence the expression *в стаята* in (Example 13b) cannot be interpreted as the GOAL (marked by an asterisk); still, it will have an English correspondence of manner of motion verb + a directional phrase.

- (13) a. [*Птицата*]_{SMOV} **ВЛИТА** [*в стаята*]_{GOAL}.
 Bird-DEF fly-PRS.3SG into room-DEF.
 ‘The bird flies into the room.’
- b. * [*Птицата*]_{SMOV} **ЛЕТИ** [*в стаята*]_{GOAL}.
 Bird-DEF fly-PRS.3SG into room-DEF.
 ‘The bird flies into the room.’

Table 7: Syntactic expression of the Self_motion FEs in Bulgarian

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>пълзя</i> (crawl)								
SELF_MOVER	40							40
AREA			4	2				6
PATH			19	1				20
GOAL			8	1				9
DIRECTION				2				2

Table 7: Syntactic expression of the Self_motion FEs in Bulgarian

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>катеря се</i> (climb)								
SELF_MOVER	48							48
AREA			2					2
SOURCE			1					1
PATH			22	1				23
GOAL			14	2				16
<i>бродя</i> (roam, wander)								
SELF_MOVER	39							39
AREA			18	2				20
SOURCE			1					1
PATH			13					13
GOAL			1					1
<i>плувам</i> (swim)								
SELF_MOVER	37							37
AREA			9	3				12
PATH			6	1				7
GOAL			5					5
<i>тичам</i> (run)								
SELF_MOVER	42							42
AREA			3					3
SOURCE			1					1
PATH			23	1				24
GOAL			11	1				12
<i>вървя</i> (walk)								
SELF_MOVER	40							40
AREA				1				1
PATH			16	2				18
GOAL			2	3				5
DIRECTION			1	1				2

The above observations are confirmed by the distribution of the patterns involving PATHS, GOALS and AREAS in the Bulgarian dataset (Table 8).

5.5 Traversing, Arriving, Departing

These semantic frames narrow down the idea of motion through profiling aspects of the general motion schema corresponding to elements of the route along which the moving object changes location: the initial stage of the motion corresponding to the SOURCE; the end-stage – associated with the GOAL, or the middle stage – corresponding to the PATH, cf. (Johnson et al. 2001: 16). Borrowed

4 The conceptualisation of the route: Non-directed and directed motion

Table 8: Valence patterns of Self_motion verbs in Bulgarian

Pattern	#	Verbs
[NP.Ext] _{SMOV} [PP] _{PATH}	86	<i>вървя, плувам, пълзя, броя, тичам, катеря се</i>
[NP.Ext] _{SMOV} [] _{PATH-INI}	45	<i>вървя, плувам, пълзя, броя, тичам, катеря се</i>
[NP.Ext] _{SMOV} [PP] _{AREA}	37	<i>плувам, пълзя, броя, тичам, катеря се</i>
[NP.Ext] _{SMOV} [PP] _{GOAL}	35	<i>вървя, плувам, пълзя, тичам, катеря се</i>
[NP.Ext] _{SMOV} [AVP] _{AREA}	6	<i>плувам, пълзя, броя</i>
[NP.Ext] _{SMOV} [AVP] _{GOAL}	6	<i>вървя, пълзя, катеря се</i>
[NP.Ext] _{SMOV} [AVP] _{PATH}	5	<i>вървя, плувам, пълзя, катеря се</i>

from Langacker (1987), profiling is understood as “the representation of the foregrounded part of a frame, the participant, prop, phase or moment which figures centrally in the semantic interpretation of the sentence within which the frame is evoked” (Fillmore et al. 2001: 16).

GOAL-profiling LUs (e.g., *arrive*, *reach*) evoke the semantic frame Arriving; SOURCE-profiling LUs (e.g. *leave*, *depart*) evoke the Departing frame; PATH-profiling LUs, such as *traverse*, *cross* correspond to the Traversing frame. Arriving and Departing are defined as subframes of Traversing: as such, each of them describes a state or transition in the conceptualisation of a complex situation referring to the sequence of transitions from the SOURCE, through the PATH, to the GOAL. Departing and Arriving are related to each other by means of the Precedes relation.

The profiling of a given FE is associated with the fact that the respective FE is central to the meaning and is always conceptually implied even if not necessarily overtly realised. In such cases it is often retrievable from the context and is thus understood and annotated as a definite null instantiation (DNI).

5.5.1 Arriving

5.5.2 Semantic description of the Arriving frame

Arriving describes directed motion towards an end point which is part of the lexical encoding of the relevant LUs: i.e. the verbs evoking the frame are GOAL-oriented verbs of inherently directed motion.

Svetlozara Leseva

Definition of the frame Arriving: An object, THEME, moves in the direction of a GOAL. The GOAL may be expressed or it may be understood from the context, but it is always implied by the verb itself.

Core frame elements: THEME, GOAL. The core FEs of the Arriving frame represent a subset of the core FEs of the Traversing frame of which it is a subframe. The FEs share the definition and semantic properties of their correspondences in the Motion frame. The profiling of the GOAL results in the backgrounding or exclusion of the remaining elements that form part of the core FEs of Traversing. SOURCE and PATH become peripheral FEs, while PATH_SHAPE, DISTANCE, DIRECTION, AREA, as defined in FrameNet, are not conceptually present in the scenario described by this frame.

5.5.3 Verbs evoking the Arriving frame

The verbs evoke the Arriving frame form a central part of the lexis of GOAL-directed motion: *appear, approach, arrive, come, crest, descend (on), enter, get, hit, make it, make, reach, return, visit*.

5.5.4 Syntactic realisation of the frame elements in the Arriving frame

The syntactic realisation of the frame elements in the Arriving frame as represented in the FrameNet corpus examples are illustrated in Table 9. The THEME is projected as a subject, while depending on the verb the GOAL may be expressed as either a prepositional or adverbial phrase – e.g. *arrive, come, return, get, make it*, or as a direct object (NP.Obj) – e.g. *approach, enter, reach, visit*.

Table 9: Syntactic expression of the Arriving FEs in FrameNet

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>approach</i>								
THEME	36		1					37
GOAL	1	29		1	6			37
<i>arrive</i>								
THEME	81							81
GOAL	1		31	11	35			78
<i>come</i>								
THEME	119							119
GOAL		2	44	16	50	3	2	117

4 The conceptualisation of the route: Non-directed and directed motion

Table 9: Syntactic expression of the Arriving FEs in FrameNet

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>enter</i>								
THEME	30	1			3			34
GOAL	3	17			10		1	31
<i>return</i>								
THEME	48				3			48
GOAL		1	17	9	21	1		49
<i>visit</i>								
THEME	24		3		2			29
GOAL	5	14	2		7	1	1	29
<i>reach</i>								
THEME	50	1			7			58
GOAL	7	48		1				56
<i>get</i>								
THEME	35							35
GOAL			16	12	7		1	36
<i>make it</i>								
THEME	22							22
GOAL			12	1	9			22

The possibility for leaving the GOAL non-overt as reflected in the considerable number of definite null instantiations (NIs in the table), stems from the fact that with some verbs this FE often receives a definite interpretation as the deictic centre and its identity is thus implied even without previous reference. This is typical for *come* and to a lesser degree for *arrive* due to their deictic nature. In this respect they are clearly distinct from *reach*, *approach*, *visit*, *get* and *make it*, which usually need an overtly expressed GOAL, as it need not be identical to the deictic centre. Examples (14a, 14b) illustrate this point.

- (14) a. [She]_{THM} **REACHED** [Rome]_{GOAL} [via Assisi]_{PATH}.
b. * [She]_{THM} **REACHED**.

5.5.5 FrameNet valence patterns

In line with the above observations, syntactically implicit GOALS represent half of the aggregated number of the GOAL-phrases (Table 10). There is a considerable number of NP GOALS, which accounts for the fact that a great deal of the verbs

Svetlozara Leseva

are transitive. In addition, AVPs are much more prominent: they make up for a third of the prepositional GOAL-phrases, while in Self_motion their number is 15% of the number of GOAL-PPs.

Table 10: FrameNet valence patterns of Arriving verbs

Pattern	#	Verbs
[NP] _{THM} [] _{GOAL-DNI}	144	<i>appear, approach, arrive, come, enter, return, visit, get, make it</i>
[NP] _{THM} [NP] _{GOAL}	126	<i>approach, enter, visit, reach, make, crest, hit</i>
[NP] _{THM} [PP] _{GOAL}	121	<i>arrive, come, return, visit, get, make it, descend (on), appear</i>
[NP] _{THM} [AdvP] _{GOAL}	46	<i>approach, arrive, come, return, reach, get, make it</i>

5.5.6 Syntactic realisation of Arriving verbs in Bulgarian

The basic verbs evoking the Arriving frame form a small but central part of the lexis of directed motion: *влизам* (enter), *връщам се* (return), *доближавам*, *доближавам се* (approach), *добирам се* (make it), *достигам* (reach), *завръщам се* (return), *наближавам*, *приближавам*, *приближавам се* (approach), *идвам*, *ида* (come), *пристигам* (arrive), *стигам* (reach), *посещавам* (visit), *прибирам се* (go home). To the exception of *посещавам*, which takes object NP GOALS and *доближавам*, *наближавам*, *приближавам*, *достигам*, *стигам* – which take either an object NP or a PP/AVP, the rest of the verbs require a PP/AVP complement. In this respect the Bulgarian verbs differ from their English counterparts, many of which take an object GOAL complement.

Table 11: Syntactic expression of the Arriving FEs in Bulgarian

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>отивам/отида</i>								
THEME	10							10
GOAL			6	1	3			10
<i>достигам/достигна</i>								
THEME	4							4
GOAL		2	2					4

4 The conceptualisation of the route: Non-directed and directed motion

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>идвам/дойда</i>								
THEME	20							20
GOAL			11		9			20
<i>пристигам/пристигна</i>								
THEME	21							21
GOAL			10		11			21
<i>стигам/стигна</i>								
THEME	15							15
GOAL		2	11	2				15
<i>влизам/вляза</i>								
THEME	23							23
GOAL			13		10			23
<i>връщам се/върна се</i>								
THEME	14							14
GOAL			7	4	3			14

As expected, the GOAL-PPs (Example 15a) predominate over NPs (Example 15b) and AVPs (Example 15c) as shown in Table 12. The possibility of leaving the GOAL syntactically unexpressed if it is construable from the context (Example 15d) is underrepresented in the sample of annotated examples.

- (15) a. [Вражески кораби]_{THM} **ИДВАТ** [към вас]_{GOAL}.
Hostile aircraft come-PRS.3PL towards you.
'Hostile spacecraft are coming your way.'
- b. Следобед [me]_{THM} **ДОСТИГНАХА** [брега]_{GOAL}.
In-afternoon-DEF they reach-PST.3PL coast-DEF.
'They reached the coast in the afternoon.'
- c. [Никой]_{THM} не **СЕ ВРЪЩА** [тук]_{GOAL}.
Nobody not REFL return-PRS.3SG-NEG here.
'No one returns here.'
- d. []_{THM} **ПРИСТИГАТЕ** []_{GOAL} *точно навреме, докторе!*
Arrive-PRS.2SG just on time, doctor!
'You arrive just on time, doctor!'

Table 12: FrameNet valence patterns of Arriving verbs in Bulgarian

Pattern	#	Verbs
[NP.Ext] _{THM} [PP] _{GOAL}	78	<i>влизам/вляза, връщам се/върна се, добирам се/добера се, доближавам (се)/доближа (се), идвам/дойда, достигам/достигна, завръщам се, завърна се, ида, навлизам/навляза, отивам/отида, прибирам се/прибера се, приближавам (се)/приближа (се), пристигам/пристигна, стигам/стигна</i>
[NP.Ext] _{THM} [] _{GOAL}	36	<i>влизам/вляза, връщам се/върна се, идвам/дойда, отивам/отида</i>
[NP.Ext] _{THM} [AVP] _{GOAL}	9	<i>връщам се/върна се, ида, отивам/отида, прибирам се/прибера се, стигам/стигна</i>
[NP.Ext] _{THM} [NP] _{GOAL}	7	<i>доближавам/доближа, достигам/достигна, наближавам/наближа, посещавам/посетя, стигам/стигна</i>

5.5.7 Departing

5.5.8 Semantic description of the Departing frame

Departing describes directed motion away from a starting point, which is encoded in the lexical meaning of the respective LUs.

Definition of the frame Departing: An object (the **THEME**) moves away from a **SOURCE**. The **SOURCE** may be expressed or it may be understood from context, but its existence is always implied by the departing word itself.

Core frame elements: **THEME**, **SOURCE**

Being a subframe of Traversing that describes the other end point of translational motion, the description of the Departing frame mirrors that of Arriving, but the profiled FE is the **SOURCE**. The profiling results in the backgrounding of the **GOAL** and the **PATH** to peripheral FEs and the removal of the remaining route FEs present in the description of Traversing (**PATH_SHAPE**, **DISTANCE**, **DIRECTION**, **AREA**) from the scenario described by Departing.

4 The conceptualisation of the route: Non-directed and directed motion

5.5.9 Verbs evoking the Departing frame

The basic verbs that evoke the Departing frame form a central part of the lexis of SOURCE-oriented directed motion: *decamp*, *depart*, *disappear*, *emerge*, *escape*, *exit*, *leave*, *skedaddle*, *vamoose*, *vanish*.

5.5.10 Syntactic realisation of the frame elements in the Departing frame

The syntactic realisation of the frame elements in the Departing FEs as represented in the FrameNet corpus examples are illustrated in Table 13. The THEME is projected as a subject (NP.Ext), while depending on the verb the SOURCE may be expressed as either a prepositional or an adverbial phrase, e.g. *disappear*, *emerge*, *vanish*, on the one hand, or as a direct object (NP.Obj), on the other: *depart*, *escape*, *exit*, *leave*.

Unlike Arriving verbs, where the verbs have a distinct preference to either NP or PP/AVP GOALS, the FrameNet data for Departing point to different distribution of NP and PP SOURCES across the verbs (Table 13), compare *depart*, where the two types of phrases are equally distributed and *leave*, which favours NP.Obj.

The Departing verbs show a similar tendency to leave the profiled element unexpressed (less prominent for the verb *leave*) if it is retrievable from the wider context and/or the movement away takes place with reference the speaker (i.e. the deictic centre).

While some Arriving verbs, such as *arrive*, *come*, *get* and *return* tend to express the GOAL as either a PP or an AVP, the Departing verbs hardly opt for AVPs, at least in the FrameNet corpus.

5.5.11 FrameNet valence patterns

While the GOAL-DNIs of the Arriving verbs represent 33% of the overall number of GOALS, syntactically implicit SOURCES are the prevalent pattern, making up for 63% of the aggregated number of the patterns with SOURCE-phrases (Table 14). In other words, judging from these data, GOAL-profiling verbs express syntactically the profiled element twice as frequently as do SOURCE-profiling verbs. This observation supports the goal-over-source asymmetry.

The number of the patterns with NP and PP SOURCES is similar, while, as noted above, AVPs, are poorly represented (Table 14).

5.5.12 Syntactic realisation of Departing verbs in Bulgarian

The Bulgarian verbs evoking the Departing frame represent the central lexis of SOURCE-oriented directed motion verbs: *заминавам* (depart), *избягвам* (escape),

Table 13: Syntactic expression of the Departing FEs in FrameNet

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>depart</i>								
THEME	77			1				78
SOURCE		14	11		52			77
<i>disappear</i>								
THEME	120							120
SOURCE			8		111			119
<i>escape</i>								
THEME	16							16
SOURCE		4	2	1	9			16
<i>vanish</i>								
THEME	69		1					70
SOURCE			12		57			69
<i>exit</i>								
THEME	32							32
SOURCE		5	5		21	1		32
<i>leave</i>								
THEME	90	1						91
SOURCE		45	7	4	29		3	88

Table 14: FrameNet valence patterns of Departing verbs

Pattern	#	Verbs
[NP.Ext] _{THM} [] _{SRC-DNI}	312	<i>decamp, exit, leave, emerge, disappear, depart, escape, vanish</i>
[NP.Ext] _{THM} [NP.Obj] _{SRC}	68	<i>exit, leave, depart, escape</i>
[NP.Ext] _{THM} [PP] _{SRC}	58	<i>decamp, exit, leave, emerge, disappear, depart, escape, vanish</i>
[NP.Ext] _{THM} [AVP] _{SRC}	5	<i>leave, escape</i>

4 The conceptualisation of the route: Non-directed and directed motion

излизам (exit), *изчезвам* (disappear), *напускам* (leave), *отивам си* (leave, go home), *тръгвам* (leave, depart), *отдалечавам се* (move away), etc. Most of the Bulgarian Departing verbs take a PP or a AVP complement, with few exceptions, *напускам* (leave) takes an NP.Obj complement.

Table 15: Syntactic expression of the Departing FEs in Bulgarian

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>напускам/напусна</i>								
THEME	39							39
SOURCE		38			1			39
<i>тръгвам/тръгна</i>								
THEME	36							36
SOURCE					36			36
GOAL			10	1				11
DIRECTION			1	1				2
<i>заминавам/замина</i>								
THEME	40							40
SOURCE			1	1	38			40
PATH			1					1
GOAL			20					20
DISTANCE			1					1
<i>излизам/изляза</i>								
THEME	39							39
SOURCE			15	1	23			39
GOAL			6	3		2		11

The data in Table 16 support the observations that apart from NP SOURCES (Example 16c), the profiled element of the Departing frame (Example 16a) tends to be left out, i.e. it is usually interpreted from the previous or the general context (Example 16b).

In addition, while the peripheral frame element SOURCE in the Arriving frame is rarely expressed (in fact not present in the data), the peripheral frame element GOAL in the Departing frame (Example 16d) was found to be quite frequently expressed and was thus annotated in the Bulgarian examples: in fact, it has as many occurrences as the profiled FE SOURCE (Table 16).

Svetlozara Leseva

Table 16: FrameNet valence patterns of Departing verbs in Bulgarian

Pattern	#	Verbs
[NP.Ext] _{THM} [] _{SRC-DNI}	53	заминавам/замина, излизам/изляза, напускам/напусна, тръгвам/тръгна
[NP.Ext] _{THM} [NP] _{SRC}	38	напускам/напусна
[NP.Ext] _{THM} [PP] _{GOAL} [] _{SRC-DNI}	35	заминавам/замина, излизам/изляза, тръгвам/тръгна
[NP.Ext] _{THM} [PP] _{SRC}	15	заминавам/замина, излизам/изляза
[NP.Ext] _{THM} [AVP] _{GOAL} [] _{SRC-DNI}	4	излизам/изляза, тръгвам/тръгна

- (16) a. []_{THM} *He* **ИЗЛИЗАЙ** [от къщи]_{SRC}.
Not go-out-IMP.2SG-NEG out-of house-DEF.
'Don't leave the house.'
- b. [Той]_{THM} **ЗАМИНА** []_{SRC} на сутринта.
He leave-PST.3SG on morning-DEF.
'He departed on the following morning.'
- c. [Тя]_{THM} **НАПУСНА** [града]_{SRC} завинаги.
She leave-PST.3SG city-DEF for good.
'She left the city for good.'
- d. След завършването [той]_{THM} **ЗАМИНА** []_{SRC} [за
After graduating, he leave-PST.3SG for
Париж]_{GOAL}.
Paris.
'After his graduation he left for Paris.'

Another fact that emerged from the data is that, even though DIRECTION and DISTANCE are not specified in the Arriving and the Departing frame, there are examples that suggest that these FEs are part of the description of the two semantic frames, even if with a peripheral status (Example 17 and Example 18, respectively).

4 The conceptualisation of the route: Non-directed and directed motion

- (17) a. [Корабѣт]_{THM} **ЗАМИНАВА** []_{SRC} [на юз]_{DIR}.
 Ship-DEF leave-PRS.3SG to south.
 ‘The ship leaves south.’
 b. [Тя]_{THM} **ЗАМИНА** [на 3000 км]_{DIST} [от дома]_{SRC}.
 She leave-PST.3SG to 3000 km from home.
 ‘She went (to live) 3,000 km away from home.’
- (18) a. [Те]_{THM} **ПРИСТИГАТ** []_{GOAL} [от юз]_{DIR}.
 They arrive-PRS.3PL from south.
 ‘They arrive from the south.’
 b. []_{THM} **ИДВАХА** [тук]_{GOAL} [отдалече]_{DIST}.
 They come-PST.3PL here from far away.
 ‘They came here from far away.’

5.5.13 Traversing

5.5.14 Semantic description of the Traversing frame

Traversing represents the complex situation of the motion of a **THEME** with respect to the different locations constituting the route.

Definition of the frame Traversing: A **THEME** changes location with respect to a salient place, which can be expressed by a **SOURCE**, **PATH**, **GOAL**, **AREA**, **DIRECTION**, **PATH_SHAPE**, or **DISTANCE**.

The frame profiles the middle section of the trajectory of motion of a moving entity, i.e. the **PATH**. Its core FEs include the **PATH** itself, as well as elements that represent either an alternative expression of the idea of space covered by the moving entity (such as **AREA**) or a characteristic feature of the **PATH**. These features may include: **DIRECTION**, which adds the dimension of spatial orientation to the non-directional **PATH**; **DISTANCE**, i.e. the length or extent of the trajectory between the starting and the ending point; **PATH_shape** – the form of the **PATH**. All of the core FEs that describe the Traversing frame are inherited from the most abstract motion frame **Motion_scenario** which is perspectivised by Traversing.

5.5.15 Verbs evoking the Traversing frame

As with Arriving and Departing, there are just a small number of non-derived verbs that evoke the frame: *ascend, circle, crisscross, cross, descend, hop, jump, leap, mount, pass, skirt, traverse*.

Svetlozara Leseva

5.5.16 Syntactic realisation of the frame elements in the Traversing frame

Table 17 illustrates the syntactic expression for a selection of Traversing verbs. The THEME is projected as the subject. Among the motion-related FEs, usually it is the profiled PATH that is expressed syntactically; its favoured realisation is either as a direct object NP, e.g. *ascend*, *cross*, *descend*, *skirt*, or as a prepositional (or adverbial) phrase, e.g. *pass* and *leap*. It can also be left unexpressed (DNI), although the number of unexpressed PATHS is much fewer than that of the profiled FEs of the Arriving and the Departing frame.

When the AREA is expressed, it may also take the place of the direct object: for most of the verbs these are single occurrences, except for *circle* and *crisscross*, whose semantics is consistent with motion along an irregular trajectory over an extended region, which predetermines their preference for the AREA over the PATH.

SOURCES and GOALS are expressed as prepositional or adverbial phrases, DIRECTION, DISTANCE, sometimes AREA (when not an object), although represented by just a few examples, are realised likewise. A small number of exceptions is found with *descend*, where some DISTANCES and DIRECTIONS are annotated as NP objects (e.g. *descended 300 m*).

PATH_SHAPES are almost always implied in the semantics of the verbs but are rarely expressed (as PPs/AVPs).

5.5.17 FrameNet valence patterns

The most frequent valence patterns (Table 18) show in even more explicit terms that across the different verbs evoking the frame, the non-overt realisation of the PATH is much rarer, especially when compared with the profiled elements of Traversing's subframes, while NPs and PPs are both well-represented, with variations across the different verbs. Another fact worth noting is that out of the remaining FEs, the GOAL is the preferred one to be expressed.

5.5.18 Syntactic realisation of Traversing verbs in Bulgarian

The central part of the Bulgarian verbs evoking the Traversing frame includes predicates such as *минавам* (pass), *кръстосвам* (crisscross), *спускам се*, *слизам* (descend), *качвам се*, including several verbs produced by means of derivation, although not necessarily transparent in the contemporary language: *изкачвам се* (ascend); *прекосявам*, *пресичам* (cross, traverse), *обикалям*, *заобикалям* (circle, skirt).

4 The conceptualisation of the route: Non-directed and directed motion

Table 17: Syntactic expression of the Traversing FEs in FrameNet

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>traverse</i>								
THEME	13		2		2			17
AREA		1						1
SOURCE			2					2
PATH	4	4	3		4	1		16
GOAL			6					6
PATH_SHAPE					17			17
DISTANCE							1	1
<i>descend</i>								
THEME	35				1			36
SOURCE			5					5
PATH	1	17	8		3			29
GOAL			9					9
PATH_SHAPE			1					1
DIRECTION			1	1				2
DISTANCE		1		1			1	3
<i>cross</i>								
THEME	53		2		2			57
AREA		1						1
SOURCE			4					4
PATH	4	26	6	4	16			56
GOAL			16				1	17
DIRECTION			4	1				5
<i>pass</i>								
THEME	20							20
AREA		1						1
SOURCE			1					1
PATH		3	14					17
DIRECTION			1					1
<i>circle</i>								
THEME	22							22
AREA		9			6			15
PATH		1	3					4
DIRECTION			1	1				2

Svetlozara Leseva

Table 18: FrameNet valence patterns of Traversing verbs

Pattern	#	Verbs
[NP.Ext] _{THM} [NP.Obj] _{PATH}	48	<i>descend, ascend, skirt, pass, cross, circle, mount</i>
[NP.Ext] _{THM} [PP] _{PATH}	40	<i>descend, ascend, skirt, pass, cross, hop, leap, circle, jump</i>
[NP.Ext] _{THM} [PP] _{GOAL} [NP.Obj] _{PATH}	14	<i>descend, ascend, skirt, cross</i>
[NP.Ext] _{THM} [NP.Obj] _{AREA} [] _{PATH_SHAPE-INC}	13	<i>crisscross</i>
[NP.Ext] _{THM} [] _{PATH-DNI}	11	<i>descend, ascend, cross</i>
[NP.Ext] _{THM} [NP.Obj] _{AREA}	10	<i>pass, circle, jump</i>
[NP.Ext] _{THM} [PP] _{SRC}	6	<i>descend, hop, jump</i>
[NP.Ext] _{THM} [] _{AREA-DNI}	6	<i>circle</i>
[NP.Ext] _{THM} [PP] _{GOAL} [] _{PATH-DNI}	5	<i>cross</i>

In addition there are a lot of Bulgarian verbs that represent lexicalisations of PATH-profiling formed by means of prefixation primarily from manner of motion verbs, which will be discussed in the next subsection along with similarly formed GOAL-profiling and SOURCE-profiling verbs.

Table 19 illustrates the syntactic realisation of several Bulgarian verbs evoking the frame Traversing. It can be noted that, like in English, for different verbs the preferred expression of the PATH may either be a direct object NP, e.g. *пребичам, прекосявам* (cross, traverse) or a prepositional (or adverbial) phrase (Example 19a), e.g. *пребика* (pass). Table 20 shows that some of the verbs that may be used both transitively and intransitively, favour the transitive (NP.Obj) realisation. The profiled element tends to be syntactically expressed, rather than left non-overt.

The verbs *качвам, изкачвам* (ascend) are always transitive (Example 19b), while *качвам се, изкачвам се, спускам се, слизам* are always intransitive (Examples 19c, 19d).

- (19) a. [Te]_{THM} **ПРЕСЯКОХА** [(през) двора]_{PATH}.
 They cross-PST.3PL (through) yard-DEF.
 ‘The boys crossed the yard.’

4 The conceptualisation of the route: Non-directed and directed motion

Table 19: Syntactic expression of the Traversing FEs in Bulgarian

	NP.Ext	NP.Obj	PP	AVP	NI	Clause	Other	Total
<i>пресичам/пресека</i>								
THEME	39							39
PATH		33	5		1			39
GOAL			2					2
<i>изкачвам/изкача</i>								
THEME	12							12
PATH		12						12
GOAL			4					4
<i>прекосявам/прекося</i>								
THEME	40							40
PATH		36	4					40
<i>преминавам/премина</i>								
THEME	15							15
PATH		4	11					15
<i>изкачваа се/изкача се</i>								
THEME	9							9
PATH			4		5			9
GOAL			6					6

- b. $[Te]_{\text{THM}}$ **ИЗКАЧИХА** $[панината]_{\text{PATH}}$.
 They climb-PST.3PL mountain-DEF.
 ‘They climbed the mountain.’
- c. $[Te]_{\text{THM}}$ **СЕ ИЗКАЧИХА** $[по панината]_{\text{PATH}}$.
 They climb-REFL.PST.3PL on mountain-DEF.
 ‘They climbed the mountain.’
- d. $[Te]_{\text{THM}}$ **СЛИЗАТ** $[по стълбите]_{\text{PATH}}$.
 They climb-PRS.3PL down stairs-DEF.
 ‘They descended the stairs.’

Although on a very small scale due to the size of the sample, the valence patterns show that GOALS are also realised syntactically (Table 20 and Example 20a). SOURCES (Example 20c) and DIRECTIONS (Example 20b) as well as combinations

Svetlozara Leseva

of motion-related FEs (Example 20c) are also attested as individual occurrences in the data.

Table 20: FrameNet valence patterns of Traversing verbs in Bulgarian

Pattern	#	Verbs
[NP.Ext] _{THM} [NP.Obj] _{PATH}	79	<i>прекосявам/прекося,</i> <i>пресичам/пресека,</i> <i>преминавам/премина,</i> <i>изкачвам/изкача</i>
[NP.Ext] _{THM} [PP] _{PATH}	23	<i>прекосявам/прекося,</i> <i>пресичам/пресека,</i> <i>преминавам/премина, изкачвам</i> <i>се/изкача се</i>
[NP.Ext] _{THM} [NP.Obj] _{PATH} [PP] _{GOAL}	6	<i>пресичам/пресека,</i> <i>изкачвам/изкача</i>
[NP.Ext] _{THM} [PP] _{GOAL} [] _{PATH-DNI}	5	<i>изкачвам се/изкача се</i>

- (20) a. [Тя]_{THM} **ПРЕКОСИ** [полето]_{PATH} [до крепостта]_{GOAL}.
 She cross-PST.3SG field-DEF to fortress-DEF.
 ‘She crossed the field towards the fortress.’
 b. [Той]_{THM} **ПРЕКОСИ** [залата]_{PATH} [по посока на
 He cross-PST.3SG hall-DEF in direction of
вратата]_{DIR}.
 door-DEF.
 ‘He crossed the hall towards the door.’
 c. [Тя]_{THM} **ПРЕСЕЧЕ** [моста]_{PATH} [от мидълсекския]_{SRC} [към
 Тя traverse-PST.3SG bridge-DEF from Middlesex-DEF to
сърейския бряг]_{GOAL}.
 Surrey-DEF shore.
 ‘She traversed the bridge from the Middlesex to the Surrey shore.’

The verb *слизам* (descend) can also co-occur with DISTANCES that may be expressed as measurement NPs (Example 21).

- (21) [Те]_{THM} **СЕ СПУСКАТ** [300 м]_{DIST}.
 They climb-PRS.3PL down 300 m.
 ‘They descend 300 m.’

4 The conceptualisation of the route: Non-directed and directed motion

5.5.19 Derivation of directional motion verbs

It has been well-established in the literature that part of the verbal prefixes in the Slavic languages yield (resultative) prefixed verbs when attached to unprefix (simplex) verbs (Beavers et al. 2010, Pantcheva 2007a,b, 2011, Palmer 2009, Spencer & Zaretskaya 1998, Svenonius 2005), as well as (Van Valin & LaPolla 1997: 178–184) for various languages, among many others. Regardless of the theoretical framework adopted and the specifics of the treatment of such verbs, the mechanism involves a verb with a simple internal (event, lexical semantic, logical) structure to which a prefix is attached so as to add a resultative subevent, thus producing a verb describing a more complex eventuality.

A typical example in the domain of motion is the prefixation of manner of motion verbs using directional prefixes, which, depending on the prefix, leads to the formation of GOAL-profiled, SOURCE-profiled or PATH-profiled verbs. As noted earlier, besides the verbs discussed in the previous sections, most of which are underived verbs with a primary directional motion meaning, there are a number of prefixed predicates derived mainly from simplex manner of motion verbs evoking frames such as Motion, Self_motion, Fluidic_motion, among others, which also evoke the frames Traversing, Arriving and Departing, and possibly other frames profiling the elements of the route of motion.

Table 21 shows the productivity of this pattern. The inventory of verbs evoking semantic frames profiling different elements of the route, is much richer than in English, where similar meanings may either be encoded by manner of motion verbs which have developed a more complex event structure and meaning (Example 22a) or expressed by means of certain syntactic constructions (Example 22b).

- (22) a. [He]_{SMOV} *was the first to FLY [the Atlantic]*_{PATH}.
 b. [He]_{SMOV} **LIMPED** [to the store]_{GOAL}.

6 Conclusions

In this chapter particular attention has been paid on the expression of the FEs that define the elements of the route traversed (SOURCE, GOAL, PATH) or region covered (AREA) by the moving entity and prominent aspects of the route such as the DISTANCE it spans, the DIRECTION it takes or its form (PATH_shape).

I showed and commented on the semantic specification, syntactic expression and valence patterns typical of manner of motion and directed motion verbs by

Svetlozara Leseva

Table 21: Prefixal derivation of directed motion verbs from manner of motion verbs in Bulgarian

Self_motion	SOURCE-profiled	GOAL-profiled	PATH-profiled
летя (fly)	отлитам (fly away), излитам, политам (fly off, take off)	долитам (fly up to), влитам (fly into)	прелитам (fly over)
хвърча (fly)	отхвърчавам (fly away), изхвърчавам (fly off)	дохвърчавам (fly (up) to)	прехвърчавам (fly over)
бягам (run)	избягвам (run away)	добягвам (run (up) to, dialect)	пробягвам, пребягвам (run, 'cover distance by running') пребягвам (run across)
тичам (run)	изтичам (run out of)	дотичвам (run (up) to)	претичвам ('cover distance by running')
пъзля (crawl)	изпъзлявам (crawl out)	допъзлявам (crawl (up) to) пропъзлявам (crawl in, crawl onto) впъзлявам (crawl into)	препъзлявам, пропъзлявам (crawl across), пропъзлявам ('cover distance by crawling')
скачам (jump)	изскачам (jump out)	доскачам (jump (up) to)	прескачам (jump, pass over)
плувам (swim)	изплувам (swim up, swim to the surface)	доплувам (swim (up) to) вплувам (swim into)	преплувам (swim across) проплувам ('cover distance by swimming')
нижа се (file)	изнизвам се (file out)		
газя (wade)		догазвам (wade (up) to)	изгазвам, преразвам ('pass through some substance by wading') изгазвам ('cross by wading')
танцувам (dance)		дотанцувам (dance (up) to)	
клатушкам се		доклатушкам се (totter (up) to)	
куцам, куцукам (limp)		докуцам, докуцукам (limp (up) to)	

4 *The conceptualisation of the route: Non-directed and directed motion*

analysing the examples in the FrameNet corpus and expanding the observations to Bulgarian examples.

Manner of motion verbs tend to express the PATH over the GOAL and especially over the SOURCE, but the particular distribution of the various patterns varies across verbs. PATH is especially prominent where complex notions of motion or trajectory are involved.

The data corroborate the observations made in the literature, that all other things being equal, there is a bias for expressing GOALS over SOURCES, a tendency which has been studied for many typologically distinct languages. In particular, if the verbs do not profile a particular aspect of the route, they tend to express GOALS over SOURCES, the intuition being that motion through space involves getting to some place, even with manner of motion verbs, and that, in this respect, the end point of the motion is a more salient feature than the starting point.

Across verbs that profile a particular aspect of the route, the profiled FE that tends to be expressed, i.e. SOURCE-profiling verbs co-occur more frequently with SOURCE expressions than verbs that do not profile this FE, GOAL-profiling verbs co-occur with GOAL expressions. While these two aspects have been of primary interest in the linguistic literature, similar observations may be made for PATH and to a lesser extent for AREA (as the examples are fewer), judging from the data.

DISTANCES and DIRECTIONS are rarely expressed and at least in some cases they show to be syntactically dependent on the PATH, as well as semantically dependent on it, as they represent elaborations on certain aspects of it (deictic or geographical orientation or the length of the route covered).

Other elements of the route may be expressed besides or instead of the profiled one. SOURCE-profiling (Departing) verbs tend to realise GOALS or PATHS, but the preference for one over the other varies across verbs and the examples are not always definitive. PATH-profiling verbs tend to favour GOALS over SOURCES. In addition, the following was observed in the Bulgarian data: peripheral GOALS may be expressed on a par with profiled SOURCES. Each of these observations warrants further investigation, especially with respect to the frequency and means of expression (including the available inventories) of various FE combinations within and across verbs and frames.

While only marked in passing, the productivity of prefixal derivation as a mechanism of deriving directed motion verbs from other motion verbs, especially from manner verbs, in Bulgarian (and other Slavic and non-Slavic languages) points to the need for these verbs to be systematically addressed within the FrameNet structure. This may also result in the definition of frame-to-frame relations that account for this systematicity.

Svetlozara Leseva

Abbreviations

AVP	Adverbial phrase	NP	Noun phrase
CNI	Constructional null instantiation	NP.Ext	Subject NP
DEF	Definite form	NP.Obj	Object NP
Dir	DIRECTION	PL	Plural
Dist	DISTANCE	PP	Prepositional phrase
DNI	Definite null instantiation	PRS	Present tense
FE	frame element	PST	Past tense
IMPF	Imperfective aspect	PWN	Princeton WordNet
INDF	Indefinite form	SG	Singular
INI	Indefinite null instantiation	SMov	SELF_MOVER
LU	Lexical unit	Src	SOURCE
NEG	Negative form	Thm	THEME

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

References

- Aske, Jon. 1989. Path predicates in English and Spanish: A closer look. In Kira Hall, Michael Meacham & Richard Shapiro (eds.), *Proceedings of the fifteenth annual meeting of the Berkeley Linguistics Society*, 1–14. Berkeley, CA: BLS.
- Baker, Collin F. 2008. FrameNet, present and future. In Jonathan Webster, Nancy Ide & Alex Chengyu Fang (eds.), *Proceedings of the First International Conference on Global Interoperability for Language Resources*. Hong Kong: City University.
- Baker, Collin F. & Christiane Fellbaum. 2009. WordNet and FrameNet as complementary resources for annotation. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, 125–129. DOI: 10.3115/1698381.1698402.
- Beavers, John, Beth Levin & Shiao Wei Tham. 2010. The typology of Motion events revisited. *Journal of Linguistics* 46. 331–377.
- Borin, Lars, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj & Dimitrios Kokkinakis. 2010. The past meets the present in Swedish FrameNet++. In *Proceedings of the 14th EURALEX international congress*, 269–281.

4 *The conceptualisation of the route: Non-directed and directed motion*

- Burchardt, Aljoscha, Katrin Erk & Anette Frank. 2005. A WordNet detour to FrameNet. In Bernhard Fisseni, Hans-Christian Schmitz, Bernhard Schröder & Petra Wagner (eds.), *Sprachtechnologie, mobile kommunikation und linguistische ressourcen*, vol. 8 (Computer Studies in Language and Speech), 408–421. Frankfurt: Lang.
- Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó & Manfred Pinkal. 2006. The SALSA corpus: A German corpus resource for lexical semantics. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2006/pdf/339_pdf.pdf.
- Candito, Marie, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël de Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathieu, Philippe Muller, Benoît Sagot & Laure Vieu. 2014. Developing a French FrameNet: Methodology and first results. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 1372–1379. Reykjavik, Iceland: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/496_Paper.pdf.
- Cristobal, Maria. 2010. Arriving events in English and Spanish: A contrastive analysis in terms of frame semantics. *Estudios de lingüística del español* 31. 1–77.
- Croft, William B., Jóhanna Barðdal, Willem Hollmann, Violeta Sotirova & Chiaki Taoka. 2010. Revising talmy's typological classification of complex events. In Hans C. Boas (ed.), *Contrastive construction grammar*, 201–235. Amsterdam: John Benjamins.
- de Lacalle, Maddalen Lopez, Egoitz Laparra & German Rigau. 2014. Predicate matrix: Extending SemLink through WordNet mappings. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 903–909. Reykjavik, Iceland: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/589_Paper.pdf.
- Dekova, Rositsa & Petya Nestorova. 2010. Formal description of some intransitive verbs of non-directed movement in the Bulgarian FrameNet. In *Proceedings of the seventh international conference formal approaches to South Slavic and Balkan languages (FASSBL), 4 – 6 october 2010, dubrovnik, Croatia*, 13–18.
- Dirven, René & Marjolyn Verspoor. 1998. *Cognitive exploration of language and linguistics* (Cognitive Linguistics in Practice 1). Amsterdam: John Benjamins.
- Fellbaum, Christiane (ed.). 1999. *WordNet: An electronic lexical database*. Cambridge: MIT Press.

Svetlozara Leseva

- Filipović, Luna. 2007. *Talking about Motion: A crosslinguistic investigation of lexicalization patterns*. Amsterdam: John Benjamins.
- Fillmore, Charles J. 1971. *Santa Cruz lectures on deixis*. Bloomington IN: Indiana University Linguistics Club.
- Fillmore, Charles J. 2003. Valency and semantic roles: The concept of deep structure case. In Vilmos Ágel, Ludwig Eichinger, Hans Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependenz und valenz: Ein internationales Handbuch der zeitgenössischen Forschung*, vol. 1, 457–475. Berlin: Walter de Gruyter.
- Fillmore, Charles J., Collin F. Baker & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the Conference COLINGACL '98, Montreal, Canada*, 86–90.
- Fillmore, Charles J., Charles Wooters & Collin F. Baker. 2001. Building a large lexical databank which provides deep semantics. In *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*, 3–26. Hong Kong, China: City University of Hong Kong. DOI: <http://hdl.handle.net/2065/12202>. <https://aclanthology.org/Y01-1001>.
- Gilardi, Luca & Collin F. Baker. 2018. Learning to align across languages: Toward multilingual FrameNet. In Tiago T. Torrent (ed.), *Proceedings of the International FrameNet Workshop 2018: Multilingual FrameNets and Constructicons*, 13–22.
- Gruber, Jeffrey Steven. 1965. *Studies in lexical relations*. Massachusetts Institute of Technology. (Doctoral dissertation).
- Gruzitis, Normunds, Gunta Nespor-Berzkalne & Baiba Saulite. 2018. Creation of Latvian FrameNet based on universal dependencies. In *Proceedings of the International FrameNet Workshop (IFNW)*, 23–27. http://lrec-conf.org/workshops/lrec2018/W5/pdf/9_W5.pdf.
- Hahm, Young Gyun, Youngsik Kim, Yousung Won, Jongsung Woo, Jiwoo Seo, Jiseong Kim, Seongbae Park, Dosam Hwang & Key-Sun Choi. 2014. Toward matching the relation instantiation from DBpedia ontology to Wikipedia text: Fusing FrameNet to Korean. In Harald Sack, Agata Filipowska, Jens Lehmann & Sebastian Hellmann (eds.), *Proceedings of the 10th International Conference on Semantic Systems, SEMANTiCS 2014, Leipzig, Germany, September 4-5, 2014*, 13–19. ACM. DOI: [10.1145/2660517.2660534](https://doi.org/10.1145/2660517.2660534).
- Ibarretxe-Antuñano, Iraide. 2004. Language typologies in our language use: The case of Basque Motion events in adult oral narratives. *Cognitive Linguistics* 15. 317–349.

4 The conceptualisation of the route: Non-directed and directed motion

- Ikegami, Yoshihiko. 1987. 'source' vs. 'goal': A case of linguistic dissymmetry. In René Dirven & Günter Radden (eds.), *Concepts of case* (Studien Zur Englischen Grammatik 4), 122–146. Tübingen: Narr.
- Imani, Zolfa & Rezvan Motavalian Naeini. 2020. Motion verbs in Persian and English: A Framenet-based contrastive analysis. *Facta Universitatis. Linguistics and Literature* 8. 195–205. DOI: [10.22190/FULL2002195I](https://doi.org/10.22190/FULL2002195I).
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge: MIT Press.
- Johnson, Christopher R., Charles J. Fillmore, Esther J. Wood, Margaret Urban, Miriam R. L. Petruck, Collin F. Baker & Charles J. Fillmore. 2001. *The FrameNet project: Tools for lexicon building*. <https://citeseerx.ist.psu.edu/pdf/0ece390b6f4e6b38c5733248992ff73f846d91aa>.
- Koeva, Svetla. 2008. Bulgarian FrameNet. Semantic-Syntactic dictionary of Bulgarian – Conceptual Model (in Bulgarian). In *Bulgarian FrameNet: Semantic-Syntactic Dictionary of Bulgarian*, 5–51. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2010. *Bulgarian FrameNet (in Bulgarian)*. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2021a. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla. 2021b. Towards expanding WordNet with conceptual frames. In *Proceedings of the 11th Global Wordnet Conference*, 182–191.
- Koeva, Svetla & Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, 203–213. Sofia, Bulgaria: Department of Computational Linguistics, IBL – BAS. <https://aclanthology.org/2022.clib-1.23>.
- Koeva, Svetla, Svetlozara Leseva, Borislav Rizov, Ekaterina Tarpomanova, Tsvetana Dimitrova, Hristina Kukova & Maria Todorova. 2011. Design and development of the Bulgarian sense-annotated corpus. In María Luisa Carrió Pastor & Miguel Ángel Candel Mora (eds.), *Proceedings of the III international congress of corpus linguistics. Information and communications technologies: Present and future in corpus analysis*, 143–150.
- Koeva, Svetla, Svetlozara Leseva & Maria Todorova. 2006. Bulgarian sense tagged corpus. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, 79–86.

Svetlozara Leseva

- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Rositsa Dekova, Tsvetana Dimitrova & Ekaterina Tarpomanova. 2012. The Bulgarian national corpus: Theory and practice in corpus design. *Journal of Language Modelling* 0(1). 65–110. DOI: 10.15398/jlm.v0i1.33.
- Kopecka, Anetta. 2010. Motion events in Polish. In Victoria Hasko & Renee Perel-mutter (eds.), *New approaches to Slavic verbs of motion* (Studies in Language Companion Series 115), 225–246. Amsterdam: John Benjamins.
- Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Landes, Shari, Claudia Leacock & Randee Tengi. 1998. Building semantic concordances. In Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*. Cambridge, MA: The MIT Press. DOI: 10.7551/mitpress/7287.003.0015.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Laparra, Egoitz & German Rigau. 2010. Extended WordFramenet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, 1214–1219.
- Leseva, Svetlozara & Ivelina Stoyanova. 2019. Structural approach to enhancing WordNet with conceptual frame semantics. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 629–637. Varna: INCOMA.
- Leseva, Svetlozara & Ivelina Stoyanova. 2020a. Beyond lexical and semantic resources: Linking WordNet with FrameNet and enhancing synsets with conceptual frames. In Svetla Koeva (ed.), *Towards a semantic network enriched with a variety of semantic relations*, 21–48. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Leseva, Svetlozara & Ivelina Stoyanova. 2020b. Towards a conceptual description of verbs. *Cybernetics and Information Technologies*. 4th ser. 20. 108–124.
- Leseva, Svetlozara & Ivelina Stoyanova. 2022a. 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)*, 214–223. Institute for Bulgarian Language.
- Leseva, Svetlozara & Ivelina Stoyanova. 2022b. Stative verbs: Conceptual structure, hierarchy, systemic relations. In Svetla Koeva, Elena Ivanova, Yovka Tishcheva & Anton Zimmerling (eds.), *Ontologiya na situatsiite za sastoyanie – lingvistichno modelirane. Sapostavitelno izsledvane za balgarski i ruski*, 68–114. Sofia: Prof. Marin Drinov Publishing House of Bulgarian Academy of Sciences.

4 The conceptualisation of the route: Non-directed and directed motion

- Leseva, Svetlozara, Ivelina Stoyanova & Maria Todorova. 2018. Classifying verbs in WordNet by harnessing semantic resources. In *Proceedings of CLIB 2018, Sofia, Bulgaria*.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Levin, Beth. 2015. Verb classes within and across languages. In Bernard Comrie & Andrej Malchukov (eds.), *Valency classes in the world's languages*, 1627–1670. Berlin: De Gruyter.
- Lindsey, Traci Speed. 2011. *Bulgarian verbs of Motion: Slavic verbs in a Balkan context*. University of California at Berkeley. (Doctoral dissertation). <https://api.semanticscholar.org/CorpusID:60635322>.
- Łozinska, Joanna. 2018. *Path and manner saliency in Polish in contrast with Russian: A cognitive linguistic study*. Leiden, The Netherlands: Brill. DOI: 10.1163/9789004360358.
- Miller, George A. 1995. WordNet: A lexical database for English. *Commun. ACM* 38(11). 39–41.
- Miller, George A., Martin Chodorow, Shari Landes, Claudia Leacock & Robert G. Thomas. 1994. Using a semantic concordance for sense identification. In *Proceedings of a Workshop on Human Language Technology, Plainsboro, New Jersey, 1993*. <https://aclanthology.org/H93-1061>.
- Miller, George A., Claudia Leacock, Randee Teng & Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of a Workshop on Human Language Technology, Plainsboro, New Jersey, 1993*. <https://aclanthology.org/H93-1061>.
- Nestorova, Petya. 2009. Argument structure of motion verbs in the Bulgarian FrameNet (settle, populate, evict, resettle) argumentna (in Bulgarian). *Research Works – Plovdiv University Paisii Hilendarski, Filology* 47. 101–108.
- Nitsolova, Ruselina. 1984. *Pragmatic aspect of the sentence in the Bulgarian literary language (in Bulgarian)*. Sofia: Narodna prosveta.
- Ohara, Kyoko, Seiko Yamaguchi Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito & Shun Ishizaki. 2004. The Japanese FrameNet project: An introduction. In <https://api.semanticscholar.org/CorpusID:15338426>.
- Palmer, Martha. 2009. SemLink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*. 9–15.
- Palmer, Martha, Claire Bonial & 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*, 13–17. Association for Computational Linguistics.

Svetlozara Leseva

- Pantcheva, Marina. 2007a. Bulgarian Motion verbs and prefixation. In CASTL, University of Tromsø. https://www.researchgate.net/publication/228797054_Bulgarian_Motion_Verbs_and_Prefixation.
- Pantcheva, Marina. 2007b. Bulgarian spatial prefixes and event structure. *Tromsø Working Papers on Language & Linguistics: Nordlyd: special issue on Space, Motion, and Result* 34(2). 320–344.
- Pantcheva, Marina. 2011. *Decomposing path: The nanosyntax of directional expressions*. University of Tromsø. (Doctoral dissertation).
- Papafragou, Anna, Christine Massey & Lila Gleitman. 2002. Shake, rattle, ‘n’ roll: The representation of motion in thought and language. *Cognition* 84. 189–219.
- Pedersen, Bolette, Sanni Nimb, Anders Søgaard, Mareike Hartmann & Sussi Olsen. 2018. A Danish FrameNet lexicon and an annotated corpus used for training and evaluating a semantic frame classifier. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA). <https://aclanthology.org/L18-1378>.
- Petruck, Miriam R. 2015. *The components of FrameNet. Tutorial at NAACL 2015*. <http://naacl.org/naacl-hlt-2015/tutorial-framenet-data/FNComponentsMRLP.pdf>.
- Petruck, Miriam R. & Gerard de Melo. 2012. Precedes: A semantic relation in FrameNet. In *Proceedings of the Workshop on Language Resources for Public Security Applications*, 45–49.
- Rakhilina, Ekaterina V. 2004. There and back: The case of Russian ‘go’. *Glossos* 5. 1–33.
- Rohde, Ada. 2001. *Analyzing path: The interplay of verbs, prepositions and constructional semantics*. Houston, TX: Rice University.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin. F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.
- Slobin, Dan. 1996. Two ways to travel: Verbs of Motion in English and Spanish. In Masayoshi Shibatani & Sandra A. Thompson (eds.), *Grammatical constructions: Their form and meaning*, 195–219. Oxford: Oxford University Press.
- Slobin, Dan. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Stronqvist & Ludo Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives*, 219–257. Mahwah, NJ: Erlbaum.
- Speed, Traci. 2015. Manner/path typology of Bulgarian Motion verbs. *Journal of Slavic Linguistics* 23. 51–81.

4 The conceptualisation of the route: Non-directed and directed motion

- Spencer, Andrew & Marina Zaretskaya. 1998. Verb prefixation in Russian as lexical subordination. *Linguistics* 36(1). 1–40.
- Stanisheva, Dina. 1985. The use of motion verbs taking into account the addressee in Bulgarian (in Russian). *Zeitschrift für Slavistik* 30(1-6). 78–84.
- Stefanowitsch, Anatol & Ada Rohde. 2004. Studies in linguistic motivation. In Günter Radden & Klaus-Uwe Panther (eds.), chap. The goal bias in the encoding of motion events, 249–268. Berlin/New York: De Gruyter Mouton.
- Stoyanova, Ivelina & Svetlozara Leseva. 2021. Semantic description of verbs for change and hierarchical organization of conceptual frames. In *Proceedings of the Annual International Conference of the Institute of Bulgarian Language Prof. Lubomir Andreychin (Sofia, 2021)*, vol. 2, 76–85.
- Subirats, Carlos. 2009. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In Hans C. Boas (ed.), *Multilingual FrameNets in computational lexicography: Methods and applications*, 135–162. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110212976.2.135.
- Svenonius, Peter. 2005. Slavic prefixes inside and outside VP. *Nordlyd (Tromsø, Norway)* 32(2). 205–253.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (ed.), *Language typology and syntactic description, vol. III: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 1991. Path to realization: A typology of event integration. *Buffalo Working Papers in Linguistics* 91. 147–187.
- Talmy, Leonard. 2000. *Toward a cognitive semantics: Typology and process in concept structuring*, vol. 2. Cambridge, MA: MIT Press.
- Taremaa, Piia. 2017. *Attention meets language: A corpus study on the expression of motion in Estonian* (Dissertationes Linguisticae Universitas Tartuensis 29). Tartu: University of Tartu Press.
- Taremaa, Piia. 2021. Verbs of horizontal and vertical motion: A corpus study in Estonian. *Finnish Journal of Linguistics* 34. 221–256. <https://journal.fi/finjol/article/view/103097>.
- Tchizmarova, Ivelina. 2015. Bulgarian verbs of change of location. *Journal of Slavic Linguistics* 15. 109–147.
- Tonelli, Sara & Daniele Pighin. 2009. New features for FrameNet – WordNet mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL'09)*, Boulder, USA.
- Torrent, Tiago T., E. Matos, Ludmila Lage, Adrieli Laviola, Tatiane Tavares, Vânia Almeida & Natália Sigiliano. 2018. Towards continuity between the lexicon and the constructicon in FrameNet brasil. In Benjamin Lyngfelt, Lars Borin, Kyoko

Svetlozara Leseva

- Ohara & Tiago T. Torrent (eds.), *Constructicography: Constructicon development across languages*, vol. 22 (Constructional Approaches to Language), 107–140. Amsterdam: Benjamins. DOI: 10.1075/cal.22.04tor.
- Trifonova, Yordanka. 1982. About the opposition ida/hodya in modern Bulgarian (in Bulgarian). *Language and Literature* 4. 105–114.
- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Verkerk, Annemarie. 2017. The goal-over-source principle in European languages. In Silvia Luraghi, Tatiana Nikitina & Chiara Zanchi (eds.), *Space in diachrony*, 1–40. Amsterdam: John Benjamins.
- Viberg, Åke. 2008. Riding, driving and traveling: Swedish verbs describing motion in a vehicle in crosslinguistic perspective. In Joakim Nivre, Mats Dahllöf & Beáta Megyesi (eds.), *Resourceful language technology: Festschrift in honor of Anna sågvall hein*, 173–201. Uppsala: Acta Universitatis Upsaliensis.
- Viberg, Åke. 2015. Motion verb typology and the expression of the endpoint of motion in Swedish. In Maria Bloch-Trojnar, Anna Malicka-Kleparska & Karolina Drabikowska (eds.), *Concepts and structures* (Studies in Linguistics and Methodology, Vol. 8), 209–229. Lublin: Wydawnictwo KUL.
- Vossen, Piek, Antske Fokkens, Isa Maks & Chantal van Son. 2018. Towards an open Dutch FrameNet lexicon and corpus. In Tiago T. Torrent, Lars Borin & Collin F. Baker (eds.), *Proceedings of the International FrameNet Workshop 2018: Multilingual Framenets and Constructicons (LREC 2018)*, 75–80.
- Wälchli, Bernhard & Fernando Zúñiga. 2006. Source-goal (in)difference and the typology of motion events in the clause. *Language Typology and Universals* 59(3). 284–303. DOI: 10.1524/stuf.2006.59.3.284.
- You, Liping & Kaiying Liu. 2005. Building Chinese FrameNet database. In *2005 international conference on natural language processing and knowledge engineering*, 301–306. <https://api.semanticscholar.org/CorpusID:10064243>.

Chapter 5

Frame semantics and verbs of contact

Maria Todorova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

The article offers a semantic description of a group of verbs from WordNet with the semantic primitive ‘verbs of contact’, which belong to the set of common vocabulary in Bulgarian. We present the result of their annotation with conceptual frames using, the information for both the semantic hierarchy from BulNet (Koeva 2021b) and WordNet respectively, and the semantic frames from FrameNet (Fillmore 1982, Fillmore et al. 1998) – original or adapted. The predicates of contact are divided into the major classes of Verbs of physical contact via motion and Verbs of physical contact in state. Using information from the hierarchical structure of WordNet, Semantic frames from FrameNet and based on the observations on the selective specifics and the syntactic realisation of the concrete meaning, conceptual subtypes of contact verbs are offered.

1 Introduction

This paper aims at a conceptual description of 450 high-frequency Bulgarian verbs categorised in WordNet (Fellbaum 1990) as verbs of contact. Their selection is based both on their participation in the set of general lexis of Bulgarian and on their thematic coverage.

Verb classes are defined in the linguistic literature as coherent sets of verbs sharing similar semantic properties, such as membership in a common semantic domain (Juffs 1996) or similar argument realisation and semantic interpretation (Fillmore 1970, Levin 1993, Kipper-Schuler 2005). Fillmore (1970) emphasises the importance of verb classes in the organisation of the verb lexicon and the study of the patterns of shared verb behaviour, as well as in the identification of the



Maria Todorova

grammatically relevant elements of meaning (Fillmore 1970: 125). At the same time, the Theory of Frame Semantics (Fillmore 1977, Ruppenhofer et al. 2016) is based on the interdependency of the lexicon and grammar of a language. It characterises the semantic and syntactic properties of predicates by relating them to semantic frames. The semantic arguments of a predicate corresponds to frame elements in frames describing their semantics.

Based on this, we study the semantics of various verbs of contacts, combining the semantic classification of verbs in WordNet with the information from FrameNet (Fillmore 1982, Fillmore et al. 1998) to group them into conceptual frames. Conceptual frames (Koeva 2021c) are abstract structures representing, on the one hand the syntagmatic relation between a predicate and its arguments, and, on the other hand, the relation of set of predicates (verbs) and sets of their arguments to the verbal and nominal synonym sets in WordNet. In this way, they generalise the application of lexical information from WordNet in the semantic frames from FrameNet.

As a result, we offer a semantic and syntactic description and classification of the Bulgarian verbs of contact compared with their English correspondences. The grouping in conceptual subtypes is based on implications about the conceptual description of the relevant verbs. The observations are based on the data from Bulgarian WordNet (BulNet) (Koeva 2021b) and from Princeton WordNet (Fellbaum 1999) respectively and on the process of enlargement of the Bulgarian FrameNet (Koeva 2010) with conceptual frames (Koeva 2021c), as well as the creation of BulFrame – a linked semantic and syntactic resource for Bulgarian (Koeva & Doychev 2022).

WordNet and FrameNet are large lexical resources that provide semantic information regarding a wide variety of features for verbs. WordNet represents a multilingual conceptual network of synonym sets (synsets) linked by means of semantic relations, such as hypernymy, antonymy, etc., and offers sets of semantic classes of verbs and nouns. FrameNet represents the semantics of lexemes by means of schematic representations (frames) describing objects, situations, or events, and their components (frame elements) in the frame semantics apparatus.

The remainder of the paper is organised as follows: Section 2 describes the data used in the process of annotation – a set of verbs of contact from WordNet and a set of semantic frames from FrameNet; Section 3 presents an overview of the related descriptions and classifications of the verbs under consideration; Section 4 discusses the semantic features of verbs of contact and their lexical semantic subtypes, and offers examples on the annotation of verbs of contact with semantic frames; Section 5 comments on the valency patterns of the studied

5 Frame semantics and verbs of contact

verbs; Section 7 sum up the observations on the results and suggests directions for future work.

1.1 Verbs of contact

The grouping of verbs into semantic classes varies depending on the theory adopted and the focus of the description. There is a vast diversity of descriptions and definitions of the verbs of contact. In general, the notion of CONTACT is understood as a “conceptual core element” of a predicate (Juffs 1996: 47). Exploring alternation patterns in English, Fillmore (1977: 125) defines two large classes: Break verbs and Hit verbs, where Hit verbs involve (often forceful) contact with an entity, without entailing a change in its state. Analysing verb valency in different languages, Levin (1993) defines a manner/result verb distinction. Verbs like *hit* that describe making surface contact with an object via forceful impact are MANNER(/means) verbs and describe ways of potentially damaging objects Levin (2015).

The set of verbs of contact in WordNet features one of the 15 lexicographer’s files, in which the verbs in WordNet are grouped according to the semantic domain to which they pertain. The set of verbs of contact is defined as “verbs of touching, hitting, tying, digging” (Miller 1995, Fellbaum 1990). It is also the largest of them, consisting of more than 2100 verbal synsets, out of all 14103 Bulgarian verbal synsets, and includes event and action verbs that share the semantic component of CONTACT or IMPACT.

This verb set type casts taxonomic framework by means of the hyponymy (troponymy) relation, which covers a number of different manner relations (Fellbaum 1990). The semantic definition of the class is fuzzy and does not really summarise the semantics of all the verbs it contains. Therefore we make an attempt to define typical subclasses within the class, using both the syntactic and semantic information from WordNet and FrameNet.

1.2 Related work

Verbs of contact are heterogeneous and overlapping as a semantic class, and thus they are less studied than other verb classes. They have been an object of research for English (Fillmore 1970, Levin 1993, Fellbaum 1990), Chinese (Gao & Cheng 2003), Swedish (Viberg 2004). Fillmore (1970) focuses on two large classes of verbs of contact, *break* and *hit*, whose members share elements of meaning and patterns of behaviour. A class of contact verbs was also defined by Levin

Maria Todorova

(1993) in her semantic classification on the basis of a number of alternations reflecting the correlation between the semantics and the syntactic behaviour of the verbs, and the interpretation of their arguments. In particular, Levin (1993: 148-156) defines a class of Verbs of contact by impact with a number of subclasses: *Hit verbs*; *Spank verbs*; *Swat verbs*; *Non-agentive verbs*. Vulchanova & Dekova (2009) represent a corpus and an empirically-derived classification of verbs of contact by impact using the Sign model formalism. Individual subtypes of the class were also described by some authors: *Physical contact verbs* (Gao 2001) and *Hit and Spank verbs of contact by impact* (Gao & Cheng 2003). These descriptions partially overlap with the classification of verbs adopted in WordNet; only a few of them are focused on the semantic frames in FrameNet of some verbs of contact.

Previous work on the conceptual semantic descriptions of Bulgarian verbs involves the analysis of verbs of change (Stoyanova & Leseva 2021) and verbs of communication (Kukova 2020), verbs of movement (Kostova 2010), mental state predicates (Stamenov 2021, 2022, Tisheva 2021, Dzhonova 2028), and a description of syntactic transformations of Bulgarian verbs (Koeva 2004, 2021a, 2022). Different stages of the study of semantic features and selectional restrictions relevant to the semantic description of Bulgarian verbs and their frame elements are explored in Svetlozara et al. (2020) and Leseva et al. (2019). As far as we know, the set of verbs denoting physical contact has not been described as a separate semantic class for Bulgarian or compared with their English correspondences so far.

2 The data analysed

As already mentioned, the analyses of the verbs in this work are based on the semantic descriptions and relational hierarchy of WordNet, and the semantic frames from FrameNet. The combined information available in the resources gives opportunity for a rich representation of the paradigmatic and syntagmatic aspects of lexical semantics (Baker & Fellbaum 2009). Thus, the semiautomatic mapping of FrameNet frames to WordNet synsets, described in Stoyanova & Leseva (2020), is used.

After the selected set of verbs of contact was extracted from WordNet, it was subsequently filtered, so as to include only verbs belonging to the general lexis of Bulgarian. The selection was made with a view to the theoretical semantic description and typology of verb predicates belonging to the set of common vocabulary in the Bulgarian language (Koeva & Doychev 2022). The collection was excerpted via the intersection of a set of 44000 English verbs selected according

5 *Frame semantics and verbs of contact*

to the AoA (age of acquisition) criterion (Brysbaert & Biemiller 2017), and a subset of verbs derived from the Bulgarian lexical-semantic network BulNet, aligned with the Princeton WordNet. The resulting verb set was additionally evaluated based on (i) a relevant label in case the corresponding synsets belong to the list of the so-called base concepts, or BCS¹; (ii) frequency information on the use of the verbs derived from the Bulgarian National Corpus (Koeva et al. 2012). The procedure of selection, described in more details in Koeva & Doychev (2022) and (Dimitrova et al. 2022), resulted in a list of over 5000 general-lexis verbs²; 804 of them are assigned the prime verb.contact, and 486 were semi-automatically assigned 107 unique frames from FrameNet. These verbs were additionally validated, and their set was filtered with a view to frequency and thematic coverage. The verbs with metaphorical or figurative meaning, as well as verbs expressing personal connections and emotional contacts were excluded.

The resulting set of 450 verbs of contact constitutes the starting set selected for annotation with semantic frames, which is described in this paper. Using this set, we propose a semantic description of verbs of contact in Bulgarian based on their frame elements, their selectional restrictions (represented in terms of semantic classes of nouns in WordNet), and their syntactic realisation in a context as well as their classification.

3 **Annotation of verbs of contact and semantic frames assignment**

The annotation of Bulgarian contact verbs with semantic frames, and the description of their frame elements and the relevant semantic restrictions are carried out by means of the software system BulFrame, specifically designed for the definition and description of conceptual frames (Koeva & Doychev 2022). The semantic restrictions imposed on the verb's arguments were aligned with a particular subtree(s) of noun synsets in WordNet, and draw on previous efforts described in Leseva et al. (2018) and Dimitrova & Stefanova (2019). The annotation of the selected verbs with BulFrame includes the following steps:

- (a) Each verb is assigned morphosyntactic information (transitive and intransitive, reflexive verbs, 3rd person verbs)

¹a WordNet subset, defined within EuroWordNet and BalkaNet projects <http://globalwordnet.org/resources/gwa-base-concepts/>

²The data is available on https://dcl.bas.bg/projects_list/enriching-wordnet/

Maria Todorova

- (b) Each verb is assigned a FrameNet frame (as is), a FrameNet frame that has been modified to better reflect the semantics of the verbs under discussion or a newly formulated frame.
- (c) For each frame element in a given frame, a linguist assigns a syntactic functions description (subject, object, adjunct).
- (d) For each frame element in a given frame, a linguist checks the validity of the general selectional restrictions assigned to it. The general semantic restrictions, describing the compatibility between semantic classes of verbs and nouns corresponding to their arguments, proposed in Leseva et al. (2019) and Dimitrova & Stefanova (2019), are aligned with the noun synsets representing the roots of the subtrees in WordNet. Those restrictions, when assigned to a frame, give a first approximation of the semantic specification of the frame elements. When a general restriction is assigned, all hyponyms of the noun synsets, representing the roots of the relevant subtrees, are checked as potential candidates for the frame element.
- (e) Each verb is examined individually in order to specify additional selectional restrictions from WordNet if needed. Specific restrictions on the lexical realisation of the frame elements are represented as individual WordNet synsets.

4 Semantic features of verbs of contact

In this Section, the semantic characterisation of verbs of contact and their division into subclasses proposed by Miller (1990) is used with a view to the WordNet hierarchy, in combination with the semantic information from FrameNet.

In order to group the selected verbs of contact into conceptual subtypes based on both the realisation of their frame elements and on their lexical semantics, we rely on the prerequisite that verbs with similar verb meanings share characteristic argument realisations.

As already mentioned, the definition of conceptual subtypes relies on the conceptual frames Koeva (2021c). A particular conceptual frame is associated with predicate(s) from a particular semantic class, and each element of the conceptual frame is associated with a set of nouns, compatible with the predicate(s).

4.1 Lexical semantic subtypes

Being the largest class of verbs in WordNet, the set of contact verbs is well-represented in the selection of Bulgarian general lexis verbs – nearly 16% of the

5 *Frame semantics and verbs of contact*

whole set (2179 synsets, labeled as *verb.contact*, out of 13766 verbal synsets in WordNet). Most of the contact verbs are hyponyms of the following verbal root synonym sets within the structure of WordNet: *fasten*, *attach*, *cover*, *cut*, and *touch*, which results in a large tree structure within the set. Based on the WordNet hypernym relation, Miller (1990: 59) define the following subgroups of contact verbs:

- (a) Verbs encoding force, intensity, or iteration of the action (*hit*).
- (b) Verbs of holding (*grab*, *squeeze*, *pinch*) and touching (*paw*, *finger*, *stroke*, *poke*).
- (c) Verbs involving an instrument or a material argument (*paint*).
- (d) Verbs involving a body part argument indicating what kind of contact action the body part is typically used for: *shoulder* (*support*, *carry*); *elbow* (*push*); *finger*, *thumb* (*touch*, *manipulate*).

The semantics of those classes corresponds to the semantics of some of the FrameNet frames. The set of WordNet verbs encoding force under the verb roots *hit* and *destroy* is described by the frames *Impact* and *Destroying* in 4.3.3; the WordNet verbs of holding and verbs involving, a body part, material, or instrument correspond to the semantic frames within the class of verbs of contact denoting displacement in 4.3.1. The State verbs of physical contact described in 4.4 belong to smaller subtrees within WordNet.

4.2 *Frame semantic subtypes*

The generalised conceptual frame Koeva (2021c) of verbs of contact, represent different actions or events where two or more entities come into physical contact with each other. This conceptual frame encompasses the semantic frames of a wide range of verbs, representing various aspects of contact and valency relations between subjects, objects, or entities.

The generalised characteristics of the frame elements summarises the core and some of non-core frame elements of groups of semantic frames. The semantics of contact is encoded in the core frame elements, representing the entities in contact involved in the situation, often referred to as the subject (the one initiating the contact) and the object (the one being contacted), but also in the frame elements representing an entity mediator for the contact – instrument or connector. The frame elements of some verbs of contact specify the direction of the contact, such

Maria Todorova

as verbs that describe one entity reaching out to touch or affect another entity as GOAL and PATH, or the manner, nature of the contact, which can vary widely.

The description of the class is based on the initial division incorporated in the frame to frame relation hierarchy (Ruppenhofer et al. 2016) of verb lexis in activities (motions) and states (Vendler 1957: 100), Dowty (1979: 40)'s proposal of lexical decomposition, and the idea of complementary notions or semantic attributes in the organisation of meaning (Löbner 2011).

Different verbs within the domain of contact encode more than one semantic attribute, and depending on the focus of classification may be grouped in more than one semantic class. This can be seen when exploring the frame semantic of the semantic classes of verbs of change, verbs of motion, and verbs of communication within WordNet as well as the verbs matching the FrameNet Frames within the semantic domains of Motion, Cause_change, Undergo_change, State, Commutative_process, Noncommutative_process. The lexical semantic domains of WordNet do not translate uniformly into frame semantic domains in FrameNet. Many verbs from the lexical domain of verb.contact are described by FrameNet frames encoding change, motion, and communication, as described in the chapters *Frame Semantics and Verbs of Motion* and *Frame Semantics and Verbs of Change*.

Taking that into consideration, we assume that the contact predicates are divided into two subgroups that combine semantic components of **Physical Contact** and **Motion**, as well as **Physical Contact** and **State**. With a view to the Basic vocabulary domain this work focuses only on verbs denoting direct physical contact, and excludes verbs described by frames Cause_change, Undergo_change, Commutative_process, Noncommutative_process.

The analysed frames are assigned to verbs of contact included in the selection of Bulgarian general lexis verbs so far. Based on the most typical frame elements in the semantic frames –THEME, BODY_PART, SOURCE GOAL, and IMPACTOR IMPACTEE, and the location and manner of contact we consider 3 subclasses of Verbs of physical contact via motion – Verbs of contact, denoting displacement, Verbs of bodily contact and Verbs of contact by impact. The most typical frame elements. They are additionally grouped according to the semantics of the frames that they represent. Some of the frames are discussed below with a view to the assignment of more refined restrictions on frame lemmatization selection.

4.3 Verbs of physical contact via motion

This verb class includes verbs denoting actions where an object or entity makes contact with another object or entity through a specific type of motion. These

5 *Frame semantics and verbs of contact*

verbs emphasise the physical interaction that occurs as a result of movement, and they are assigned one of the following FrameNet frames³: *Becoming_attached*, *Body_movement*, *Breaking_off*, *Cause_fluidic_motion*, *Closure*, *Destructing*, *Detaching*, *Dispersal*, *Filling*, *Fluidic_motion*, *Food_gathering*, *Gathering_up*, *Grinding*, *Make_noise*, *Manipulate_into_shape*, *Placing*, *Removing*, *Reshaping*, *Undressing*, *Processing_materials*.

These frames can be additionally specified based on the semantic relationship between the frame elements and their syntactic realisation. In many cases, part of the synsets sharing the same FrameNet frame belong to the same (or to semantically close) WordNet subtrees. In these cases, the topmost synset more or less complies with the restrictions for the frame, whereas its hyponyms may impose more specific requirements. Based on the motion types combined with contact manner, we subdivide verbs of contact via motion into verbs denoting displacement and verbs of contact by impact. The subclasses that we offer represent the verbs and frames within the data we analysed and do not claim to cover or enlist all possible semantic areas in the class.

4.3.1 *Verbs of contact, denoting displacement*

This subclass includes verbs denoting attaching, detaching, placing, removing, filling, and emptying. They share common frame elements and restrictions based on the point of physical contact. As a whole, these frames involve the movement of an entity the *THEME* either directed to a particular *PLACE* or originating from *SOURCE*. Their core frame elements share similar general restrictions – their *AGENTS* are volitional; the *CAUSE* denotes a physical entity or eventuality; the *ITEM* is a physical object; the *GOAL* – a physical entity or container; and the *CONNECTOR* – a physical entity. The semantics of the point of physical contact defines three main subgroups: verbs of contact on or along a surface, verbs of contact with a container, verbs of contact with a body.

4.3.1.1 *Verbs of contact on or along a surface*

Verbs of contact on or along a surface describe actions where an object or entity comes into contact with, moves across, or follows the contour of a surface. Considering the manner of contact motion, they denote direct physical contact between two objects or entities, such as in *докосвам* (touch), *държа* (hold); slight

³We used the FrameNet data available in September 2023: <https://framenet.icsi.berkeley.edu/>

Maria Todorova

physical contact that involves indirect physical interaction like, such as in *чет-кам* (brush), *ожулвам* (graze); gentle contact like *галия* (caress), *сгушвам се* (nuzzle), *потупвам* (pat); or exploratory contact, such as in *опипвам* (probe), *натискам* (poke).

These verbs often emphasise the physical interaction between the object or entity and a surface. Thus, the most characteristic semantics of frame element of those verbs is the “surface”, represented by SOURCE, GOAL, and CONNECTOR, where the contact occurs, or along which the movement takes place. It is a key component in understanding the spatial aspect of the action described by these verbs. Table 1 represents the frames and their frame elements within this set as well as examples of lexical units. The frame elements of the frames PLACING (Example 1a) and ATTACHING (Example 1b) are illustrated below.

- (1) a. [*Жената*]_{AGE} **ТРИЕ** [*масата*]_{GOAL} [*с гъба*]_{INS}.
 Woman-DEF is rubbing table-DEF with sponge.
 ‘The woman wipes the table with a sponge.’
- b. [*Той*]_{AGE} **ВЪРЗАЛ** [*с връв*]_{CONN} [*разваления зъб*]_{GOAL}.
 He tied with string rotten-DEF tooth.
 ‘He tied the rotten tooth with a string.’
- c. [*Майсторът*]_{AGE} **ЛАКИРА** [*с лак*]_{THM} [*новата маса*]_{GOAL}.
 Craftsman-DEF varnished with varnish new-DEF table.
 ‘The craftsman varnished the new table.’

The FrameNet frames Filling and Emptying describe two semantic situations each. On one hand, they represent the contact of a THEME with a container (see Table 2) and, on the other hand, they represent covering areas with an object, several objects, or a substance, or removing it from the surface. Thus we split the frames Filling and Emptying between the verbs of contact on along surface and verbs of contact with a container. An argument for this is the difference in the core frame elements, representing the surface of the contact – GOAL and CONTAINER. For the frames Filling and Emptying which describe verbs of contact on along surface, the frame element THEME usually expresses that the substance lying on the surface of the GOAL as shown in Example 1c. The THEME imposes additional selectional restrictions on the frame element - direct object in Bulgarian, as shown by the verb eng-30-01269008-v: {лакирам} (varnish, ‘cover with varnish’). Its AGENT is a volitional human being, usually a qualified person, while the THEME is a particular kind of substance, described from the synset eng-30-04521987-n: {лак} (varnish). The GOAL is eng-30-00002684-n: {physical object} or eng-30-08660339-n: {surface}.

5 Frame semantics and verbs of contact

Table 1: Verbs of contact on or along a surface

Frame	Definition	Frame elements	Lexical units
Attaching	An AGENT attaches an ITEM to a GOAL by manipulating a CONNECTOR	AGENT; GOAL; ITEM; CONNECTOR	the verb root eng-30-01354673-v: { <i>срѣзам</i> }, (tie, 'connect, fasten, or put together two or more pieces') ^a and its hyponyms
Detaching	An AGENT detaches an ITEM from a SOURCE	AGENT; SOURCE; ITEM	the verb root eng-30-01298668-v: { <i>махам</i> }, (detach, 'cause to become detached or separated; take off') and its hyponyms.
Placing	An AGENT places a THEME at a location – the GOAL, which is profiled	AGENT; GOAL; THEME	the verb root eng-30-01249724-v: { <i>трия</i> }, (rub, 'move over something with pressure') and its hyponyms – { <i>четкам</i> } (brush); { <i>прекарвам</i> } (gauge); { <i>изтърквам</i> } (scrub); { <i>мажа</i> } (smear); { <i>стържа</i> } (scrape) etc.
Removing	An AGENT causes a THEME to move away from a location, the SOURCE, which is profiled	AGENT; SOURCE; THEME	the verb root eng-30-01532589-v: { <i>изчиствам</i> }, (clean, 'make clean by removing dirt, filth, or unwanted substances from') and its hyponyms – { <i>четкам</i> } (brush); { <i>мета</i> } (sweep); { <i>изпирам</i> } (wash) etc.

^aThe BulNet aligned with the English WordNet and other languages is available online: <http://dcl.bas.bg/bulnet/>

Maria Todorova

4.3.1.2 Verbs of contact with a container

This group of verbs represents the interaction of an object or entity with a container, which may be a box, bag, jar, vessel, or any other object used to hold or store items. Table 2 describes the frames within this set, as well as their frame elements, and examples of lexical units.

Table 2: Verbs of Contact with a Container

Frame	Definition	Frame elements	Lexical units
Filling	Filling CONTAINERS with some thing, things or substance, the THEME.	AGENT; CONTAINER; THEME	the verb root eng-30-01332730-v: {зaпyшaм; зaпyшa} (fill up, ‘fill or stop up’) and its hyponyms
Emptying	An AGENT causes a THEME to move away from a location, the SOURCE, which is profiled	AGENT; SOURCE; THEME	the verb root eng-30-01488313-v: {изпpазвaм; изпpазня} (empty, ‘remove’) and its hyponyms

The most characteristic frame element of those verbs is the “container” or “receptacle”, and specifies the particular container with which the contact is made. The nature and characteristics of the container lead to narrower selectional restrictions of verbs, that elaborate on the more general ones assigned to the frame. The AGENT corresponds to the WordNet root synset eng-30-00007846-n: {*person*}; the THEME corresponds to the WordNet root synset eng-30-00002684-n: {*physical object*} or to eng-30-00001740-n: {*entity*}, and the GOAL matches the synset eng-30-03094503-n: {*container*} A specific valency pattern is represented from the CONTAINER which can be direct, or indirect object, and it is designated GOAL of motion of the THEME which is indirect object.

4.3.2 Verbs of bodily contact

Within this subgroup, verbs describe actions that involve various ways in which objects and body parts come into contact with one another’s surface. This subclass includes the verbs assigned the FrameNet frame Manipulation, which describes *the manipulation of an ENTITY by an AGENT, the ENTITY is not deeply or*

5 Frame semantics and verbs of contact

permanently physically affected, nor is it overall moved from one place to another. The most characteristic frame element of those verbs is the BODY_PART_SURFACE, where the contact occurs or along which the movement takes place. Often it is not expressed explicitly, and is a part of the verb meaning, or the AGENT's manipulation of an ENTITY may be further specified as being localised to some part of the ENTITY, a LOCUS. The part of the Agent's body being used to manipulate the Entity may also be expressed, as shown in Example 2.

Based on the manner of the contact, verbs within this group can denote direct physical contact: *докосвам* (touch), *държа* (hold), *стискам* (squeeze); light and gentle touch: *гала* (caress), *гъделичкам* (tickle); caring or affectionate touch: *прегръщам* (hug), *целувам* (kiss); forceful or aggressive touch: *удрям* (slap), *ципя* (pinch).

- (2) [Майката]_{AGE} **ДЪРЖИ** [лъжицата]_{ENT} [с ръка]_{LOCUS}.
 Mother-DEF holds spoon-DEF with hand.
 'The mother holds the spoon with a hand.'

Selectional restrictions: the AGENT corresponds to the WordNet root synset eng-30-00007846-n: {*person*}; the ENTITY corresponds to the WordNet root synset eng-30-00002684-n: {*physical object*} or eng-30-00001740-n: {*entity*}.

More specific restrictions for the core frame elements of verb synsets assigned the frame MANIPULATION, which are hyponyms of the synset eng-30-01216670-v: {*хващам*} (hold, 'have or hold in one's hands or grip'), are used as shown in Example 3 below.

- (3) a. hyponym: {*стискам*} (grasp, 'hold firmly')
 b. hyponym: {*притискам се*} (clutch, 'hold firmly, usually with one's hands')
 c. hyponym: {*люлея*} (cradle, 'hold gently and carefully')
 d. hyponym: {*склучвам*} (interlace, 'hold in a locking position')
 e. hyponym: {*улавям*} (trap, 'hold or catch as if in a trap')

The restrictions on the AGENT of the root verb and a part of its hyponyms differ: for some verbs the AGENT is a volitional human being, corresponding to the WordNet root synset eng-30-00007846-n: {*person*} (Examples 3b, 3d), while in other cases the verbs may allow their Agent to be an animal, corresponding to the WordNet root synset eng-30-08660339-n: {*animal*} (Examples 3a, 3b), or BODY_PART, corresponding to eng-30-03183080-n: {*body part*}, as in (Example 3e).

The restrictions on the frame element ENTITY are also not consistent in all the discussed members of the tree. ENTITY may be either an animate (Example 3c) or an inanimate physical object (Example 3d).

Maria Todorova

4.3.3 Verbs of contact by impact

Verbs of contact by impact denote a strong or forceful manner of physical contact. They include a wide set of verbs, where the most typical are the verbs represented by the FrameNet frames *Impact* and *Destroying*. The frame *Impact* represents *an IMPACTOR in motion, making sudden, forcible contact with the IMPACTEE, or two IMPACTORS both move, mutually making forcible contact*. The frame *Destroying* represents *a DESTROYER (a conscious entity) or CAUSE (an event, or an entity involved in such an event) affecting a PATIENT negatively so that the PATIENT no longer exists*. The core frame elements in those frames share similar general semantic characteristics, so more specific selectional restrictions can not be defined – the *IMPACTOR* and the *IMPACTEE*, as well as *Destroyer* and *PATIENT* may be physical entities or eventualities, devices, or persons. The broad thematic range of frame elements of those verbs can be illustrated by the verbs belonging to the WordNet subtree stemming from eng-30-01236164-v: {удрям} (hit, ‘hit against; come into sudden contact with’) in Example 4a, 4b, whose semantics is represented by the FrameNet frame *Impact*, and verbs belonging to the WordNet subtree stemming from eng-30-01564144-v: {унищожавам} (destroy, ‘damage irreparably’) (Example 4c, 4d).

- (4) a. hyponym: {сблъскам се} (shock, ‘collide violently’)
- b. hyponym: {разбивам се} (crash, ‘undergo damage or destruction on impact’)
- c. hyponym: {опустошавам} (devastate, ‘cause extensive destruction or ruin utterly’)
- d. hyponym: {унищожавам} (ruin, ‘destroy completely; damage irreparably’)

More rigid selectional restrictions on their frame elements impose some thematic groups within the class of verbs of contact via impact. Such are some of the verbs of digging which describe actions related to excavating or removing material from the ground are represented by the FrameNet frames *Self_motion*, *Planting* and *Mining*. They can denote fundamental excavating or digging into the ground: *копая* (dig.v), *ровя* (burrow.v); gardening: *садя* (plant.v), *разкопавам* (till.v); mining and extraction of valuable resources, or materials: *вадя* (extract.v); burial in graves: *заравам* (bury.v), *законавам* (inter.v), *позребвам* (entomb.v).

5 Frame semantics and verbs of contact

The frame *Self_motion* represents a living being – a *SELF_MOVER*, who moves under its own direction along a *PATH*; the frame *PLANTING* describes an *AGENT* who puts the *THEME* into the *GROUND* for the purpose of growing and the frame *Mining* represents a *MINER* who attempts to obtain a desirable *RESOURCE*, rocks and minerals, located in a *PLACE* being mined, by digging or tunneling in the ground. These conceptual frames share the semantics of forceful contact with the ground, represented by the frame elements *PLACE*, *GROUND*, *PATH*.

4.4 State verbs of physical contact

This verb class includes the verbs assigned the FrameNet frames: {*Locative_relation*, *Being_wet*, *Distributed_position*, *Posture*, *Spatial_contact*, *Surrendering_possession*, *Surrounding*, *Scouring*. These frames describe an *AGENT* (Protagonist), *ITEM*, *THEME*, *FIGURE*, or another entity being on, in or in contact with an area or a substance *LOCATION*. Within the FrameNet frame to frame hierarchy, most of them are subframes of the frame *State*, its subordinate *Locative_relation*, and *Spatial_contact*, described by the definition: *A FIGURE is located in contact with a GROUND or relative to a GROUND location*. Based on the manner of the contact, they denote spatial or location-based contact and describe how objects are situated in relation to each other: *rest*, *place*, *position*, *situate*. Table 3 represents the frames and their frame elements within this set as well as examples of lexical units.

Selectional restrictions: Some verbs impose more specific selectional restrictions on their *AGENT* such as the hyponyms of eng-30-01547001-v: {*лежа*} (*lie*) in Example 5 below.

- (5) a. hyponym: {*нека се*} (sunbathe, ‘expose one’s body to the sun’)
- b. hyponym: {*изтягам се*} (sprawl, ‘sit or lie with one’s limbs spread out’)
- c. hyponym: {*излягам се*} (recumb, ‘lean in a comfortable resting position’)
- d. hyponym: {*покривам*} (overlie, ‘lie upon; lie on top of’)
- e. hyponym: {*почивам*} (repose, ‘lie when dead’)
- f. hyponym: {*нрпичам се*} (bask, ‘be exposed’)

For some of them the Agent may be a volitional human being corresponding to the WordNet root synset eng-30-00007846-n: {*person*} (Example 5e, 5f) as well as an animal (Example 5a, 5b, 5c, 5d), corresponding to the WordNet root synset eng-30-08660339-v: {*animal*}. The frame element *LOCATION* is an adjunct in Bulgarian and can be omitted, and is thus not discussed here.

Maria Todorova

Table 3: State verbs of physical contact

Frame	Definition	Frame elements	Lexical units
Locative_relation	A FIGURE – an entity or event is located relative to a GROUND location.	FIGURE GROUND	eng-30-01466978-v: {границы} (border, ‘lie adjacent to another or share a boundary’)
Posture	An AGENT supports their BODY_PART in a particular LOCATION.	AGENT; LOCATION; BODY_PART	verbs from the WordNet subtree stemming from eng-30-01547001-v: {лежа} (lie, ‘be lying, be prostrate; be in a horizontal position’)

5 Valency patterns

The observations on the syntactic behaviour of the studied verbs led to the delineation of several general syntactic constructions within the group:

- (a) *NP (pro-drop subject) Verb NP (direct object – THEME/ITEM) PP (non-obligatory indirect object – to/on/over NP GOAL/SOURCE).*

This syntactic structure is typical for verbs selecting a Theme as an object within the frame Emptying, for instance *разля* (pour) – *Разля чая по масата* (‘She spilled the tea over the table’).

- (b) *NP (pro-drop subject) Verb NP (direct object – Goal) PP (non-obligatory indirect object – with NP THEME).*

This pattern is found with verbs taking the frame element GOAL as an object within the frame Filling, for instance *намажа* (spread) – *Намаза филията с масло* (‘She spread the slice with butter’).

- (c) *NP (pro-drop subject) Verb NP (direct object – CONTAINER) PP (non-obligatory indirect object – with NP THEME).*

This type of structure is typical for verbs selecting the frame element CONTAINER as an object within the frame Filling, such as *натоваря* (load) – *Натовариха камиона с тухлите* (‘They loaded the truck with the bricks’).

6 Discussion

The process of annotation raised some interesting questions regarding the language-specific lexicalisation patterns of some Bulgarian verbs as compared with their English counterparts. This led to the conclusion that different word formation mechanisms across Bulgarian and English, such as derivation, compounding, and conversion, as well as lexical gaps, reflect differences in the semantic structure of lexemes.

The syntactic expression of some of the frame elements differs in the two languages. The obligatoriness of the syntactic realisation depends on the point of contact between the core frame elements. The English verbs of contact that encode one of the frame elements in their morphological structure – e.g. the instrument (knife), the resultant shape (slice), the covering material (paint), the container (box, bag), etc. – have a different lexical expression in Bulgarian. Not all the Bulgarian correspondences have the frame element incorporated in their word structure. For example, the English verb *eng-30-01364483-v: cream*, ‘put on cream, as on one’s face or body’ – has no one-word correspondence in Bulgarian and is translated as the expression *намазвам с крем* (cover with cream), where *крем* (cream) is the THEME, compare Example 6:

- (6) [Тя]_{AGE} **НАМАЗА** [лицето си]_{GOAL} [с крем]_{Thm}.
 She covered face-DEF REFL with cream.
 ‘She creamed her face.’

On the other hand, some of the Bulgarian verb hyponyms express a specific manner by means of prefixation, for example *разрязвам* (*eng-30-01552519-v: {cut}*, ‘cut into pieces’). Such predicates lexicalise a meaning component which specifies a scale of motion or state and contact, and they do not have full one-word correspondences in English. These and other similar cases have necessitated the modification of FrameNet frames or the definition of further specifications.

7 Conclusions and future work

The research described in this paper offers a semantic description of some verbs from WordNet with the semantic primitive ‘contact verbs’, which belong to the set of common vocabulary in Bulgarian. Only the most frequent cases with larger coverage in BulNet were commented on. With regard to the information obtained for both the semantic hierarchy from BulNet and WordNet, and the conceptual

Maria Todorova

frames that describe them from FrameNet, the class of verbs of contact is grouped into two major classes of predicates: Verbs of physical contact via Motion and Verbs of physical contact in state. The selection of the data and its description is based on the assumption that the conceptual semantic of verbs of contact involves actions or events where two or more entities come into physical contact with each other. This verb set encompasses a wide range of verbs that describe various aspects of contact between subjects, objects, entities or areas. The core elements of this conceptual frame include entities in contact – the one initiating the contact and the one being contacted; Directionality of contact – one entity reaching out to touch or affect another entity; Manner of contact – direct, slight, forceful, gentle, exploratory. Semantic subgroups based on the frame semantics are offered within the Verbs of Physical Contact via Motion and Verbs of Physical Contact in State, along with the description of the syntactic properties and the definition of more specific selectional restrictions for each verb. The work makes conclusions on the internal semantic organisation of the verbs within the domain of verbs with the semantic attribute Physical contact. The analysis contributes to constructing a theoretically and empirically coherent approach to usage data and investigating its specificities.

In the future a more detailed analysis on certain subclasses within the set of verbs of contact is needed, with focus on their syntactic realisation and alternations, which will lead to the enrichment of both WordNet and FrameNet.

As the proposed analysis is based on multilingual resources such as WordNet and FrameNet, some of the observations may also be useful for other languages and may contribute to the implementation of NLP applications aimed at automatic semantic analysis, word sense disambiguation, language understanding and generation, machine translation, etc.

Abbreviations

Age	AGENT	NP	Noun phrase
Conn	CONNECTOR	PP	Prepositional phrase
Ent	ENTITY	Thm	THEME
Ins	INSTRUMENT	VP	Verb phrase

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

References

- Baker, Collin F. & Christiane Fellbaum. 2009. WordNet and FrameNet as complementary resources for annotation. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, 125–129. DOI: 10.3115/1698381.1698402.
- Brysbaert, Marc & Andrew Biemiller. 2017. Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods* 49. 1520–1523. DOI: 10.3758/s13428-016-0811-4.
- Dimitrova, Tsvetana & Valentina Stefanova. 2019. On hidden semantic relations between nouns in WordNet. In *Proceedings of the Tenth Global WordNet Conference*, 54–63.
- Dimitrova, Tsvetana, Maria Todorova & Valentina Stefanova. 2022. Research on the basic verbal vocabulary in Bulgarian for students in the initial stage of education through online games. *Pedagogika-Pedagogy* XCIV. 896–913.
- Dowty, David. 1979. The semantics of aspectual classes of verbs in English. In *Word meaning and montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*, 37–132. Dordrecht: D. Reidel. DOI: 10.1007/978-94-009-9473-7_2.
- Dzhonova, Marina. 2028. Bulgarica – studia et argumenta. Jubilee Collection in Honor of the 65th Anniversary of prof. Ruselina Nitsolova. r Bulgarische bibliotek. In vol. 13, chap. Argument structure of sentences containing predicates taking an indirect object Experiencer of the type ‘spi mi se’ (in Bulgarian), 231–240. Münche: O. Sagner.
- Fellbaum, Christiane. 1990. English verbs as a semantic net. *International Journal of Lexicography* 3. 278–301. DOI: 10.1093/ijl/3.4.278.
- Fellbaum, Christiane. 1999. The organization of verbs and verb concepts in a semantic net. *Text, Speech and Language Technology* 6. 278–301.
- Fillmore, Charles J. 1970. The grammar of hitting and breaking. In Roderick Jacobs and Peter Rosenbaum (ed.), *Readings in English Transformational Grammar*, 120–133. Waltham, MA: Ginn.
- Fillmore, Charles J. 1977. Grammatical relations. In Peter Cole & Jerrold M. Sadock (eds.) (Syntax and Semantics 8), chap. The Case for Case Reopened, 59–81. Brill. DOI: 10.1163/9789004368866_005.
- Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Publishing.
- Fillmore, Charles J., Collin F. Baker & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the Conference COLINGACL '98, Montreal, Canada*, 86–90.

Maria Todorova

- Gao, Hong. 2001. Notions of Motion and contact for physical contact verbs. In A. Holmer, J. Svantesson & Åke Viberg (eds.), *Proceedings of the 18th Scandinavian Conference of Linguistics*, vol. 2, 193–209. Lund: University of Lund.
- Gao, Hong & Chin-Chuan Cheng. 2003. Verbs of contact by impact in English and their equivalents in Mandarin Chinese. *Language and Linguistics* 4.3. 485–508.
- Juffs, Alan. 1996. *Learnability and the lexicon. Theories and second language acquisition research*. Amsterdam: John Benjamins.
- Kipper-Schuler, Karin. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania. (Doctoral dissertation).
- Koeva, Svetla. 2004. Syntactic and Semantic Description of Bulgarian Diatheses (in Bulgarian). In *Bulgarian Linguistics. Cognitive Grammar of Bulgarian and French – Description and Formalisation*, 182–231. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Koeva, Svetla. 2010. *Bulgarian FrameNet (in Bulgarian)*. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2021a. Stative predicates: Properties and linguistic tests (in Bulgarian). *Zeszyty Cyrylo-Metodiańskie* 10. 106–138.
- Koeva, Svetla. 2021b. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla. 2021c. Towards expanding WordNet with conceptual frames. In *Proceedings of the 11th Global Wordnet Conference*, 182–191.
- Koeva, Svetla. 2022. The system of diatheses in Bulgarian and stative predicates. In Svetla Koeva, Elena Ivanova, Yovka Tisheva & Anton Zimmerling (eds.), *Ontology of stative situations – linguistic modeling. A contrastive Bulgarian-Russian study*, 117–160. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Koeva, Svetla & Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the fifth international conference computational linguistics in bulgaria*, 203–214. <https://aclanthology.org/2022.clib-1.23>.
- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Tsvetana Dimitrova, Rositsa Dekova & Ekaterina Tarpomanova. 2012. The Bulgarian national corpus: Theory and practice in corpus design. *Journal of Language Modelling* 0. 65–110. DOI: 10.15398/jlm.v0i1.33.
- Kostova, Nadezhda. 2010. *Basic motion verbs in Bulgarian*. Sofia: Avangard Prima.
- Kukova, Hristina. 2020. Verbs for communication, frame elements and semantic restrictions (on BulNet synsets). *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2020)* 2. 233–241.

5 Frame semantics and verbs of contact

- Leseva, Svetlozara, Ivelina Stoyanova, Hristina Kukova & Maria Todorova. 2018. Integrating subcategorization information in WordNet's relational structure. *Bulgarian Language* 2. 13–40.
- Leseva, Svetlozara, Ivelina Stoyanova, Msria Todorova & Hristina Kukova. 2019. A theoretical overview of conceptual frames and semantic restrictions on frame elements. *Linguistique Balkanique* LVIII(2). 172–186.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Levin, Beth. 2015. Verb classes within and across languages. In Bernard Comrie & Andrej Malchukov (eds.), *Valency classes in the world's languages*, 1627–1670. Berlin: De Gruyter.
- Löbner, Sebastian. 2011. Concept types and determination. *Journal of Semantics* 28. 279–333.
- Miller, George A. 1990. Nouns in WordNet: A lexical inheritance system. *International Journal of Lexicography* 3(4). 245–264.
- Miller, George A. 1995. WordNet: A lexical database for English. *Commun. ACM* 38(11). 39–41.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin. F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.
- Stamenov, Maksim. 2021. Groups of verbs for internal psychological experience. An attempt at classification (in Bulgarian). In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2021)*, vol. 2, 69–75. Professor Marin Drinov Publishing House of BAS.
- Stamenov, Maksim. 2022. Single-component verbal sentences, impersonal verbs, predicatives and subjective forms in Bulgarian (in Bulgarian). *Bulgarian Language Addendum*. 158–178.
- Stoyanova, Ivelina & Svetlozara Leseva. 2020. Towards a Semantic Network Enriched with a Variety of Semantic Relations. In Svetla Koeva (ed.), chap. Beyond Lexical and Semantic Resources: Linking WordNet with FrameNet and Enhancing Synsets with Conceptual Frames, 21–48. Sofia: Professor Marin Drinov Publishing House of BAS.
- Stoyanova, Ivelina & Svetlozara Leseva. 2021. Semantic description of verbs for change and hierarchical organization of conceptual frames. In *Proceedings of the Annual International Conference of the Institute of Bulgarian Language Prof. Lubomir Andreychin (Sofia, 2021)*, vol. 2, 76–85.
- Svetlozara, Leseva, Ivelina Stoyanova, Maria Todorova & Hristina Kukova. 2020. A semamtic description of the combinability between verbs and nouns (on material from Bulgarian and English). *Chujdoezikovo obuchenie* 47. 115–128.

Maria Todorova

- Tisheva, Yovka. 2021. Observations on the syntactic realisation of predicates for emotional states (in Bulgarian). *Zeszyty Cyrylo-Methodiańskie* 10(1). 114–136.
- Vendler, Zeno. 1957. Verbs and times. *Philosophical Review* 66(2). 143–160.
- Viberg, Åke. 2004. Physical contact verbs in English and Swedish from the perspective of crosslinguistic lexicology. In Karin Aijmer & Bengt Altenberg (eds.), *Advances in corpus linguistics: Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)* (Language and Computers 49), 327–352. Brill. DOI: 10.1163/9789004333710_020.
- Vulchanova, Mila & Rositsa Dekova. 2009. On the Encoding of lexical information: Events and their lexicalization in English and Bulgarian. *Bulgarian Language* LVI. 84–96.

Chapter 6

Frame semantics and verbs of emotion

Hristina Kukova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

The intersection of lexical semantics and syntax has been an important area of linguistics for some time now. Verbs, as the core of the lexicon, provide the key to exploring the interaction between syntax and semantics as well as to understanding the nature of the lexicon. The focus of the research is on verbs of emotion in the Bulgarian language and their frame semantics. An overview of both theoretical and empirical observations comprises the general aim of the study. Neutral, positive and negative verbs of emotion are discussed and the results are summarised. The analysis is based on lexical units' semantic and partly morphological information extracted from WordNet (Fellbaum 1999b) as well as the semantic and syntactic specificities with which the emotion verbs under study are represented in FrameNet (Fillmore et al. 1998, Ruppenhofer et al. 2016). Five semantic frames are documented and they were selected on the basis of their high frequency and the wide variety of lexical units that evoke them. The description includes grammatical features of the lexical units, semantic and syntactic restrictions that verbs impose on frame elements, mapping of frame elements to a Wordnet noun synset or subtree, reflecting the realization of the frame elements in context. The status of the frame elements, which is essential for the realization of a lexical unit, is retrieved from FrameNet.

1 Introduction

The aim of the study is to present emotion verbs belonging to the Bulgarian common vocabulary set. We apply the methodology of Frame semantics in order to outline various constructions, in which verbs of emotion participate. We also use the BulNet semantic network to extract their distinctive meanings. Thus, we



Hristina Kukova

present the verbs under study within specific WordNet synsets, which include lexical and morphological information. We then describe each predicate within the semantic frame it evokes together with its frame elements (FEs) and their selectional restrictions expressed in terms of specific WordNet synsets or subtrees. We assume that the lower levels (hyponyms) of the selected subtree can also occupy the FE position.

In this research we rely on the usage-based approach and provide evidence for the importance of context in the semantic analysis and frame profiling. The analysis of the corpus data contributes to constructing a theoretically and empirically coherent approach to describing semantic and syntactic specificities of verb classes.

The key objectives of this study are: (i) systematisation of main theoretical findings on emotion verbs; (ii) analysis of the semantic frames and their frame elements; (iii) demonstration of how syntactic realisations can be predicted by lexical semantics within a given verb class; (iv) highlighting the importance of the interaction between semantics (lexical-semantic properties) and syntax (syntactic behaviour).

We build on what is known from Levin's study (Levin 1993) who classifies verbs of psychological state into 4 major subclasses on the basis of both intuitive semantic grouping and their participation in valence alternations. We consider the transitivity / intransitivity *желая* (*wish.v*) / *страхувам се* (*fear.v*) of the verbs and the possibility of taking the EXPERIENCER as a grammatical subject – *обичам* (*love.v*), or object – *харесва ми* (*appeal to.v*) in a sentence to further subdivide them. This division is reflected in a verb's evoking *Experiencer_focused_emotion* – *завиждам* (*envy.v*), or *Stimulate_emotion* and *Cause_to_experience* semantic frames – *изненадвам* (*surprise.v*), *дразня* (*annoy.v*).

If we take *съжалявам* (*regret.v*) as an example of a verb that evokes the *Contrition* frame, we can see that in most cases the position of the EXPERIENCER is occupied in context by the synsets *eng-30-00007846-n: {човек} (person.n)* or *eng-30-07950920-n: {социална група} (social group.n)*. The ACTION FE “marks expressions that indicate a prior action that the EXPERIENCER has come to feel bad about” and can be encoded both as a PP (with the prepositions *за* (*for.prep*) and *заради* (*because of.prep*)) or as a clause (with the help of the conjunctions *че* (*that.conj*), *защото* (*for.conj*)) and the interrogative pronouns *как* (*how.adv*), *къде* (*where.adv*), *какво* (*what.adv*), *кой* (*who.adv*), *колко* (*how much/many.adv*). The prepositions *за* (*for.prep*) and *заради* (*because of.prep*) in turn take an object, which can vary between the following synsets: *eng-30-00029378-n: {събитие} (event.n)*, *eng-30-00030358-n: {действие} (act.n)*, *eng-30-00037396-n: {действие} (action.n)*, or *eng-30-05770926-n: {умствена дейност} (thought process.n)*. We

6 *Frame semantics and verbs of emotion*

take into consideration the possible selectional restrictions a verb imposes on its frame elements and group verbs further into subclasses.

The study is based on corpus data; unless otherwise indicated, examples are taken from the Bulgarian National Corpus (Koeva et al. 2012).

The remainder of the paper is organized as follows. Section 2 deals with the notion of a Conceptual frame and its prerequisites. A detailed explanation is provided on the resources that were used. In Section 3 previous studies and motivation are outlined. It focuses on the description of the class of verbs of emotion and on the various criteria for classification. In the same section and throughout the paper differences in classification systems serve as a basis for distinction between subclasses of verbs of emotion. Section 4 provides an overview of linguistic descriptions of Bulgarian verbs of emotion and their specifics. Section 5 focuses on the semantic features of verbs of emotion in Bulgarian. It encompasses descriptions of different semantic frames and their frame elements. Section 6 summarises the results of this study and concludes the paper.

2 Resources

WordNet is a lexical-semantic network suitable for machine processing, originally developed at the Princeton University by a team led by George Miller (Miller 1995, Fellbaum 1999b). The Bulgarian version of WordNet – BulNet – contains over 100,000 synsets (Koeva 2021a).

While BulNet was used to present the semantic and paradigmatic features of predicates, the main resource used for their “semantic and syntactic combinatory possibilities” is FrameNet (Ruppenhofer et al. 2016: 7). FrameNet was started under the guidance of Charles Fillmore in 1997 (Fillmore et al. 1998) and is essential for both theoretical linguistic research and practical natural language processing.

Semantic frames represent the conceptual structure of an event or an object and its participants. Frame elements can be seen as semantic roles. They can be core and non-core, the former being essential for the realisation of the particular semantic frame, while the latter are mostly descriptive (concerning time, place, etc.). Lexical units are lemmas that give rise to a situation (frame). Each sense of the same word is encoded as a separate lexical unit and a different semantic frame is evoked by it.

As Koeva and Doychev state, “a Conceptual frame defines a unique set of syntagmatic relations between verb synsets representing the frame and noun synsets expressing the frame elements” (Koeva & Doychev 2022: 203). Building on the information we extract about the verb sense from the WordNet and the syntactic

Hristina Kukova

and semantic restrictions it imposes on its FrameNet frame elements, we generate a grid of possible combinations. All verbs under study are taken in each particular sense separately and their frame elements can be an NP, PP, AdvP, AccCl (obligatory accusative clitic), DatCl (obligatory dative clitic) or a clausal element (S or small clause). We use the BulFrame web-based system to create and visualise conceptual frames (Koeva & Doychev 2022).

With a view to the exhaustive description of the Bulgarian verbal lexical units the following information has been used:

- (a) From FrameNet: core and non-core frame elements from the semantic frames, their semantic types, the sets of verbal lexical units, connected with a specific semantic frame and the Inheritance relation between frames.
- (b) From WordNet: hypernym-hyponym relations organising noun and verb synsets into hypernym trees and the semantic classes to which noun and verb synsets belong.

The BulFrame web-based system, developed at the Department of Computational Linguistics at the Bulgarian Academy of Sciences, has been used to create, edit, view and verify the conceptual frames (Koeva & Doychev 2022).

Most of the language material was excerpted from the Bulgarian National Corpus (Koeva et al. 2012), created at the Institute for Bulgarian Language “Prof. Lyubomir Andreychin”. The Bulgarian National Corpus consists of a monolingual part, which contains 240,000 texts or 1.2 billion words, and 47 parallel corpora.

3 Previous studies and motivation

3.1 Methodology

S. Koeva notes the necessity of a formal description of syntagmatic relations in WordNet (Fellbaum 1999b, Koeva 2021a) by introducing the notion of the conceptual frame to define a set of verbs exhibiting unambiguous syntagmatic relations with nouns (Koeva 2021c: 182). Leseva et al. have also studied the possibility to integrate data from WordNet, FrameNet and VerbNet, proposing thus “a system of semantic relations reflecting thematic relations between predicates and their potential arguments within the framework of WordNet” (Leseva et al. 2018). Our approach, based on Frame semantics (Fillmore et al. 1998, Koeva 2010), combines both the abstract syntactic layer and the projection of semantic and pragmatic relations on the possible Frame elements. In the context of their specifics, emotion verbs provide a very interesting field for multi-disciplinary research.

6 *Frame semantics and verbs of emotion*

As frame analysis is very sensitive and prone to mistakes, the decision making is delegated to human experts. To facilitate the process, we have adopted a procedure, which we describe below.

Step 1: We choose the relevant verb sense (literal), which evokes a specific frame, out of a number of synsets.

Step 2: We check whether all the core frame elements are relevant for Bulgarian and/or whether additional one should be included. We can either choose from the existing FEs where appropriate or insert a completely new one, giving it a name.

Step 3: We define the possible selectional restrictions by (a) choosing from a list of noun synsets for NPs; (b) indicating the prepositions for PPs; (c) specifying the conjunctions that can introduce the dependent clauses.

The frame-semantic analysis with its flexibility and versatility can contribute to a number of NLP tasks and applications and to improving language understanding.

- (i) Frame semantics provides a framework for Semantic Role Labelling, i.e. identifying and labelling the different roles played by entities in a sentence. This process is crucial for tasks like question-answering, information extraction, and machine comprehension.
- (ii) Sentiment Analysis. Frame semantics aids in a deeper understanding and analyzing the underlying emotions and attitudes expressed in a text. By capturing the semantic frames associated with sentiment, sentiment analysis models can understand the implicit information in a more nuanced manner.
- (iii) Text Classification. By considering semantic frames and their associated meanings, models can identify the implicit information in context and capture the intended meaning of a text, leading to more accurate and nuanced text classification.
- (iv) Machine Translation. Frame semantics assists in transferring meaning from one language to another by capturing the semantic frames and their semantic roles. This approach goes beyond word-to-word translation and ensures that the intended meaning of the source sentence is preserved in the target language, generating more accurate translations. There have

Hristina Kukova

also been proposed metrics for machine translation evaluation based on frame semantics (e.g. Czulo et al. 2019).

- (v) Information Retrieval and Question-Answering. Frame semantics aids in improving search engine results and question-answering systems. By understanding the frames and semantic roles in queries and documents, these systems can retrieve more relevant information and provide accurate answers by matching semantic frames and roles.
- (vi) Knowledge Graph Construction. Frame semantics is useful in constructing knowledge graphs by identifying the relationships between entities based on semantic frames and their FEs. It helps in organizing and representing the structured knowledge from an unstructured text, contributing to tasks like knowledge extraction and knowledge representation.

Frame semantics plays a crucial role in corpus studies as well. It provides a framework for analyzing and understanding the meaning and structure of texts within a given corpus. It can influence corpus studies in various ways:

- (vii) Semantic Analysis. By identifying and labelling semantic frames and their FEs, corpus studies can uncover patterns and relationships between entities, actions, and events, leading to a deeper understanding of the underlying semantics within the corpus.
- (viii) Semantic Annotation. Frame semantics provides a systematic approach for annotating corpora with semantic information. Corpus studies can leverage frame-based annotation schemes to mark up frames and their FEs in texts, thereby enabling more detailed analysis and, in turn, facilitating the development of machine learning models for various NLP tasks.
- (ix) Comparative Studies. Frame semantics enables comparative studies of different corpora or subsets within a corpus. Researchers can explore variations in frame usage across different genres, time periods, or languages, shedding light on how meaning and semantic structure vary in different contexts. This helps in studying linguistic and cultural differences, diachronic changes, or genre-specific semantic patterns.
- (x) Semantic Similarity and Clustering. By applying frame semantics to corpus studies, researchers can measure the semantic similarity and cluster texts based on their frame-based representations. This facilitates tasks like document clustering, topic modelling, and information retrieval, where a better understanding of semantic relationships between texts is essential.

6 *Frame semantics and verbs of emotion*

- (xi) **Corpus-based Lexical Semantics.** Frame semantics aids in corpus-based studies of lexical semantics. By analyzing lexical items in the context of semantic frames and their FEs, corpus studies can reveal the nuances and contextual meanings associated with words, leading to the identification of polysemy, homonymy, and semantic shifts within the corpus.
- (xii) **Corpus-based Frame Compilation.** Corpus studies contribute to the compilation of frame databases or resources. By analyzing large corpora, researchers can identify recurring semantic frames, frame-evoking lexical units, and their roles, which serve as valuable data for building or expanding frame resources.

Overall, frame semantics provides a rich representation of the meaning and structure of language, allowing NLP models to have a deeper understanding of texts and perform a wide range of tasks more effectively. It also provides a theoretical and practical foundation for corpus studies, enabling researchers to delve deeper into the semantics of texts, compare different corpora, uncover patterns, and advance our understanding of language structure and meaning within a given corpus.

The aim of this paper is, therefore, to demonstrate that formal linguistic facts related to the emotion verbs' complementation in Bulgarian are actually clear manifestations of the semantics of the predicates, and their linguistic categorization.

3.2 **Verbs of emotion**

Emotions can be defined as experiences or states triggered off by a certain event, situation, action, other people, our thoughts, expectations and plans (Belaj & Faletar 2011: 155). In view of this phenomenon we attempt to relate the complexity of the syntax of emotions to the variety of their semantics as demonstrated in Section 1.

In one recently published psychological encyclopedic manual (Strickland 2000: 218) emotions are defined as “a reaction, both psychological and physical, subjectively experienced as strong feelings, many of which prepare the body for immediate action. In contrast to moods, which are generally longer lasting, emotions are transitory, with relatively well-defined beginnings and endings. They also have valence, meaning that they are either positive or negative. Subjectively, emotions are experienced as passive phenomena. Even though it is possible to exert a measure of control over one's emotions, they are not initiated – they happen to people.”

Hristina Kukova

As far as the linguistic field is concerned, there have been published a number of studies dealing with the description of emotion words, starting with Wierzbicka (Wierzbicka 1971, Wierzbicka 1972: 57, Wierzbicka 1980: 142, Wierzbicka 1986 and Iordanskaja 1970, Iordanskaja 1973, Iordanskaja 1986). Wierzbicka was the first to observe that unlike other language groups, Slavic languages tend to use verbs to speak of emotions, which holds true for the Bulgarian language as well. Her early works include attempts to formalize emotions, defining emotion words in natural language and referring to typical situations that evoke particular emotional states. Both Wierzbicka and Iordanskaja put forward the concept of evaluation of the situation by X for the description of emotion words in linguistic semantics. Anna Zaliznjak (Zaliznjak 1983, Zaliznjak 1985) deals with what she calls “predicates of internal state”, establishing the distinction of the assertion and presupposition in their definitions. Lakoff (Lakoff 1987) and Kövecses (Kövecses 1988) pay attention to the uses of emotion expressions and metaphors in a given language, in order to describe a conceptual model of the corresponding emotion – as it is perceived and expressed in actual speech.

Another widely disputed issue throughout the studies of verbs of emotion and specifically among Slavic authors is the verbs’ reflexivity or mediality. The most influential account of the Slavic verbs under discussion is offered by A. Wierzbicka (Wierzbicka 1988, Wierzbicka 1995). The author states that these verbs in Russian and Polish with *-sja* and *-się* respectively are reflexive forms on the basis that they indicate “emotions to which people ‘give themselves’ almost voluntarily and which they outwardly express” (Wierzbicka 1988: 253). As the author claims, expressing emotions by reflexive verbs implies that they are “treated not as arising by themselves but by the speaker’s conscious thoughts about the event” (Wierzbicka 1995: 22). Moreover, she outlines the syntactic distinction between voluntary (with EXPERIENCER in nominative and the *-sja* verb), involuntary (with dative EXPERIENCER and an adverbial predicative) and neutral emotion (with nominative EXPERIENCER and an adjectival predicative) (Wierzbicka 1988: 253–254). A. Bedkowska-Kopczyk challenges Wierzbicka’s views and considers this type of verbs in Slovene middle voice verbs (Bedkowska-Kopczyk 2014).

As far as Bulgarian language is concerned, the *ce* particle can be involved in rather complex relations between words and constructions. It can represent both a word-forming and a morphological marker and can bear a passive (Example 1a), a medial (Example 1b) or a reflexive (Example 1c) meaning (Tisheva 2022: 100–103).

6 Frame semantics and verbs of emotion

- (1) a. *Пациентът не трябва да **СЕ БЕЗПОКОИ** (от никого).*
 Patient-DEF not should to REFL disturb-3.SG.PRS (by nobody).
 ‘The patient should not be disturbed (by anybody).’
- b. *Пациентът не трябва да **СЕ БЕЗПОКОИ**.*
 Patient-DEF not should to REFL disturb-3.SG.PRS.
 (да изпитва безпокойство)
 (to experience worry)
 ‘The patient should not worry. (experience worries)’
- c. *Синът ми вече **СЕ МИЕ** сам.*
 Son-DEF my already REFL wash-3.SG.PRS alone-MASC.SG.
 ‘My son can already wash himself on his own.’

Tisheva & Dzhonova (2022: 76) also address this polemical issue in their research on syntactic characteristics of emotion predicates ¹ in Bulgarian. According to the authors “*se* is a marker for middle voice construction and does not indicate reflexivity, it occupies the direct object position and those verbs could have only PP or a complement clause as their second argument.”

Since the current study focuses on the semantic and syntactic features of the verbs under discussion together with their possible complements as imposed by the verb sense, we will not deal with this particular aspect of the verb description. As in most cases the verbs used with and without the reflexive *ce* involve literals from different WordNet synsets, they will have different meaning and, respectively, heterogeneous frame elements’ restrictions.

A large number of studies have been carried out on different language material in the last 20 years involving the emotion verbs, their organisation in FrameNet and their semantic specifications.

Taking the emotion concepts as a basis, Ruppenhofer describes the evolution and the development of FrameNet analyses over time due to application-oriented goals. Taking into account different linguistic theories and approaches (dimensional, categorical, meaning-oriented, etc.), the author illustrates how fine-grained distinctions of lexical units lead to formulating new semantic frames or dividing one frame into two (Ruppenhofer 2018). The explanation of the steps and motivation underlying the conceptualisation and the development of the frame organisation holds a specific value for emotion frame descriptions and their detailed understanding. Thus, the *Experiencer_Subj* and *Experiencer_Obj* verbs

¹The authors explore both verbs and other constructions, based on adjectives, adverbs or nouns (predicatives).

Hristina Kukova

were initially grouped by valence criteria whereas in the latest version the semantics of the verbs is also considered.

Subirats and Petruck analyse Spanish predicates, describing events of surprise. They compare the Spanish lexical units to those in English in order to outline the similarities and differences in the lexicalisation patterns of the two languages. They use annotation of Spanish verbs with the help of the FrameNet frames to summarise the different syntactic realizations (Subirats & Petruck 2003). As the Bulgarian grammatical and syntactic realization bears a lot more similarities to the Spanish than to the English one², it was especially useful to get acquainted with their experience.

Subirats continued his work in the field and together with Sato (Subirats & Sato 2004) presented the Spanish FrameNet and the web-application for using it FrameSQL, which handles two-language information and makes it possible to compare semantic structures of two lexicons more easily.

Ghazi et al. make an attempt to automatically detect the emotion STIMULUS. They compile a dataset with manually labeled emotion stimuli and then apply sequential learning methods to a complementary dataset with no labels of stimuli but containing such (Ghazi et al. 2015).

All these studies lay the groundwork for our research and have influenced the observations made in the central part of this paper, where we will explore the nature of categorization of emotions and the way it is formally reflected in grammatical and semantic structure, in particular in the restrictions on emotion verb complements.

3.3 Classifications

The typological description of emotion verbs has also proved a point of interest for various authors in different studies. In this section we will give a brief overview of their approaches.

Starting with emotion words in general, Kövecses proposes a division into expressive and descriptive emotion words (Kövecses 2003), including emotionally charged comments and statements of agreement and disagreement in the former group, whereas in the latter he poses the terms which name specific emotional experience. Tisheva also distinguishes between the lexical and grammatical means for emotional attitude of the speaker / author on the one hand and the naming

²It shows closeness in agreement, question formation, negation, usage of prepositions, and above all word order.

6 *Frame semantics and verbs of emotion*

of emotional states, relations or evaluation, on the other (Tisheva 2021: 115). Having in mind their division, we will deal further only with the descriptive emotion words.

Tisheva maintains that duration is a basic concept to draw the line between emotions and feelings. According to the author “while emotions are spontaneous responses to specific inner or outer stimuli, feelings bear greater resilience and durability and always include evaluation of the object they are targeted at” (Tisheva 2021: 115).

Rozwadowska & Willim (2016) generalise that “there is general agreement in the philosophical, psychological and linguistic literature that emotions are internal mental states which arise from evaluative cognitions (cf., among others, Dowty 1991; Jackendoff 2009; Johnson-Laird & Oatley 1989; Lewis et al. 2010; Ortony et al. 1988). The evaluations can concern causes, objects or goals, among others”.

Most of the linguistic classifications take as their basis the above stated psychological aspects of emotions, dividing them thus into positive and negative ones according to their basic tone. Scherer (Scherer 2005) captures three distinctive features of emotions, namely: intensity, duration and the capacity to encourage reaction and creates a typology of affective phenomena as given below:

- (a) emotion: a relatively brief response to (the evaluation of an external or internal) STIMULUS event, e.g. *angry, sad, joyful, fearful, ashamed, desperate*,
- (b) mood: a diffuse affect state pronounced as of low intensity but relatively long duration, often without apparent cause, e.g. *cheerful, gloomy, depressed*,
- (c) interpersonal stance: affective stance taken toward another person in a specific interaction, e.g. *distant, warm, supportive, contemptuous*,
- (d) attitude: relatively enduring, effectively colored beliefs, preferences, and predispositions towards objects or persons, e.g. *liking, hating, desiring*,
- (e) personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person, e.g. *nervous, hostile, jealous, envious*.

Lyashevskaya and Paducheva, on the other hand, classify verbs of emotion on the basis of their semantic structure and the consistency of verbal operational functors included in each meaning. Thus they divide the verbs under question

Hristina Kukova

into five distinct groups: event, feeling, attitude, state, and feature (Lyashevskaya & Paducheva 2011).

In the current research, we are not going to focus so much on the semantic differentiation, but rather on the syntactic realization of the verbs and their FE's semantic specificity, which plays a crucial role within the frame-semantic analysis.

Considering the grammatical features of the verbs, Johnson-Laird & Oatley (Johnson-Laird & Oatley 1989) refer to two types of emotion verbs (they speak also of emotion nouns and adjectives): **emotional relations**, e.g., *to love*, *to fear*, and **causatives**, e.g., *to annoy*, *to frighten*. Their observation correlates with the two types often described in the linguistic literature, comprising SE verbs and OE verbs respectively. Syntactic structures in which the EXPERIENCER is the subject encode the verb class of emotional relation, while the structures in which the EXPERIENCER is encoded as the grammatical object name the verb class of causatives. The former are cross-linguistically referred to as the **Subject-EXPERIENCER** verbs (SE), whereas the latter are known as **Object-EXPERIENCER** verbs (OE) arising from the theories of argument structure (Dowty 1991; Levin & Rappaport Hovav 2005). Fellbaum joins this direction of linguistic description, saying that emotion predicates “fall into two grammatically distinct classes: those whose subject is the animate EXPERIENCER and whose object (if there is one) is the SOURCE (*fear*, *miss*, *adore*, *love*, *despise*); and those whose object is the animate EXPERIENCER and whose subject is the SOURCE (*amuse*, *charm*, *encourage*, *anger*)” (Fellbaum 1999a: 297).

The main division in Slavic languages follows the definition of the two groups of emotion verbs on the basis of the syntactic expression of the EXPERIENCER participant as a subject or a direct or indirect object (Croft 1993: 55, Ovsjannikova 2013: 21, Tisheva & Dzhonova 2022: 75).

On the basis of this observation, three main subtypes are generally distinguished for Slavic languages: (i) SE verbs (Example 2a), (ii) OE verbs with the EXPERIENCER in the accusative case (Example 2b), and (iii) OE verbs with the EXPERIENCER in the dative case (Example 2c). This fact has been maintained by a number of Slavic linguists: for Russian – Sonnenhauser (2010), for Polish – Biały (2005) and Rozwadowska et al. (2007), for Bulgarian – Slabakova (1996), among others.

- (2) a. *Аз наистина ХАРЕСВАМ вампирите.*
 I really like-1.SG.PRS vampire-PL.DEF.
 ‘I really like vampires.’

6 Frame semantics and verbs of emotion

- b. Тази постоянна светлина почва да ме
 This-FEM.SG constant-FEM.SG light start-3.SG.PRS to I-ACC
ДРАЗНИ.
 annoy-3.SG.PRS.
 ‘This constant light is starting to annoy me.’
- c. Мисля, че това му ХАРЕСВА.
 Think-1.SG.PRS that it he-DAT appeal-3.SG.PRS.
 ‘I think he likes it.’

A number of authors also observe the possibility to form diathetic pairs of verbs, “in which the STIMULUS-subject verb is transitive, while its EXPERIENCER-subject counterpart is an intransitive reflexive verb” (Ovsjannikova & Say 2020: 121) marked with a reflexive pronoun or suffix. For the Bulgarian language the system of diathesis and alternations is studied by Koeva (2022).

4 Bulgarian verbs of emotion

Bulgarian verbs of emotion, which are traditionally considered part of the larger psychological class of verbs, form an intriguing set. In her research from 2008, R. Nitsolova proposes to consider verbs like *обичам* (love.v), *мразя* (hate.v), *ненавиждам* (detest.v) and others as “mental predicates for emotional attitude” (Nitsolova 2008b: 265). S. Koeva further subdivides them into predicates for emotional reaction or assessment, (i) which are expressed by verbs *харесвам* (like.v), *съжалявам* (regret.v), *радвам се* (be glad.v), *страхувам се* (fear.v), *тревожа се* (worry.v) or (ii) constructions like *благодарен съм* (be grateful.v), *яд ме е* (be mad.v), *срам ме е* (be ashamed.v), *тъжно ми е* (be sad.v) (Koeva 2019: 62–63).

Considering the argument structure of the verbs and predicative expressions for emotions in the Bulgarian language A. Dineva states that there are four types, namely: (i) one-argument constructions, realising only an EXPERIENCER, such as *вълнувам се* (be excited.v), *тъжно ми е* (be sad.v), *страх ме е* (be afraid.v), *спокоен съм* (be calm.v); (ii) two-argument structures, expressing the EXPERIENCER and the STIMULUS with causative verbs, such as *радвам* (rejoice.v), *натъжвам* (sadden.v), *ядосвам* (make angry.v), *изненадвам* (surprise.v), or (iii) the EXPERIENCER and the Object with verbs for attitude such as *обичам* (love.v), *уважавам* (respect.v), *харесвам* (like.v), *ценя* (appreciate.v), *обожавам* (adore.v) and (iv) verbs with three arguments – an EXPERIENCER and alternating arguments, expressing the STIMULUS and the ОБЪЕКТ *Книгата ми харесва* – *Харесвам книгата* (Dineva 2000).

Hristina Kukova

Tisheva distributes the verbs of emotion into two groups on the grounds of the semantic role of the subject in a sentence: subject-Stimulus verbs and subject-Experiencer verbs, forming thus converse couples (*безпокоя* – *безпокоя се* (*worry.v*), *радвам* – *радвам се* (*rejoice.v*), *обиждам* – *обиждам се* (*insult.v*) and so forth). The state verbs of emotion involved in these oppositions are reflexive as to their form and therefore intransitive. In the Bulgarian language they are regarded as medial verbs. As a general rule the emotion state verbs with *се* take the STIMULUS as a PP, while the causatives encode the STIMULUS as an NP. Tisheva notes that not all emotion verbs fall into these couples. A number of authors state that verbs like *боя се* (*be afraid.v*), *наслаждавам се* (*enjoy.v*), *страхувам се* (*fear.v*) are not used without the reflexive *се*, whereas *тъжа* (*grieve.v*), *тъгувам* (*sorrow.v*), *тържествувам* (*triumph.v*) do not have a counterpart with *се* (Koeva 1996: 24, Nitsolova 2008a: 232, Tisheva 2022: 101–102). This is one of the reasons why the common semantic model, which includes the subject and object of emotion, cannot be expressed with a universal structural equivalence.

Dzhonova and Tisheva state that these converse couples can depict one and the same situation and possess two identical valences, though filled with different actants. Causatives impose the EXPERIENCER semantic role to the direct object, while the stative predicates attribute it to the subject (Tisheva & Dzhonova 2022: 394). According to the authors, state verbs of emotion in Bulgarian are viewed as primary and causative predicates as semantically derived, according to Van Valin and LaPolla's classification of predicates (Van Valin & LaPolla 1997).

Stamenov divides the Bulgarian verbs of internal psych experiences into 12 groups on the basis of the semantic roles that each of them requires within its realisation in context (Stamenov 2021: 70). Besides verbs of emotion, Stamenov's structural classification contains mental and perception verbs. Of the 12 types that the author outlined we have picked out 7 that contain emotion verbs (at least one) and are relevant for the purpose of our research:

- (i) Intransitive verbs whose lexical meaning expresses the inseparable unity of the ACTOR and the EXPERIENCER: *копнея* (*crave.v*), *тъжа* (*grieve.v*);
- (ii) Transitive verbs with an EXPERIENCER and STIMULUS or OBJECT: *обичам* (*love.v*), *мразя* (*hate.v*), *обожавам* (*adore.v*), *харесвам* (*like.v*);
- (iii) Verbs, expressing the semantic structure of a STIMULUS and a specific effect (on the EXPERIENCER) with a predicate of the CAUSE + DEVERBAL NOUN type: *възторгвам* / *възторгвам се* (*enrapture.v* / *go into raptures.v*) – *предизвиквам* / *преживявам възторг* (*cause.v* / *experience rapture.v*),

6 Frame semantics and verbs of emotion

възхищавам / *възхищавам се* (*cause admiration.v* / *admire.v*), *вълнувам* / *вълнувам се* (*excite.v* / *experience excitement.v*);

- (iv) Verbs with the possibility of attaching clitics – *домъчнява* (*ми*) (*start to feel unhappy.v*), *причернява* (*ми*) (*start to feel unwell.v*) – *Четенето ми доскучава* (*Reading makes me bored*) / *Доскучавам на Петьо с въпросите си* (*I'm boring Petyo with my questions*);
- (v) Verbs for ambient inner state with obligatory accusative EXPERIENCER: *до-страшава ме* (*start to feel scared.v*), *доядява ме* (*start to feel angry.v*);
- (vi) Verbs for inner psychological state with obligatory dative EXPERIENCER, which allow for a second indirect object STIMULUS / THEME: *дожелява ми* (*start to feel pity.v*), *докривява ми* (*start to feel sad.v*);
- (vii) Reflexiva tantum verbs, in which the verb action is directed back to its subject EXPERIENCER: *любвам се* (*revel.v*), *срамувам се* (*feel ashamed.v*), *страхувам се* (*fear.v*);

The author also states that different senses of verbs can fall into distinct groups in his classification.

In Section 5 we deal with five of these subclasses of verbs, taking into consideration their frequency and coverage in the Bulgarian language. First, we analyse the two uppermost hypernyms of the common emotion verbs {*изпитвам, чувствам*} / {*feel, experience*}. Then we inspect the main transitive verbs expressing attitude such as *обичам* (*love.v*), *мразя* (*hate.v*), *харесвам* (*like.v*). We also deal with causatives *веселя* (*rejoice.v*), *радвам* (*gladden.v*), *плаша* (*scare.v*), which are subdivided on the basis of the entity evoking the emotion – either an AGENT or a STIMULUS. And finally, we present the stative and inchoative verbs which are the causative verbs' medial counterparts formally expressed with the verb and the reflexive *се* – *веселя се* (*rejoice (oneself).v*), *радвам се* (*gladden (oneself).v*), *плаша се* (*scare (oneself).v*). We used the WordNet definitions where it was necessary to distinguish between different senses.

5 Frames and semantic features

The EXPERIENCER and the STIMULUS are the two obligatory participants in an emotion event. The EXPERIENCER has been in the centre of much research and is known as necessarily involving a sentient participant – usually a human or an

Hristina Kukova

animate being. The behaviour and syntactic expression of the STIMULUS has received considerably less attention, as its diverse nature can scarcely be specified. For the predicates under consideration in our work the STIMULUS affects the EXPERIENCER, changing the emotions he or she experiences. This general scenario designates emotion verbs and predetermines both the possible syntactic structures within a sentence and the morphological and semantic restrictions that are imposed on the situation participants.

In the current research, we will consider some of the most frequent verbs of emotion. The study is based on their semantic frame representation, which builds on the FrameNet and WordNet structures. That is why the different subclasses of emotion verbs will be presented below with brief definitions, taken from FrameNet, if suitable, and slightly modified, if not.

We will outline first the core frame elements within the Emotions frame as most of the semantic frames under study inherit them by virtue of the relations between frames. The definition of Emotions non-lexical frame is that ‘*An EXPERIENCER has a particular emotional STATE, which may be described in terms of a specific STIMULUS that provokes it, or a TOPIC which categorizes the kind of STIMULUS. Rather than expressing the EXPERIENCER directly, it may (metonymically) have in its place a particular EVENT (with participants who are EXPERIENCERS of the emotion) or an EXPRESSOR (a body-part of gesture which would give an indication of the EXPERIENCER’s state to an external observer).*’ Both the core and the non-core Emotion frame elements are presented systematically in Table 1.

We will consider five main semantic frames demonstrating the syntactic specifics of five subclasses of emotion verbs. Firstly, we will investigate the uppermost hypernyms of the emotion verbs *узнавам:1*, *чувствам:1* (*feel:8*, *experience:3*), which are represented by the Feelings frame (Section 5.1). In Section 5.2 we will deal with the Experienced_focused_emotion frame, which is comprised of transitive verbs, expressing attitude. Thirdly, we will explore the Cause_to_experience and Stimulate_emotion frames, which represent causative verbs of emotion, having an AGENT or a STIMULUS as a subject (Section 5.3). And finally, in Section 5.4, we will examine the Emotion_directed frame, which represents the stative and inchoative verbs, formed from the causatives from Section 5.3 and the reflexive *се*.

5.1 Feeling

Definition: In this frame an EXPERIENCER experiences an EMOTION or is in an EMOTIONAL_STATE. There can also be an EVALUATION of the internal experiential state.

6 *Frame semantics and verbs of emotion*

Table 1: The Emotion frame elements.

(a) Core frame elements

Event	The occasion or happening in which EXPERIENCERS in a certain emotional state participate.
EXPERIENCER	The person or sentient entity who experiences or feels the emotion.
Expressor	It marks expressions that indicate body part, gesture or other expression of the EXPERIENCER that reflects his or her emotional state.
State	The abstract noun that describes a more lasting experience by the EXPERIENCER .
STIMULUS	The person, event, or state of affairs that evokes the emotional response in the EXPERIENCER .
Topic	The general area in which the emotion occurs. It indicates a range of possible STIMULUS .

(b) Non-core frame elements

Degree	The extent to which the EXPERIENCER’s emotion deviates from the norm for the emotion.
Empathy-target	The individual or individuals with which the EXPERIENCER identifies emotionally and thus shares their emotional response.
Reason/Explanation	The Explanation is the explanation for why the STIMULUS evokes a certain emotional response.
Manner	Any description of the way in which the EXPERIENCER experiences the STIMULUS which is not covered by more specific FEs. Manner may also describe a state of the EXPERIENCER that affects the details of the emotional experience.
Parameter	A domain in which the EXPERIENCER experiences the STIMULUS.

Hristina Kukova

The verbs *чувствам* (*feel.v*) and *изпитвам* (*experience.v*) typically evoke this semantic frame. The only synset that contains those verbs and is marked by the semantic prime *verb.emotion* is presented in Example 3 and is illustrated in Example 4:

- (3) a. BG {изпитвам; изпитам; чувствам; почувствам; преживея; преживявам; осезавам} ('изживявам емоционално състояние или афект (по отношение на някого или нещо)')
- b. EN {*feel*; *experience*} ('undergo an emotional sensation or be in a particular state of mind')

- (4) Започнали да **ЧУВСТВАТ** любов към венерианките и същата загриженост, както и към себе си.

'They started to feel love towards the Venusian women and the same concern they felt towards themselves.'

The core frame elements are EMOTION, EMOTIONAL_STATE, EVALUATION, and EXPERIENCER. The observations on Bulgarian-language material show that the transitive verbs (*чувствам* (*feel.v*), *изпитвам* (*experience.v*)) only take an EXPERIENCER and EMOTION FEs, whereas intransitive ones (*чувствам се* (*feel (one-self).v*)) encode an EXPERIENCER and an EMOTIONAL_STATE or EVALUATION (in rare cases).

EMOTION – the EMOTION is the feeling that the EXPERIENCER experiences. There are a lot more examples with *изпитвам* (*experience.v*) than with *чувствам* (*feel.v*) with a direct object position filled with a hyponym of the {emotion} synset (Examples 5 and 6).

- (5) Често **ИЗПИТВАМ** [*завист*]_{ЕМОТ} към човешките същества.

Often feel-1.SG.PRS envy to human-DEF beings.

'I often feel envious of human beings.'

- (6) Елейн изобщо не **ИЗПИТВАШЕ** [*гордост*]_{ЕМОТ}.

Eleyn at all not feel-3.SG.IMP pride.

'Elaine felt no pride at all.'

EXPERIENCER – the EXPERIENCER experiences the EMOTION or is in the EMOTIONAL_STATE. The position of the EXPERIENCER is generally occupied by a literal belonging to the eng-30-00007846-n: {*person*} synset or its hyponyms. It is expressed in a sentence by an NP, which functions as a subject. It can also be used metaphorically with a part of the body, usually the *heart*, which has the potential to serve as an expressor of one's feelings (Example 7).

6 Frame semantics and verbs of emotion

- (7) [...] *и да каже онова, което ЧУВСТВА* [сърцето ѝ]_{BODP}.
 [...] and to tell-3.SG.PRS that what feel-3.SG.PRS heart her.
 ‘[...] and to say what her heart feels.’

The *se* counterpart of *чувствам* (*feel.v*) – *чувствам се* (*feel (oneself).v*) is reflexive concerning its form and, respectively, intransitive. It does not take a direct object and, therefore, does not encode an EMOTION. In order to realise its meaning, it needs the other core frame element – EMOTIONAL_STATE.

EMOTIONAL_STATE – the EMOTIONAL_STATE is the state the EXPERIENCER is in. The EMOTIONAL_STATE can be expressed by an adjective/participle as in (Examples 8 and 9) describing the EXPERIENCER or by an adverb (Example 10), indicating the manner in which the EXPERIENCER feels.

- (8) *ЧУВСТВАМ СЕ* [свободен]_{EMOS}.
 Feel-1.SG.PRS free.
 ‘I feel free.’
- (9) *В момента тя СЕ ЧУВСТВАШЕ* [предадена]_{EMOS}.
 In moment-DEF she feel-3.SG.IMP betrayed.
 ‘Right now she felt betrayed.’
- (10) *Дали ще СЕ ЧУВСТВАШ* [отвратително]_{EMOS}, може.
 Whether will feel-2.SG.FUT disgustingly, maybe.
 ‘Will you feel disgusted, maybe.’

5.2 **Experiencer_focused_emotion**

Definition: The words in this frame describe an EXPERIENCER’s emotions with respect to some CONTENT. Although the CONTENT may refer to an actual, current state of affairs, quite often it refers to a general situation which causes the emotion.

Experiencer_focused_emotion is a semantic frame that encodes the EXPERIENCER as a subject and the CONTENT as a direct object and is well-represented in Bulgarian. This semantic frame encompasses verbs like *харесвам* (*like.v*), *обичам* (*love.v*), *мразя* (*hate.v*), *ненавиждам* (*detest.v*), *обожавам* (*adore.v*), *съжалявам* (*feel sorry.v*), *презирям* (*despise.v*), among others. As Y. Tisheva specifies “unlike the usage of some mental predicates (Nitsolova 2008b: 264), with the verb *обичам* (*love.v*) the negation does not affect the choice of lexical elements that can occupy the syntactic positions, but only the interpretation of the meaning of the whole

Hristina Kukova

sentence” (Tisheva 2021: 117). This observation can be spread over the verbs comprising this subclass with the exception of *ненавиждам* (*detest.v*), which bears negation within its structure and does not allow for a second negative element.

The core frame elements within the *Experiencer_focused_emotion* in FrameNet are the EXPERIENCER, the CONTENT, the EVENT and the TOPIC. For the sake of the description of Bulgarian verbs we will use a modified frame, taking into consideration only the EXPERIENCER and the CONTENT as the other two core frame elements are generally combined with other parts of speech. The EVENT generally describes noun phrases and the TOPIC gives additional information that was not found with the verbs under discussion. That is why the latter two core FE will not be discussed hereon.

We will consider firstly the two main meanings of the verb *обичам* (*love.v*) as reflected in the BulNet lexical-semantic resource (Examples 11 and 12). They both encode the CONTENT as a direct object or as a *da*-clause. We believe that the other verbs from the group follow the same syntactic constructions.

- (11) a. BG {обичам} (‘изпитвам силна привързаност и симпатии към някого или свързаност с и удоволствие от нещо’)
Тя обича шефа си и работи усърдно за него; Обичам френската кухня.
- b. EN {love} (‘have a great affection or liking for’)
She loves her boss and works hard for him; I love French food.
- (12) a. BG {обичам} (‘харесвам много, изпитвам удоволствие от нещо’)
Обичам да готвя.
- b. EN {love} (‘get pleasure from’)
I love cooking.

EXPERIENCER – The EXPERIENCER experiences the emotion or other internal state. The EXPERIENCER FE position is generally filled with a subtree of {*person*} (Example 13), but can also be encoded as {*animal*} and its hyponyms (Example 14).

- (13) [*Победителят*]_{EXP} **ОБИЧАШЕ** да вкарва голове.
 Winner-DEF love-3.SG.IMP to score goals.
 ‘The winner loved to score goals.’
- (14) [*Кучето*]_{EXP} **ОБИЧА** стариците.
 Dog-DEF love-3.SG.PRS old women.
 ‘The dog loves old women.’

6 Frame semantics and verbs of emotion

The EXPERIENCER is usually expressed by a singular noun, as shown in Examples 13 and 14. When the subject EXPERIENCER is in plural, it marks a specific group, which comes as a collective subject of the emotion as in Example 15.

- (15) [Магьосниците]_{EXP} **ОБОЖАВАТ** *неразкритите тайни*.
 Magician-PL.DEF adore-3.PL.PRS undiscovered-PL.DEF secrets.
 ‘Magicians adore undiscovered secrets.’

In addition, the EXPERIENCER’s position is often occupied by {душа} (*soul*) or {сърце} (*heart*) synsets together with expressive modifiers or quantifiers to reveal the point where the feeling is concentrated (Example 16).

- (16) [Душата ми]_{EXP} до болка те **ОБИЧА**.
 Soul-DEF my to pain you-ACC love-3.SG.PRS.
 ‘My soul loves you painfully.’

Examples with metonymic shifts of the type *Barcelona loves partying* (meaning the people of Barcelona) actually show that a great number of lexical units can possibly occupy a certain position if they can express the same semantic role. Our aim is to outline the syntactic regularities and that is why such occasional examples lie beyond the scope of this study.

CONTENT – CONTENT is what the EXPERIENCER’s feelings or experiences are directed towards or based upon. The CONTENT differs from a STIMULUS because the CONTENT is not construed as being directly responsible for causing the emotion. The CONTENT FE is commonly expressed by a noun of the {*person*} or {*animal*} subtrees, as is shown in Example 17, but the position of this FE can generally be occupied by any {*entity*} hyponym, alluding to a specific human being (e.g. one’s voice, as shown in Examples 18 and 19).

- (17) Искрено да **ОБИЧА** [родителя]_{CONT} [...]
 Sincerely to love-3.SG.PRS parent-DEF [...]
 ‘To sincerely love the parent [...]’
- (18) **ОБИЧАМ** [звука на гласа ти]_{CONT}!
 Love-1.SG.PRS sound-DEF of voice-DEF your!
 ‘I love the sound of your voice!’
- (19) Библиотекарят **ОБИЧАШЕ** [театъра]_{CONT}.
 Librarian-DEF love-3.SG.IMP theatre-DEF.
 ‘The librarian loved theatre.’

Hristina Kukova

There are examples in the corpus where the CONTENT is also conveyed metaphorically as in Example 20, where the glass actually symbolises the EXPERIENCER's attitude towards drinking.

- (20) *Чичо Ташко ОБИЧАШЕ [чашката]_{CONT} [...]*
 Uncle Toshko love-3.SG.IMP glass-DEF [...]
 'Uncle Toshko loved to drink [...]

The CONTENT of the emotion can also be expressed in Bulgarian by a subordinate clause, introduced by the conjunction *да* (Example 21), interrogative pronoun *как* (Example 22) and relative pronoun *когато* (Example 23). When the object position is filled with a clause, there is no structural dependency between the arguments of the predicates in the main and the subordinate clause.

- (21) *Повече ѝ ХАРЕСВАШЕ [да прочете]_{CONT} [...]*
 More she-DAT LIKE-3.SG.IMP to read [...]
 'She liked more to read [...]
- (22) *И МРАЗЕШЕ [как я гледаш]_{CONT} [...]*
 And hate-3.SG.IMP how she-ACC look_at-2.SG.PRS [...]
 'And she hated how you looked at her [...]
- (23) *Хората ОБИЧАТ*
 Person-PL.DEF love-3.PL.PRS
[когато някой се нуждае от свобода]_{CONT} [...]
 when somebody need-3.SG.PRS of freedom [...]
 'People like it when someone needs freedom [...]

This usage should be distinguished from the one, where *когато*-clause is used for conflicting circumstances as in Example 24. The CONTENT position in this sentence is filled with a direct object accusative pronoun *ме* (*me*).

- (24) *[...] как можеш да [ме]_{CONT} ОБИЧАШ, когато*
[...] how can-2.SG.PRS to I-ACC love-2.SG.PRS, when
едва снощи се срещнахме [...]
only last night REFL met-2.PL.PRS [...]
 '[...] how can you love me when we met only last night [...]

English verbs show similar usage to Example 23 when projecting the non-core frame element CIRCUMSTANCES with the help of a finite *wh*-complement, which is typically preceded by a null NP object (Example 25).

6 Frame semantics and verbs of emotion

- (25) *I HATE it when you do that.*

An adjunct *когато*-clause is also used in Example 26, as *обичам* (*love.v*) takes a complement *да*-clause, which occupies the position of the CONTENT. The proximity or remoteness of a phrase/clause and the verb does not affect the logical structure of the sentence.

- (26) *Не ОБИЧАМ, когато ѝ говоря колко много я обичам,*
 Not love-1.SG.PRS when she-DAT tell-1.SG.PRS how much I love her,
 [*тя да мълчи*]_{CONT}.
 she to be silent.
 ‘I don’t like her keeping silent when I tell her how much I love her.’

Finally, we are going to examine a more specific sense of the verb *обичам* (*love.v*) as in Example 27, as it demonstrates high frequency of usage.

- (27) a. BG {*обичам*} (‘влюбен съм, изпитвам любов към някого’)
Тя искрено обичаше съпруга си.
 b. EN {*love*} (‘be enamored or in love with’)
She loves her husband deeply.

Within this particular meaning of the verb, the positions of the EXPERIENCER and the CONTENT are semantically restricted to {*person:1*} synset and its hyponyms only. Moreover, as Tisheva states, they should reflect a single individual, so both FE should be expressed by singular nouns considering their grammatical form (Tisheva 2021: 124) or implying allocation when expressed with a plural form. The usage of the defined form in singular or plural usually shows generic usage. Example 28 is from Tisheva’s paper.

- (28) *Майката ОБИЧА децата си.*
 Mother-DEF love-3.SG.PRS child-PL.DEF REFL-POSS.
 ‘A mother loves her children.’

5.3 Stimulate_emotion and Cause_to_experience

We will consider these two semantic frames and the lexical units that evoke them in parallel as they show great similarity concerning the sentence structure and the situation participants, displaying differences only with regard to one of the frame elements. Both frames denote two core frame elements which are conventionally expressed.

Hristina Kukova

Stimulate_emotion's definition is "Some phenomenon (the STIMULUS) provokes a particular emotion in an EXPERIENCER." Its core frame elements are an EXPERIENCER and a STIMULUS, defined as follows:

EXPERIENCER: the EXPERIENCER reacts emotionally or psychologically to the STIMULUS.

STIMULUS: the STIMULUS is the event or entity which brings about the emotional or psychological state of the EXPERIENCER.

Within the *Cause_to_experience* frame an EXPERIENCER and an AGENT can be pointed out as core frame elements and the definition of the frame is "An AGENT intentionally seeks to bring about an internal mental or emotional state in the EXPERIENCER".

AGENT: the AGENT is an external argument of the target word and purposefully arouses an emotional state.

EXPERIENCER: the EXPERIENCER is the person the AGENT causes to have a particular emotional state.

The semantic and syntactic restrictions of the EXPERIENCER frame element will be identical for both semantic frames. It should be an *{animate being}* (Example 29), but more often than not the position is limited to a *{person}* NP.

- (29) *Изведнѹж СТЯСКАМЕ* [заек]_{EXP} [...]
 Suddenly startle-1.PL.PRS rabbit [...]
 'Suddenly, we startle a rabbit [...]'

The examples with an explicit presence of the STIMULUS of the emotion and an unexpressed EXPERIENCER comprise an interesting case (Example 30). In her study on the predicative construction *it is known*, Nitsolova notes that "the place of the EXPERIENCER in the semantic structure is actually occupied with a multitude of epistemic subjects. The set of epistemic subjects includes at least the speaker himself, who wants to include the listener as well" (Nitsolova 2001: 175). This observation can also be applied to the unexpressed EXPERIENCER of the causative predicates of emotion: the object is present in the semantic structure of the predicate and represents a multitude of individuals.

- (30) [Москва]_{AGE} ПЛАШИ, че ще разкрие истината [...]
 Moscow threaten-3.SG.PRS that will reveal-3.SG.FUT truth-DEF [...]
 'Moscow threatens to reveal the truth [...]'

6 Frame semantics and verbs of emotion

Both semantic frames can be evoked by verbs such as *ужасявам* (*terrify.v*), *плаша* (*scare.v*), *разстройвам* (*upset.v*), *веселя* (*rejoice.v*), *радвам* (*gladden.v*), *успокоявам* (*comfort.v*), *вълнувам* (*excite.v*), *забавлявам* (*entertain.v*), *стряскам* (*startle.v*), which are causative and, correspondingly transitive. The direct object position is occupied by the EXPERIENCER, whereas the subject may be either animate or inanimate. If the source of emotion is animate, it will get agent-like interpretation and refer to the Cause_to_experience frame; while if it is inanimate, it will be projected as the STIMULUS of the emotion and appertain to Stimulate_emotion frame.

Within the Cause_to_experience frame the AGENT can only be presented with the {*person*} synonym set or its hyponyms. The Stimulate_emotion frame can encode any {*entity*} hyponyms in its subject position with the exception of {*person*}. Apart from that, the STIMULUS can also be encoded as a clause. As Koeva notes, the complementisers in Bulgarian are represented by the conjunctive links *че*, *да*, *как*, and *демо* (Koeva 2021b: 18). This statement can be applied to predicates of emotion whose complement clauses representing the STIMULUS are generally introduced by the one of those complementisers. Some of the verbs allow for all types of clauses, whereas some verbs do not demonstrate any usage in the corpus with some of them. Table 2 shows the distribution of the possible conjunctions with the predicates under study as presented in the Bulgarian National Corpus³.

The results from the corpus search show that *веселя* (*rejoice.v*), *разстройвам* (*upset.v*) and *стряскам* (*startle.v*) can have only a NP in the subject position. Table 2 shows that *радвам* (*gladden.v*) is the only verb that allows for all four conjunctions. *Ужасявам* (*terrify.v*), *успокоявам* (*comfort.v*) and *забавлявам* (*entertain.v*) can have *че*- and *да*-clauses (Example 31), but do not show usage with the other two complementisers. *Плаша* (*scare.v*) can be used with *че*-, *да*- and *демо*-constructions in the subject position, and *вълнувам* (*excite.v*) – with *че* and *как* complementisers (Example 32).

- (31) *Винаги го ЗАБАВЛЯВАШЕ, [че арабите им*
 Always he-ACC entertain-3.SG.IMP that Arab-PL.DEF they-DAT
*вярваха]*_{STIM}.
 believe-3.PL.IMP.
 ‘It always amused him that the Arabs believed them.’
- (32) *Изобищо не ме ВЪЛНУВА [как изглежда]*_{STIM}!
 At all not I-ACC CARE-3.SG.PRS how look like-3.PL.PRS!
 ‘I don’t care at all what it looks like!’

³We have documented the results from the corpus-based search, although the values are not always identical with our linguistic-intuition ones.

Hristina Kukova

Table 2: The distribution of causative verbs and possible complementisers.

verbs	<i>че</i> (<i>that</i>)	<i>да</i> (<i>to</i>)	<i>как</i> (<i>how</i>)	<i>демо</i> (<i>as/for/that</i>) ^a
<i>ужасявам</i> (<i>terrify</i>)	+	+	–	–
<i>плаша</i> (<i>fear</i>)	+	+	–	+
<i>разстройвам</i> (<i>upset</i>)	–	–	–	–
<i>веселя</i> (<i>rejoice</i>)	–	–	–	–
<i>радвам</i> (<i>gladden</i>)	+	+	+	+
<i>успокоявам</i> (<i>comfort</i>)	+	+	–	–
<i>вълнувам</i> (<i>excite</i>)	+	–	+	–
<i>забавлявам</i> (<i>entertain</i>)	+	+	–	–
<i>стряскам</i> (<i>startle</i>)	–	–	–	–

^aAccording to the Dictionary of Bulgarian Language *демо* is a conjunction formed by an adverb or a relative pronoun. It has a variety of functions in a sentence, that is why more than one possible translation is presented in the table.

The STIMULUS clause can also be introduced with the intensifying modifier *колко* (*how much / many.adv*) as in Example 33.

- (33) **УЖАСЯВА** *ме* [*колко е сериозен*]_{STIM}.
 Terrify-3.SG.PRS I-ACC how much be-3.SG.PRS serious-MASC.SG.
 ‘It terrifies me how serious he is.’

Apart from the subject clauses (Examples 31 and 33), the STIMULUS frame element can also be introduced by a *c*-PP (Examples 34). The subordinate clause applies to the PP, and not to the verb in this case.

- (34) *Неведнъж я беше СТЯСКАЛ* [*с това, което знаеше*]_{STIM}.
 Not once she-ACC startle-3.SG.PLUSQ [with this which know-3.SG.IMP].
 ‘He had startled her with what he knew more than once.’

5.4 Emotion_directed

The Emotion_directed frame encompasses stative and inchoative subject-EXPERIENCER psych verbs which are marked as having formally the reflexive *се*

6 Frame semantics and verbs of emotion

and a medial usage. It comprises of verbs such as *ужасявам се* (*feel/become terrified.v*), *плаша се* (*fear.v*), *разстройвам се* (*feel/become upset.v*), *веселя се* (*re-joyce.v*), *радвам се* (*be glad.v*), *успокоявам се* (*calm down.v*), *вълнувам се* (*be excited.v*), *забавлявам се* (*entertain.v*), *стряскам се* (*be startled.v*) and others. We will take into consideration the above mentioned verbs as the *се*-counterparts of the ones we explored in Section 5.3.

Definition: this frame describes an EXPERIENCER who is feeling or experiencing a particular emotional response to a STIMULUS or about a TOPIC. There can also be a CIRCUMSTANCES FE under which the response occurs or a REASON why the STIMULUS evokes the particular response in the EXPERIENCER.

The core frame elements are the EVENT, EXPERIENCER, EXPRESSOR, REASON, STATE, STIMULUS, TOPIC. We will slightly modify this semantic frame for Bulgarian verbs's sake by excluding the EVENT, EXPRESSOR and STATE FEs as we deal with no adjectives or nouns evoking the semantic frame.

EXPERIENCER – The EXPERIENCER is the person or sentient entity that experiences or feels the emotions.

We haven't found any examples in the corpus with complement clauses in subject position. If a subject is explicitly present in the sentence, the syntactic realisations of the EXPERIENCER include nominal phrases mainly of the {*person*} subtree. There are rare instances with animate non-person beings which in that case will be occupied by {*animate being*} subtree (Example 35). As we have mentioned beforehand, Bulgarian as a pro-drop language, allows the subject position to be empty.

- (35) *Животните СЕ ПЛАШЕХА, но назад не можехме да*
 Animal-PL.DEF scare-3.PL.IMP but back not can-3.PL.IMP to
се върнем.
 go back-1.PL.PRS.
 'The animals were scared, but we could not go back.'

Metonymic transfers make it possible for non-animate objects to take the position of the subject, although few cases illustrated that type in the corpus (Example 36 and 37).

- (36) *Как СЕ ВЕСЕЛИ [градът]_{EXP}?*
 How rejoice-3.SG.PRS city-DEF]
 'How does the city have fun?'

Hristina Kukova

- (37) [...] *търсенето беше прекратено и [селото]_{EXP}*
[...]_{search-DEF} call off-3.SG.PST.PS and village-DEF
СЕ УСПОКОИ.
calm down-3.SG.PST.
‘[...] the search was called off and the village calmed down.’

STIMULUS – The STIMULUS is the person, event, or state of affairs (excluding REASON) that evokes the emotional response in the EXPERIENCER. As the last example (Example 37) shows, the STIMULUS of the emotion can be syntactically unexpressed. When this element of the emotional scenario is expressed, it is traditionally projected into a subordinate clause.

As in Section 5.3 we checked all the possible combinations of verbs and complementisers and present them in Table 3⁴. The differences from causative verbs’ usage are encircled.

Table 3: The distribution of ce-verbs and possible complementisers.

verbs	<i>че</i> (that)	<i>да</i> (to)	<i>как</i> (how)	<i>демо</i> (as/for/that)
<i>ужасявам се</i> (feel terrified)	+	+	⊕	–
<i>плаша се</i> (fear)	+	+	⊕	⊖
<i>разстройвам се</i> (feel upset)	–	–	–	–
<i>веселя се</i> (rejoice)	⊕	–	–	–
<i>радвам се</i> (be glad)	+	+	+	+
<i>успокоявам се</i> (calm down)	+	⊖	–	–
<i>вълнувам се</i> (be excited)	+	⊕	+	–
<i>забавлявам се</i> (entertain)	+	+	–	–
<i>стряскам се</i> (be startled)	⊕	–	–	–

The verbs *ужасявам се* (feel/become terrified.v), *плаша се* (fear.v), *вълнувам се* (be excited.v) show examples with the first three complement types as marked in Table 3 (Example 38). *Разстройвам се* (feel/become upset.v) does not take subordinate clauses of any type. *Веселя се* (rejoice.v), *успокоявам се* (calm down.v) and *стряскам се* (be startled.v) allow clause complements with *че* only (Example 39). The compatibility of *радвам се* (be glad.v) with subordinate conjunctions is

⁴We have documented the results from the corpus-based search, although the values are not always identical with our linguistic-intuition ones.

6 Frame semantics and verbs of emotion

significantly wider – it can be used with all four of them, according to our empirical material (Example 40). And finally, *забавлявам се* (*entertain.v*) can be used with the two most frequent conjunctions *че* and *да* (Example 41), but not with the other two.

- (38) **УЖАСЯВА СЕ** [да е далеч от теб]_{STIM}.
Be terrified-3.SG.PRS to be-3.SG.PRS far from you-ACC.
'(He) is terrified of being away from you.'
- (39) [...] а пихме и слагахме трапези да СЕ ВЕСЕЛИМ,
[...] but drink-1.PL.PST and set-1.PL.PST table-PL to rejoice-1.PL.PRS
[че те си отидоха]_{STIM}.
that they go away-PL.PST.
'[...] but drank and set tables to rejoice that they had gone.'
- (40) [...] че я обича много и СЕ РАДВА [демо
[...] that she-ACC love-3.SG.PRS much and be glad-3.SG.PRS that
всичко свърши]_{STIM} [...]
everything end-3.PL.PST [...]
'[...] that he loves her very much and is glad that everything ended [...]
- (41) И тя СЕ ЗАБАВЛЯВАШЕ [да ме учи]_{STIM}.
And she entertain-3.SG.IMP to I-ACC teach-3.SG.PRS.
'And she had fun teaching me.'

There are occasional instances in the corpus, where the complement clause is introduced with the intensifier *колко* (*how much / many.adv*) (Example 42).

- (42) [...] и СЕ РАДВАХА [колко хубав обещава
[...] and rejoice-3.PL.IMP how much fine promise-3.SG.PRS
да бъде денят]_{STIM}.
to be-3.SG.PRS day-DEF.
'[...] and rejoiced at how fine the day promised to be.'

Apart from a subordinate clause the *STIMULUS* can also be introduced by a dative clitic argument (Example 43) or a *на-*, *за-*, *от-* or *с*-PP (Example 44).

Negative-emotion verbs – *ужасявам се* (*feel / become terrified.v*), *плаша се* (*fear.v*), *разстройвам се* (*feel / become upset.v*), tend to take an *от*-PP, while positive ones prefer *на-* or *за*-PPs.

- (43) [...] да [му]_{STIM} СЕ РАДВАМ скришом.
[...] to he-DAT rejoice-3.SG.PRS secretly.
'[...] and enjoyed him secretly.'

Hristina Kukova

- (44) *Обикновено СЕ УСПОКОЯВАМЕ* [*с известния факт*]_{STIM} [...].
 Usually calm down-1.PL.PRS with known-DEF fact [...].
 ‘We usually calm down at the well-known fact [...].’

The *c*-PP appears a lot more frequently when denoting another individual or individuals, who the EXPERIENCER shares emotional response with. Within the frame structure it is marked as an EMPATHY_TARGET and is a non-core frame element (Example 45).

- (45) *Върви да СЕ ЗАБАВЛЯВАШ* [*с Едуард*]_{EMPT}.
 Go-SG.IMPER to have fun-2.SG.PRS with Edward.
 ‘Go have fun with Edward.’

REASON – the REASON is the explanation why the Stimulus evokes a certain emotional response. Another possible syntactic construction within this semantic frame is for both the STIMULUS and the REASON to appear together in one sentence. In these cases the STIMULUS is marked by a PP and the REASON – with a complement clause. Koeva points out that in these cases an inner left dislocation is observed – an argument from the subordinate clause can appear in object position with the main predicate. It can be explicitly expressed in the subordinate clause as well and is coreferent with the object in the main clause. No such examples were found in the corpus, but there are such in the Internet (Example 46).

- (46) *РАДВАМ СЕ* [*на децата*]_{STIM} [*че ходят на училище с удоволствие*]_{REAS}.
 Be glad-1.SG.PRS to child-PL.DEF that go-3.PL.PRS to school with pleasure.
 ‘I am happy for the children that they attend school with pleasure.’

6 Conclusions

The study is dedicated to outlining the semantic and syntactic behaviour of verbs of emotion together with their arguments. The used resources were described in Section 2, discussing their main characteristics and structures. The methodology of the work was pointed out in Section 4, followed by a description of the class of emotion verbs and different typological approaches dealing with them.

For the purposes of the study a set of the most frequent verbs of emotion were selected and their semantic frames were discussed. The main focus extended over

6 Frame semantics and verbs of emotion

five semantic frames, namely Feeling, Experiencer_focused_emotion, Stimulate_emotion, Cause_to_experience and Emotion_directed.

Minor semantic frames (e.g. Worry, Fear, Emotion_heat and others), covering fewer lexical units were left outside the scope of our study and will be examined in future research.

All of the semantic frames under consideration were characterised in the means of the lexical units that evoke them; their core frame elements and the possible representations they may have concerning their syntactic and semantic expression. The Feeling frame was presented with its core frame elements EXPERIENCER, EMOTION, EMOTIONAL_STATE, and EVALUATION. It was observed that the transitive verbs encode an EMOTION as a direct object, whereas the intransitive *чувствам се* (*feel (oneself).v*) involves the EMOTIONAL_STATE or EVALUATION within the sentence. The Experiencer_focused_emotion was slightly modified with a view to the description of Bulgarian verbs and the CONTENT and the EXPERIENCER were taken as the core frame elements. Different types of encoding the CONTENT were illustrated. The semantic frames Stimulate_emotion and Cause_to_experience showed similar characteristics: they both include causative verbs and have two core frame elements, one of which is the EXPERIENCER. The second core frame element is semantically expressed as a STIMULUS in the former semantic frame and as an AGENT, in the latter. The potential conjunctions, interrogative and relative pronouns that can introduce a frame element were subjected to search in the corpus and the results were exemplified and tabulated for clarity. The Emotion_directed frame encompasses the medial counterparts of the Stimulate_emotion and Cause_to_experience frames, presenting stative and inchoative emotion verbs.

Lexical units of the Feeling frame are neutral concerning the emotion they designate and their complements express the positive or negative connotation. The verbs themselves bear the semantics of a positive or negative emotion within the other four semantic frames that were discussed above.

To conclude, each semantic frame consists of a collection of frame elements that represent the semantic components or roles associated with it. The role of each frame element within a specific semantic frame is crucial to accurately represent the semantic structure and the frame conceptualisation. FEs help in capturing the relationships, roles, and interactions between the different participants and components within a semantic frame. They provide a detailed representation of a frame's conceptual content, allowing for more precise and nuanced linguistic analysis and understanding. The semantic analysis of the frame elements of Emotions frame and its five subframes under study makes it possible to collectively predict all the arguments of a semantic frame with respect to the stated

Hristina Kukova

linguistic constraints. The corresponding facets of the scenario represented for each semantic frame are a set of possible values from an inverted tree or subtree from WordNet. Sorting possible semantic components of words into groups of shared semantic type (hypernyms) contrasts with investigating the semantic argument structure of sentences using specific words.

This in-depth analysis and the manual approach to assigning semantic and syntactic information to the core frame elements provides new insights and deeper understanding of the syntactic behaviour of the verbs and their environment. Although the manual check and selection is rather time-consuming, among the strengths of the method is that it involves a precise alignment of data from various resources which are quite asymmetrical to be aligned automatically.

Abbreviations

AccCl	Obligatory accusative clitic	FE	Frame element
AdvP	Adverbial phrase	NP	Noun phrase
DatCl	Obligatory dative clitic	PP	Prepositional phrase
Emos	EMOTIONAL_STATE	Reas	REASON
Emot	EMOTION	S	Subordinate clause
EmpT	EMPATHY_TARGET	Stim	STIMULUS
Exp	EXPERIENCER		

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

References

Będkowska-Kopczyk, Agnieszka. 2014. Verbs of emotion with se in Slovene: Between middle and reflexive semantics. A cognitive analysis. *Cognitive Studies / Études cognitives* 14. 203–218.

Belaj, Branimir & Goran Tanacković Faletar. 2011. Cognitive foundations of emotion verbs complementation in Croatian. *Suvremena lingvistika* 37(72). 153–169.

Biały, Adam. 2005. *Polish psychological verbs at the lexicon-syntax interface in cross-linguistic perspective*, vol. 282. Frankfurt am Main: Peter Lang.

6 Frame semantics and verbs of emotion

- Croft, William. 1993. Case marking and the semantics of mental verbs. In James Pustejovsky (ed.), *Semantics and the lexicon* (Studies in Linguistics and Philosophy 49), 55–72. Dordrecht: Kluwer. DOI: 10.1007/978-94-011-1972-6_5.
- Czulo, Oliver, Tiago T. Torrent, Ely Edison da Silva Matos, Alexandre Diniz da Costa & Debanjana Kar. 2019. Designing a frame-semantic machine translation evaluation metric. In *Proceedings of the human-informed translation and interpreting technology workshop (HiT-IT 2019)*, 28–35.
- Dineva, Aneta. 2000. Valencey and semantic roles when expressing emotions in Bulgarian (in Bulgarian). *Bulgarian Language* 2. 1–25.
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.
- Fellbaum, Christiane. 1999a. The organization of verbs and verb concepts in a semantic net. In Patrick Saint-Dizier (ed.), *Predicative forms in natural language and in lexical knowledge bases*, vol. 6 (Text, Speech and Language Technology), 93–110. Dordrecht: Springer.
- Fellbaum, Christiane (ed.). 1999b. *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Fillmore, Charles J., Collin F. Baker & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the Conference COLINGACL '98, Montreal, Canada*, 86–90.
- Ghazi, Diman, Diana Inkpen & Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In *Proceedings of the Computational Linguistics and Intelligent Text Processing: 16th International Conference, CICLing 2015, Cairo, Egypt, Part II* 16, 152–165. Springer.
- Iordanskaja, Lidija N. 1970. Searching for the lexicographic interpretation of a group of Russian words designating feelings (in Russian). *Machine Translation and Applied Linguistics* 13(3). 26.
- Iordanskaja, Lidija N. 1973. Tentative lexicographic definitions for a group of Russian words denoting emotions. In F. Kiefer (ed.), *Trends in Soviet theoretical linguistics* (Foundations of Language, Supplementary Series 18), 389–410. Holland: D. Reidel.
- Iordanskaja, Lidija N. 1986. Russian expressions denoting physical symptoms of emotions an example of two-argument lexical functions. *Lingua* 69(3). 245–282.
- Jackendoff, Ray. 2009. *Language, consciousness, culture: Essays on mental structure*. Cambridge: MIT Press.
- Johnson-Laird, Philip Nicholas & Keith Oatley. 1989. The language of emotions: An analysis of a semantic field. *Cognition and emotion* 3(2). 81–123.

Hristina Kukova

- Koeva, Svetla. 1996. Classification of Bulgarian verbs (in Bulgarian). *Bulgarian Language* 6. 22–28.
- Koeva, Svetla. 2010. *Bulgarian FrameNet (in Bulgarian)*. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin.
- Koeva, Svetla. 2019. Complements in Bulgarian. In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2019)*, 57–68.
- Koeva, Svetla. 2021a. The Bulgarian WordNet: Structure and specific features. *Papers of Bulgarian Academy of Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla. 2021b. Towards a typological analysis of complements in Bulgarian (in Bulgarian). In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2021)*, vol. 2, 13–27.
- Koeva, Svetla. 2021c. Towards expanding WordNet with conceptual frames. In *Proceedings of the 11th Global Wordnet Conference*, 182–191.
- Koeva, Svetla. 2022. The system of diatheses in Bulgarian. In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2022)*, 80–91. Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences.
- Koeva, Svetla & Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, 203–213. Sofia, Bulgaria: Department of Computational Linguistics, IBL – BAS. <https://aclanthology.org/2022.clib-1.23>.
- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Rositsa Dekova, Tsvetana Dimitrova & Ekaterina Tarpomanova. 2012. The Bulgarian national corpus: Theory and practice in corpus design. *Journal of Language Modelling* 0(1). 65–110. DOI: 10.15398/jlm.v0i1.33.
- Kövecses, Zoltán. 1988. *The language of love: The semantics of passion in conversational English*. Lewisburg: Bucknell University Press.
- Kövecses, Zoltán. 2003. *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge: Cambridge University Press.
- Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Leseva, Svetlozara, Ivelina Stoyanova, Hristina Kukova & Maria Todorova. 2018. Integrating subcategorisation information in WordNet’s relational structure. *Bulgarian Language* 2. 13–40.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.

6 Frame semantics and verbs of emotion

- Levin, Beth & Malka Rappaport Hovav. 2005. *Argument realization*, vol. 10. Cambridge: Cambridge University Press.
- Lewis, Michael, Jeannette Haviland-Jones & Lisa Feldman Barrett. 2010. *Handbook of emotions*. New York: Guilford Press.
- Lyashevskaya, Olga & Elena Paducheva. 2011. Ontological categories of the names of emotions (in Russian). *Research and Technology Information, Series 2: Information Processes and Systems* 5. 23–31.
- Miller, George A. 1995. WordNet: A lexical database for English. *Commun. ACM* 38(11). 39–41.
- Nitsolova, Ruselina. 2001. Modern Linguistic Theories. A Guide in Syntax. In Svetla Koeva (ed.), chap. Meaning and Grammatical Specifics in Complex Sentences with Predicates for Knowledge in Bulgarian (in Bulgarian), 174–192. Plovdiv: Plovdiv University Press.
- Nitsolova, Ruselina. 2008a. *Bulgarian grammar. Morphology (In bulgarian)*. Sofia: Publishing house of Sofia University.
- Nitsolova, Ruselina. 2008b. Problems of the complex sentences with complements in Bulgarian (in Bulgarian). *South Slovene Philology* 59. 64.
- Ortony, Andrew, Gerald Clore & Allan Collins. 1988. *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CBO9780511571299.
- Ovsjannikova, Maria. 2013. Encoding and semantic properties of stimulus in Russian: Verbs of anger and beyond. *Russian linguistics* 37(1). 21–33.
- Ovsjannikova, Maria & Sergey Say. 2020. The instrumental case in the diachrony of Russian reflexive verbs of emotion: From cause to content. *Scando-Slavica* 66(1). 118–143.
- Rozwadowska, Bożena & Ewa Willim. 2016. The puzzle of psychological verbs: (Why) are emotion verbs grammatically special. In Marek Kuźniak, Bożena Rozwadowska & Michał Szawerna (eds.), *From motion to emotion: Aspects of physical and cultural embodiment in language* (Łódź Studies in Language 45), 11–27. Berlin: Peter Lang.
- Rozwadowska, Bożena, Peter Kosta & Lilia Schürcks. 2007. Various faces of the psych-phenomenon in Polish. In Peter Kosta & Lilia Schürcks (eds.), *Linguistic investigations into formal description of Slavic languages: Contributions of the sixth European conference held at Potsdam University, November 30–December 2, 2005*, 557–575. Berlin: Peter Lang.
- Ruppenhofer, Josef. 2018. The treatment of emotion vocabulary in FrameNet: Past, present and future developments. In Alexander Ziem, Lars Inderelst & Detmer Wulf (eds.), *Frames interdisziplinär: Modelle, Anwendungsfelder, Methoden*, 95–122. Düsseldorf: Dusseldorf University Press.

Hristina Kukova

- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin. F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.
- Scherer, Klaus. 2005. What are emotions? And how can they be measured? *Social science information* 44(4). 695–729.
- Slabakova, Roumyana. 1996. Bulgarian psych verbs: A case for distributed morphology. *McGill Working Papers in Linguistics* 11. 85–104.
- Sonnenhauser, Barbara. 2010. The event structure of verbs of emotion in Russian. *Russian Linguistics* 34(3). 331–353.
- Stamenov, Maksim. 2021. Groups of verbs for internal psychological experience. An attempt at classification (in Bulgarian). In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2021)*, vol. 2, 69–75. Professor Marin Drinov Publishing House of BAS.
- Strickland, Bonnie. 2000. *The gale encyclopedia of psychology*. Morth America: Gale/Cengage Learning.
- Subirats, Carlos & Miriam R. Petruck. 2003. Surprise: Spanish FrameNet. In *Proceedings of CIL 2003*, vol. 17. Matfyzpress, Prague.
- Subirats, Carlos & Hiroaki Sato. 2004. Spanish framenet and framesql. In *Proceedings of the 4th International Conference on Language Resources and Evaluation. Workshop on Building Lexical Resources from Semantically Annotated Corpora. Lisbon (Portugal)*.
- Tisheva, Yovka. 2021. Observations on the syntactic realisation of predicates for emotional states (in Bulgarian). *Zeszyty Cyrylo-Metodiańskie* 10(1). 114–136.
- Tisheva, Yovka. 2022. Positive emotions and predicates: The joy (in Bulgarian). *Papers of the Institute for Bulgarian Language* 35. 90–121.
- Tisheva, Yovka & Marina Dzhonova. 2022. Syntactic characteristics of emotive predicates in Bulgarian: A corpus-based study. In *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, 75–80.
- Van Valin, Robert D. & Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.
- Wierzbicka, Anna. 1971. *Kocha–łubi–szanuje: Medytacje semantyczne*. Warszawa: Wiedza Powszechna.
- Wierzbicka, Anna. 1972. *Semantic primitives*. Frankfurt am Main: Athenäum-Verlag.
- Wierzbicka, Anna. 1980. *Lingua mentalis: The semantics of natural language*. Sydney/New York: Academic Press. 368.
- Wierzbicka, Anna. 1986. Human emotions: Universal or culture-specific? *American anthropologist* 88(3). 584–594.

6 *Frame semantics and verbs of emotion*

- Wierzbicka, Anna. 1988. *The semantics of grammar*, vol. 18. Amsterdam: John Benjamins.
- Wierzbicka, Anna. 1995. Everyday conceptions of emotion: A semantic perspective. In James A. Russell, José-Miguel Fernández-Dols, Antony S. R. Manstead & Jane C. Wellenkamp (eds.), *Everyday Conceptions of Emotion: An introduction to the psychology, anthropology and linguistics of emotion*, 17–47. Dordrecht: Kluwer.
- Zaliznjak, Anna A. 1983. Semantics of the verb бояться in Russian (in Russian). *Papers of the Academy of Sciences of the Soviet Union* 42(1). 59–66.
- Zaliznjak, Anna A. 1985. *Functional semantics of predicates for internal state (based on material from French)*. (Doctor of Science Thesis (in Russian)).

Chapter 7

Basic verb vocabulary: An empirical approach to argument structure and word associations

Valentina Stefanova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

Maria Todorova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

Tsvetana Dimitrova 

Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences

The article discusses the results of a pilot survey on basic verb vocabulary conducted through an online experiment in the form of language tasks employing associative stimuli, thematic stimuli, and contextual and textual stimuli. An analysis revolves around observations on the performance of the tasks by two focus groups of children – 7- to 10-year-olds and 11- to 14-year-olds. Semantic frame representations of the selected verbs are employed to explain the specifics of the respondents' competence in recognising the selected verb meanings and the arguments associated with a specific verb meaning.

1 Introduction

The article discusses the results of a survey on basic verb vocabulary in Bulgarian conducted via an online experiment in the form of five types of language



Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova. 2024. Basic verb vocabulary: An empirical approach to argument structure and word associations. In Svetla Koeva (ed.), *Universality of semantic frames and language specific Bulgarian data*, 293–325. Berlin: Language Science Press. DOI: ??

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

tasks among two focus groups – children from the primary stage of education (7- to 10-year-olds and 11- to 14-year-olds). The analysis and observations on the performance of the tasks reveal the respondents' basic competences for understanding the selected verb meanings and confirms or rejects hypothesis of basic vocabulary set.

When formulating language tasks and selecting the lexical information they test, we assumed that the respondents of all focus groups had acquired basic vocabulary in their educational and family environment and could understand and use a set of words and expressions associated with different spheres of human activity. The observed data that was part of the experiment was selected following two large language resources – (Princeton) WordNet (Fellbaum 1990) and the Bulgarian wordnet (BulNet) (Koeva 2021a).

The discussion in the article revolves around the verbs with the highest selection frequency by respondents, which we analyse according to their semantic frame representations to shed light on the realisation of specific meanings. The observations are organised around the description of the general structure of the target verb groups, semantic frame elements. Based on these observations, conclusions are drawn as to which semantic frames can describe the verbs.

Our work draws upon the hypothesis that children in the focus age groups have mastered the basic vocabulary, which includes lexical units that are freely and spontaneously used in everyday language practice. Children have been exposed to this part of the basic vocabulary in the course of their schooling and in their family environment.

Firstly, we attempt at confirming the validity of the hypothesis for the target verbs by measuring the frequency of the respondents' answers. Secondly, our aim is to determine the semantic frames which may be evoked by the target verbs, via observations on the specific frame elements recognised by the children to confirm their understanding of the verbs' meaning. The respondent's knowledge of the selected verbs is analysed through Frame Semantics (Fillmore 1982) and the semantic-syntactic description of the concrete frames in FrameNet.

The pilot experiment involved language tasks, described in more detail in Section 3. The tasks within this pilot study aim at examining whether (and which of) the verbs from the set belong to the basic vocabulary, on one hand, and on the other hand – whether the form of the experiment (and specific tasks) is appropriate for testing the respondents' (children of relevant school age) linguistic knowledge, experience and intuition about the use of the verbs, and the ability to match them with the appropriate elements associated with them.

The article is organised as follows: in Section 2, we discuss some definitions along with our motivation; Section 3 gives an outline of the experiment which

7 *Basic verb vocabulary: An empirical approach*

has been more thoroughly described in previous papers; Section 4 discusses the structures of the basic vocabulary verbs that have been investigated via the language tasks (the discussion first focuses on the thematic tasks, followed by the association tasks, and the contextual tasks.).

2 Related work and motivation

Basic vocabulary build-up and acquisition has been addressed in numerous linguistic works in recent decades with a solid theoretical as well as applied background (Meara 1980, Arnaud & Bejoint 1992, Schmitt 2007, Cohen 1990).

The process of language acquisition of a person is related to both communicative abilities and skills in mastering different levels and areas of linguistic knowledge like understanding and producing sentences and longer pieces of narrative (Karmiloff & Karmiloff-Smith 2001, Vulchanova & Vulchanov 2021).

The core of the lexical system is organised around active vocabulary (Georgiev & Duridanov 1959, Boyadzhiev 2002) consisting of lexical units that are freely and spontaneously used in everyday language practice, in all spheres of life, throughout a variety of styles, in both oral and written speech (Boyadzhiev 2002). The words that belong to the basic vocabulary are known by all native speakers of a language, serve to derive other words, and are usually inherited from earlier language stages. The basic vocabulary includes nouns denoting basic activities and states, such as eating, drinking, sleeping, lying, sitting, etc. (Georgiev & Duridanov 1959).

There are various studies and experiments of first language acquisition mechanisms in different conditions and at different ages (Vulchanova & Vulchanov 2021, Stoianova 2014, Stenger & Avgustinova 2021). The applied approaches to the study of the basic vocabulary in Bulgarian are only a few most of them involved mainly children in preschool age (Popova 2020, Andonova 2021) and people learning Bulgarian as a foreign language (Dimchev 2005, Nisheva 2013, Burov 2000).

As far as we know, the language competence and vocabulary of Bulgarian children during the first school years, in the initial stage of education, 1st – 4th grade (between the ages of 6 and 11), which are the focus of the interest of this article, haven't been subject of a separate study. According to Vulchanova et al. (2020: 682) this is the second stage of children language learning when words are integrated it into the network of lexical representations. We focus on the period of the first school years, also because it is considered that knowledge about words, grammar and discourse uses are built as independent components

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

of language competence, which can already be identified separately (Vulchanova & Vulchanov 2021).

In recent decades, strong focus has been placed on technologies for language learning and teaching. The methods used by applied linguistics (Carter 1998) and works on task-based language teaching (Dalpanagioti 2021, Dolgova & Tyler 2019) emphasise the usage-based nature of linguistic knowledge and language and give opportunity for different angles of language investigations.

Furthermore, various studies on language acquisition have employed approaches to language learning and teaching via computationally represented contextual information and the theory of Frame Semantics and FrameNet. Jódar-Sánchez (2018) demonstrate the application of Frame Semantics' principles for learning second language lexis as well as for better understanding and/or acquisition of the learner's first language lexicon. Blanco-Carrión (2006) underlines that frames, being concepts, were organised around human experience.

Employing the theory of Frame Semantics Torrent et al. (2022) consider that the notion of frame includes context as one possible source of information in language comprehension. They report on two experiments: (i) the identification of frame-evoking lexical units in sentences, and (ii) a methodology for domain adaptation in Neural Machine Translation that leverages frames and qualia for representing sentence-level context.

Although technologies are widely used in education in Bulgaria, the Bulgarian basic vocabulary and learner's competence with respect to employing available language resource have not been researched so far.

3 The experiment

The experiment was carried out in the form of language games in an online environment¹. Each game variant includes a combination of the language tasks, which are described in each section below with a discussion on results. The tasks use different types of stimuli – verbal (target verbs), associative (images), and contextual (grammatical and selective) ones. The verb stimuli are associated with thematic areas related to universal human needs and basic human activities such as nutrition, body movement, daily household activities, personal interests – leisure, favourite pastimes; knowledge of the surrounding world (weather, seasons, climate), plants and animals, etc. (thematic areas are also covered in the WordNet semantic classes (Miller et al. 1990)). Picture stimuli feature clear and recognisable objects that can be associated with actions and states represented

¹<https://ibl.bas.bg/igrasglagoli/>

7 Basic verb vocabulary: An empirical approach

by the verbs. They were selected from databases of free images, and a set of 750 graphical images Duabeitia et al. (2018).

We analyse respondents' knowledge of verb meaning according to the results in the execution of the tasks (association tasks (types 1 and 2), thematic tasks (types 3 and 4), and a contextual task (type 5), discussed in detail in Section 4.

In the two association tasks, the respondents had to choose (at least) one or more verbs according to their association with a picture stimulus and/or simple text stimuli (words, phrases, sentences). In the thematic tasks, the respondents had to choose five out of ten verbs, based only on their membership in a specific semantic class (according to WordNet). In the contextual task, the respondents had to fill in the gaps in a short text by choosing from a list of verbs, given below, and taking into consideration the contextual, grammatical, and selection specifics of the sentence and of the text as a whole.

We compare the respondent's answers with the semantic frames of the target verbs to evaluate the respondents' competence of the situational context and to observe their reactions to linguistic context and accurate use of language units and forms.

The results from the experiment were collected in Excel tables, and the answers were calculated as a percentage of 100%. For example, in a task with four possible answers (verbs) (in task type 1), the results had the following distribution: *hug.v* 34/82.90%, *caress.v* 2/4.90%, *cuddle.v* 4/9.80%, *squeeze.v* 1/2.40% (number of responses / percentage of all responses given). We determine whether the hypothesis that a verb belongs to the basic vocabulary is correct or not based on the percentage of the answers given. We have used correlation scores to analyse some of the results of the thematic and association tasks.

3.1 The experiment dataset and the target verbs

The verbs used in the experiment were selected from a set of verbs compiled according to basic vocabulary criteria as described in Koeva & Doychev (2022):

- (1) the verb's place in the structure of the Bulgarian WordNet (e.g., synonymy and hyponymy-hypernymy relations between verb synonym sets);
- (2) the membership of synonym sets to the subsets of base concepts (BCS) in WordNet²;

²The set of base concept synsets has been defined by the EuroWordNet and the BalkaNet projects <http://globalwordnet.org/resources/gwa-base-concepts/>

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

- (3) the proximity of the synonym set to the root of the lexical semantic tree with respect to the hierarchical substructure in WordNet in which the corresponding synonym set is included;
- (4) the frequency of the verbs in the Bulgarian National Corpus (BulNC) (Koeva et al. 2012) (per 1 million words) in texts of different domains and genres;
- (5) the membership of the verbs in a list of meanings with an estimate of their age of acquisition according to the study "Test-based age-of-acquisition norms for 44 thousand English word meanings" (Brysaert & Biemiller 2017).

Additional criteria for the selection of verbs for children are:

- (6) verbs belonging to selected semantic classes (Table 1) in WordNet (Miller et al. 1990);
- (7) the frequency of the verbs in a small corpus of textbooks for 7-year-olds children on Bulgarian language, fine arts, music, technology.

For example, verbs as *наранявам* ('injure')³, *отслабвам* ('reduce/lose weight'), *режа* ('cut') are selected as target verbs because of their high frequency in BulNC; they belong to the set of BCS in WordNet, and are either at or near the root of the WordNet lexical-semantic tree; they are also found in the 4th grade textbooks.

In addition, we had to take into account the relatively short attention span of children and set some limits on the number and the variety of verbs that were selected in the tasks. Thus, we have selected 188 verbs, of which 112 were targeted in the tasks (considered the 'correct' ones).

To provide coverage of the basic vocabulary requirement for inclusion of everyday activities and states, such as eating, drinking, sleeping, lying, sitting, etc., we have manually selected target verbs, using the information of semantic class of verbs in WordNet, distributed as shown in Table 1.

Verbs such as verbs of communication – *write, indicate, teach, pause, hum*; verbs of cognition – *surprise, think, decide, read*; stative verbs – *deserve, miss, be, wait*, are less covered, with few examples of creation verbs (*play*) and emotion verbs (*desire*).

The tasks involve different verb classes (instantiating different frames) which may evoke specific elements associated with them. In associative tasks, verbs of

³The Bulgarian examples are followed by their translation equivalents.

7 Basic verb vocabulary: An empirical approach

Table 1: The distribution of target verbs within WordNet semantic classes

Semantic Class	Number	Type of Tasks ^a	Examples
verb.contact	34	Associative Tasks	вися ‘hang.v’, изрязвам ‘carve.v’, копя ‘dig.v’, клякам ‘squat.v’, намазвам ‘spread.v’, etc.
verb.motion	39	Associative and Contextual Tasks	бягам ‘run.v’, излитам ‘take off.v’, скачам ‘jump.v’, навеждам се ‘bend.v’, сядам ‘sit down.v’, etc.
verb.consumption	10	Associative, Thematic Tasks	дъвча ‘chew.v’, закузвам ‘have breakfast.v’, пия ‘drink.v’, ям ‘eat.v’, etc.
verb.body	11	Thematic and Contextual Tasks	дишам ‘breathe.v’, ранявам ‘injure.v’, поруменявам ‘blush.v’, спя ‘sleep.v’, ставам ‘get up.v’, etc.
verb.change	18	Associative, Contextual Tasks	залязвам ‘fade.v’, осветявам ‘brighten.v’, раста ‘grow.v’, ври ‘boil.v’, цъфтя ‘bloom.v’, etc.
verb.weather	8	Thematic and Contextual Tasks	блести ‘shine.v’, духа ‘blow.v’, ръси ‘dew.v’, etc.
verb.perception	10	Contextual Tasks	гледам ‘look.v’, помиришвам ‘smell.v’, слушам ‘listen.v’, чувам ‘hear.v’, etc.

^aThe type of task in which the verb is used – Associative, Contextual, Thematic

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

motion and verbs of contact dominate, while in the thematic tasks the preference is given to verbs of weather, body, and consumption. The contextual tasks combine verbs from different thematic areas (e.g., verbs of consumption, verbs of change, verbs of contact, motion verbs).

For the analysis of verb associations, we rely on the associative knowledge of a frame and its frame elements. Thus, in the associative tasks the frame elements are offered as picture stimuli which were chosen as possible realisations of a frame element AGENT is a person or an animal, THEME is a physical object (a ship, a train, candle, a light bulb), INSTRUMENT is again a physical object (a shovel, scissors), LOCATION is a bench, etc.

The observations on verbs in context in the contextual tasks rely on sentences (and texts) within particular thematic areas. Context allows us to observe the frame elements – both core and non-core – in specific sentences.

4 Analysis and semantic description of basic vocabulary verbs

The analysis is based on three steps: 1) confirmation (or not) of the basic vocabulary hypothesis (using the percentage ratio); 2) observations on verb meaning competence using the semantic-syntactic structure of the target verbs; 3) observations on the statistical correlation of the results from the thematic and the association tasks.

In order to investigate the semantic structure of the target verbs and the respondents' knowledge about the verb, we use the information for semantic frames from FrameNet and lexical information for nouns from WordNet.

The semantic frames are schematic representations of situations and their participants – actors, circumstances, and other roles – which are elements of the frame. The frame elements have a name, a definition, and a semantic type, plus a specification of the relations between them (Fillmore 2000).

In FrameNet, all verbs sharing same semantic frame analyses evoke the same situation and share common frame elements and frame element relations (Ruppenhofer et al. 2016). For example, the verbs *hate*, *adore*, *despise*, *grieve*, *like*, *love* are analysed by the frame Experience_focused_emotion (Koeva 2021b).

We also investigate the semantic and syntactic compatibility between verbal predicates and the nouns that are realisations of their frame elements using the idea of conceptual frames as defined in (Koeva 2021a). They are abstract structures that determine the semantic and syntactic compatibility between verbal predicates and the concrete nouns evoking their frame elements. A particular

7 Basic verb vocabulary: An empirical approach

conceptual frame can be associated with a particular semantic class, and each frame element of the conceptual frame is associated with a set of nouns that are compatible with the predicate (Koeva 2021b).

Our observations on identification of frame-evoking lexical units in sentences lay on the hypothesis of conceptual frames that any of the noun synsets in WordNet can be a lexical representation of frame elements, thus the highest synset in a hyponym-hypernym tree can be chosen to describe the frame element (Koeva & Doychev 2022). For example, we may predict that {food} may evoke the frame element FOOD of the frame Apply_Heat. Thus, {food} and its hyponyms {meat}, {pasta}, {breakfast food}, {fish}, {seafood} can be analysed as elements of the frame evoked by the verb *boil*. Of course, not all synsets in a given tree would be appropriate for the specific frame element of a given frame.

Therefore, in the analysis of semantic classes of the target verbs below, we rely on the description of semantic frames, and comment on the pilot results with a view to the target verbs' semantic frames, their semantic and syntactic combinability with concrete lexical realizations of the frame elements.

The structure of a conceptual frame consists of three sections: lexical, morphological, and frame (syntactic and semantic) (Koeva & Doychev 2022). We will focus on the last section, where verbs are represented with semantic (as semantic restrictions to a given frame element, based on the WordNet noun synsets) and syntactic (as phrases – candidates for frame elements in Bulgarian) frame representation. As mentioned in (Koeva & Doychev 2022), semantic restrictions to a given element follow a certain context, and selective restrictions act between a predicate evoking a specific semantic frame, and its frame elements. Therefore, the two may differ for each case.

4.1 Analysis of thematically related verbs

The thematic tasks use the notion of a thematic area as defined with respect to the semantic class of the verbs. For example, the thematic area covering universal human needs and basic human activities such as *nutrition, body movement, health*, is defined by the usage of verbs related to bodily functions and form (*injure, lose weight, breathe*) or verbs of consumption (*drink, swallow, eat*). The target verbs are selected according to the WordNet semantic classes (for example, verb.body or verb.consumption), and the respondents had to choose five out of ten verbs according to a simple description of the thematic area. For example: (i) verbs representing actions or situations in the thematic area *weather* are extracted from the set of verbs labeled classified as verb.weather in WordNet; (ii) verbs for physical states or actions in the thematic area *human or animal body*

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

are extracted from the set of verbs classified as *verb.body* in WordNet; (iii) verbs in the thematic area of *using, buying or eating something* are extracted from the set of verbs labeled with the semantic class *verb.consumption* in WordNet. Below we also comment on the verb groups which are part of the experiment with emphasis on their semantic description and representation.

4.1.1 Target verbs related to bodily functions and form

The general semantics of the thematic area can be illustrated by verbs, such as *контузвам* ('injure'), *лекувам* ('cure'), and *отслабвам* ('reduce (lose) weight'), which correspond: to the frame *Experience_bodily_harm*: An EXPERIENCER is involved in a bodily injury to a BODY_PART; to the frame *Cause_bodily_experience*: An AGENT or CAUSE affects a BODY_PART of an EXPERIENCER so that the quality of the EXPERIENCER's perceptual experience of the BODY_PART is qualitatively changed; and to the frame *Cure*: a HEALER is treating and curing an AFFLICTION (injury, disease, or pain) of the PATIENT, whose BODY_PART is the specific area which is treated. All three frames share the frame element BODY_PART.

The verb *дишам* ('breathe') with the semantic class *verb.body* is under discussion, as the body (or parts of it) is not explicitly realised frame element, but is implicit in the meaning of the verb 'for man and animal – I take air through my nose and mouth into my lungs and let it out after absorbing the oxygen necessary for the body' (see DBL⁴). The verb has the highest frequency of use in the Bulgarian National Corpus compared to the other four target verbs. The semantic frame that the verb evokes is *Breathing* with core frame elements AGENT and AIR.

The verb *поруменявам* ('blush') has low frequency, does not occur in selected textbooks and was not among the targeted verbs, but the results of the experiment show that it was recognised as a verb related to the human body. Verbs associated with the *Body_description_part* frame, typically evoke an Individual through a salient body part. The frame element FIGURE profiles the specific body part.

The results show that the score is higher if a verb's meaning is related to an experience directly affecting the body. The target verbs activate a physical dimension of the human body and its state, while the other (non-target) verbs evoke the person as a subject who can perform the action indicated by the verb, such as to sprinkle something on an object (*поръсвам* ('sprinkle') is *verb.contact* analysed

⁴Dictionary of the Bulgarian Language, <https://ibl.bas.bg/rbe/>

7 Basic verb vocabulary: An empirical approach

by the frame Filling, with core frame elements AGENT, CAUSE, GOAL, THEME); to eat some food (*янавам* ('eat') is verb.consumption analysed by the frame Ingestion, with core frame elements INGESTOR, INGESTIBLES); to understand something with one's mind (*разбурам* ('understand') is verb.communication analysed by the frame Awareness, with core frame elements COGNIZER, CONTENT, EXPRESSOR, TOPIC); to draw with my hand (*чертая* ('draw') is verb.contact analysed by the frame Create_physical_artwork, with core frame elements CREATOR, REPRESENTATION); to fasten a garment on my body (*закончавам* ('fasten') is verb.contact with frame Closure and core frame elements AGENT, CONTAINER_PORTAL, CONTAINING_OBJECT, FASTENER).

Table 2⁵ shows the result: the third and the fourth column give the frame and the frame elements in FrameNet.

4.1.2 Target verbs of weather

For this task we used the thematic area related to the environment.

The verb *зърму* ('thunder'), 'WN: be the case that thunder is being heard'), and the verb *пъсу* ('sprinkle') in the third person singular in the sense of 'DBL: rain, snow, etc. – fall, rain continuously' (DBL⁶) have low frequency both in the Bulgarian National Corpus and in the textbooks, but receive a high score among the answers, as can be seen in Table 3. The verb can be analysed by the frame Make_noise, with core frame elements NOISY_EVENT, SOUND, and SOUND_SOURCE.

The features shared by the verbs, labeled under the semantic class verb.weather, are environmental conditions – temperature, precipitation, wind, and sun as SOURCE of weather – which activate a particular PLACE, AREA, and TIME, and respondents made an unambiguous choice of verbs related to weather.

4.1.3 Target verbs of consumption

The results for the verbs in the semantic class verb.consumption (Table 4) confirm the assumption of the experiment that the respondents have acquired the basic vocabulary and have developed an active vocabulary covering different areas of human activity. The hypothesis that the selected verbs belong to the basic vocabulary, which contains lexical units that are used freely and spontaneously in everyday language practice, is confirmed. The high percentage of chosen verbs that belong to the same thematic area, suggests that the respondents can freely

⁵The second and third columns of the table show the results for 7-to 10-year-olds and 11- to 14-year-olds.

⁶Dictionary Of The Bulgarian Language, <https://ibl.bas.bg/rbe/>

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

Table 2: Results and the frames of the target verbs related to bodily functions and form

Verb	7-10 yrs	11-14 yrs	Frame	Frame elements
контузвам 'injure.v'	14.60%	16.80%	Experience_ bodily_harm	BODY_PART; EXPERIENCER
отслабвам 'lose weight.v'	16.00%	18.90%	Cause_bodily_ experience	AGENT; BODY_PART; CAUSE; EXPERIENCER
лекувам 'cure.v'	10.00%	11.50%	Cure	AFFLICTION; BODY_PART; HEALER; MEDICATION; PATIENT; TREATMENT
поруменявам 'blush.v'	11.90%	9.50%	Body_descrip- tion_part (manually cho- sen)	FIGURE; INDIVIDUAL
дишам 'breathe.v'	17.80%	20.00%	Breathing	AGENT; AIR

use the set of words and expressions, associated with the set of frame elements of the semantic frame Ingestion with core frame elements INGESTOR (person or animal that eats or drinks) and INGESTIBLES (objects that are ingested). In the case of the verb *дъвча* ('chew'), the frame profiles the process rather than the consumption: in the frame Grinding a subject frame element GRINDER realises cause GRINDING CAUSE of an object PATIENT to be broken into smaller pieces.

4.1.4 Results of the experiment on thematically related verbs

As a result of the experiment, it was found that more than 80% of the participants (for both age groups) recognised the verbs of weather (*вали* ('rain') – 22.40%⁷ / 22.30%; *духа* ('blow') – 22.40% / 21.90%; *гърми* ('thunder') – 21.80% /

⁷The percentages are given in the following order: 7- to 10-year-olds / 11- to 14-year-olds.

7 Basic verb vocabulary: An empirical approach

Table 3: Results and the frames of the target verbs related to weather

Verb	7-10 yrs	11-14 yrs	Frame	Frame elements
вали ‘rain.v’	22.40%	22.30%	Precipitation	PRECIPITATION; PLACE; TIME
духа ‘blow.v’	22.40%	21.90%	Motion	AREA; DIRECTION; DISTANCE; GOAL; PATH; SOURCE; THEME
гpee ‘shine.v’	19.70%	20.50%	Light_movement	EMITTER; SOURCE; BEAM; PATH; GOAL
гърми ‘thunder.v’	21.80%	21.90%	Motion_noise	AREA; GOAL; PATH; SOURCE; THEME
ръси ‘sprinkle.v’	11.60%	8.70%	Precipitation	PLACE; PRECIPITATION; TIME

21.90%; *зpee* (‘shine’) – 19.70% / 20.50%; *ръси* (‘sprinkle’) – 11.60% / 8.70%, against 0% and 2% for the other verbs *jump*, *sleep*, *dream*, *decide*, *carry*), as well as the verbs of consumption (eating and drinking) (*пия* ‘drink’ 20.20% / 20.70%; *ханвам* (‘eat’), *дъвча* (‘chew’), *гълтам* (‘swallow’), and *зпуза* (‘nibble’) with the same result for the first age group 19.20% and for the second age group 20.70%, 21.10%, 19.10% and 17.10%). It is noted that the verb *ръси* (‘sprinkle’) means ‘rain, snow, etc. – fall, rain heavily and continuously’ is not clearly recognised as a weather-related verb, with a percentage score of 11.60% / 8.70%. Variations in choice are observed for verbs that activate semantic frames for actions related to bodily functions and form, with percentage distributions between 10% and 20% for the target verbs *контузвам* (‘injure’), *отслабвам* (‘lose weight’), *дишам* (‘breathe’), *поруменявам* (‘blush’), *лекувам* (‘cure’), and between 5% and 15% for the rest of the verbs: *sprinkle*, *bite*, *understand*, *draw* and *fasten*.

The results give us reason to consider the target verbs as part of the basic vocabulary of both age groups.

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

Table 4: Results and FrameNet realisation of target verbs related to consumption

Verb	7-10	11-14	Frame	Frame elements
пия ‘drink.v’	20.20%	20.70%	Ingestion	INGESTOR; INGESTIBLES
гълтам ‘swallow.v’	19.20%	19.10%	Ingestion	INGESTOR; INGESTIBLES
хапвам ‘eat.v’	19.20%	20.70%	Ingestion	INGESTOR; INGESTIBLES
дъвча ‘chew.v’	19.20%	21.10%	Grinding	GRINDING_ CAUSE; GRINDER; PATIENT
гриза ‘nibble.v’	19.20%	17.10%	Ingestion	INGESTOR; INGESTIBLES

The Pearson correlation coefficient is used to measure the correlation of responses between the two groups. The number of responses, correlation and diagram can be seen in Figures 1, 2 and 3.

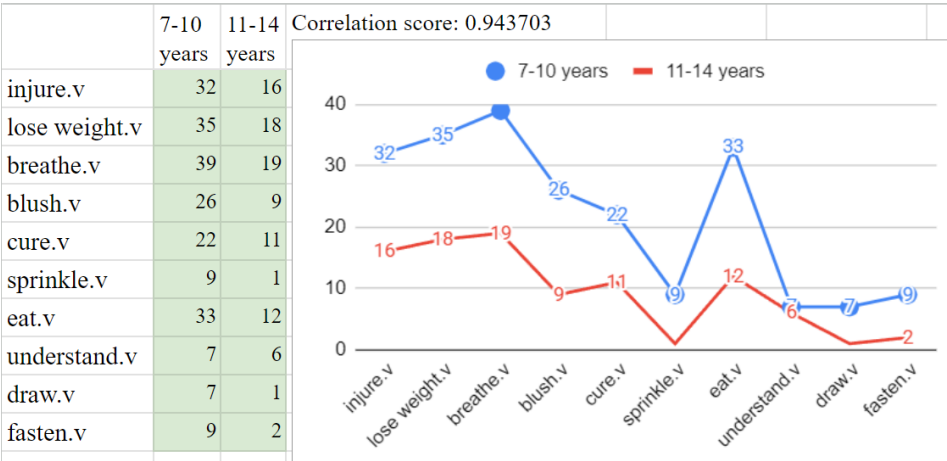


Figure 1: Correlation score of verbs related to bodily functions and form

7 Basic verb vocabulary: An empirical approach

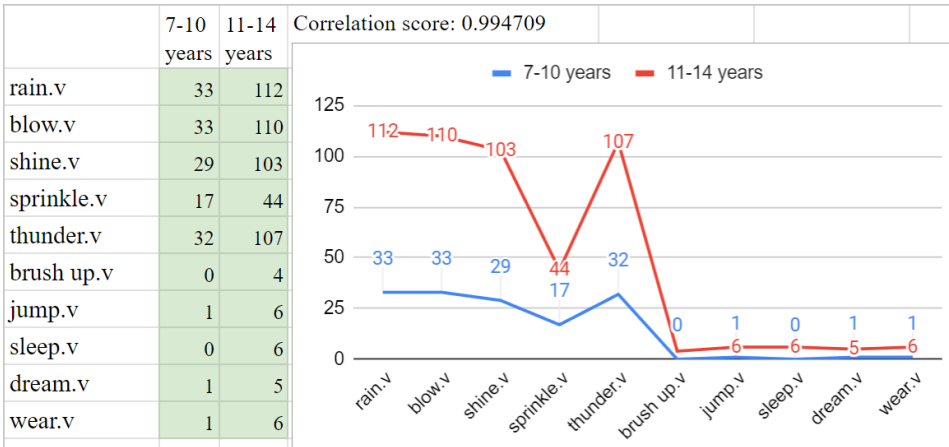


Figure 2: Correlation score of verbs of weather

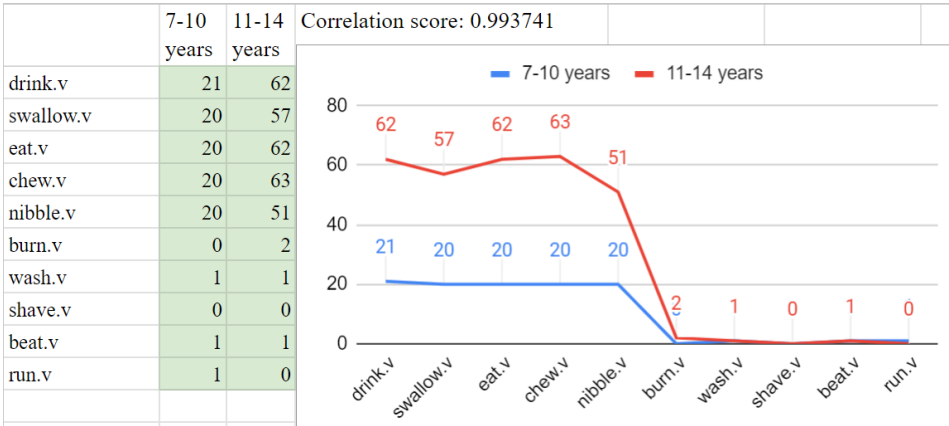


Figure 3: Correlation score of verbs of consumption

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

The result for the three groups of verbs: verbs related to bodily functions and form, verbs of weather, and verbs of consumption, is $R = 0.943702806$, $R = 0.994709117$, and $R = 0.993740859$, respectively.

This is a strong positive correlation, which means that high X variable scores go with high Y variable scores (and vice versa).

This gives us reason to conclude that high selection values in a given group increase or at least confirm the values of the other group. If a verb has a high percentage ratio in 11- to 14-year-olds, the verb may be part of the basic vocabulary for that group, but also it will be part of the basic vocabulary of the 7- to 11-year-olds, and vice versa. In other words – if a verb is mastered at the age of 7-10, then it is more likely to be part of a person’s basic verb vocabulary at the age of 11-14. And vice versa – if a verb is part of the basic verb vocabulary at the age of 11-14, it will be part of the basic verb vocabulary at a younger age.

On the other hand, the correlation between the frequency of use of the target verbs and the responses for each age group is different. The coefficient shows a moderate positive correlation between verb’s frequency of use and responses of verb of weather, and also between verb’s frequency of use and responses in the 11-14 age group of verbs related to bodily functions and form, while in the other age group the correlation is defined as technically a positive one albeit weak. With verbs of consumption, the correlation also differs between the two age groups: the correlation in the 7-10 age group is technically a negative one, albeit weak, while the correlation in the age group 11-14 is again weak, but technically a positive one (see Figure 4). Hence, the correlation between the frequency of use and the responses is not a significant factor in determining the dependence between the two variables.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	verb.weather	frequency	7-10 years	11-14 years		verb.bodily	frequency	7-10 years	11-14 years		verb.consumption	frequency	7-10 years	11-14 years
1	rain.v	39.5	33	112		breathe.v	51.61	39	19		eat.v	323.88	20	62
2	shine.v	22.29	29	103		cure.v	24.66	22	11		drink.v	40.79	21	62
3	blow.v	10.49	33	110		injure.v	22.1	32	16		swallow.v	7.44	20	57
4	thunder.v	3.96	32	107		lose weight.v	8.18	35	18		chew.v	4.6	20	63
5	sprinkle.v	0.31	17	44		blush.v	4.38	26	9		nibble.v	2.12	20	51
6	Correlation score		0.51693	0.55848		Correlation score		0.48371	0.52761		Correlation score		-0.1401	0.38324

Figure 4: Correlation score between frequency and responses

4.2 Analysis of target verbs via associative stimuli

Word associations are among the key storage mechanisms in recall (Glanzer 1972). Considering the intuitive nature of associations between the meaning of the verb, the semantic frame it evokes and its frame elements, we deem these types of tasks as the least difficult ones.

Association tasks are organised in two types.

4.2.1 Target verbs evoked by picture stimuli

In the first type, four verbs are associated with a given picture, where at least one of the verbs refers to the main sense and is assumed to be part of the basic vocabulary, without additional encoding of the manner of action (which may be encoded by prefixes, suffixes, etc.). Results confirm that respondents prefer verbs from the core vocabulary, where the picture stimulus represents an element of the respective frames (mostly core frame elements, but non-core ones as well).

For example, the picture stimulus of a running man is most often associated by respondents with the verbs *тичам* ('run') (51.2% among 7- to 10-year-olds; and 49.6% among 11- to 14-year-olds) and *бягам* ('run') (39.5% among 7- to 10-year-olds; and 44.7% among 11- to 14-year-olds). Both verbs belong to the synset {тичам; бягам} – {run}, verb.motion, with the definition 'WN: move fast by using one's feet, with one foot off the ground at any given time'. The frame is Self_Motion; the picture stimulus activates the core frame element – SELF_MOVER.

The picture stimulus of a shovel for the non-core frame element INSTRUMENT is associated with the verb *копая* ('dig') in 70% of responses, followed by *рина* ('shovel') in 19.2%, while the prefixed verbs *разривам* ('dig up') and *прекопавам* ('dig up') are chosen by less than 10% of the respondents. The picture stimulus of a bench for the core frame element LOCATION is associated with the verb *седна* ('sit'), which may instantiate the frame Change_posture (it is preferred to the prefixed verbs *поседна* ('sit down') and *приседна* ('sit down') and the manner verb *клекна* ('squat')).

The verb *светя* ('light, glisten') is associated with the picture stimulus of a light bulb, illustrating either both the SOURCE and EMISSION core elements of the Emanating frame, or the core elements of FIGURE and LIGHT of the frame Location_of_light. The verb is preferred over *осветявам* ('illuminate'), *светвам* ('light up'), *блесвам* ('shine').

Hesitancy among respondents, demonstrated by a more heterogeneous distribution of choices, is related to ambiguity or other possible association with the picture stimuli. For example, the picture stimulus of a dog standing upright on all

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

fours is most often associated with the derivative *изчаква* ('await') (40%) among 7- to 10-year-olds, less often with the non-derived verb *чака* ('wait') (24.4%), followed by the manner verb *дебна* ('lurk') (17.8%), while the 11- to 14-year-olds prefer the non-derivative *чака* ('wait') and *дебне* ('lurk') (both with 36%), possibly due to the association of the dog with aggression (the verb *дебна* ('lurk') can be described by the frame Attack). The picture corresponds to the PROTAGONIST core element of the Waiting frame.

Among all the tasks, clear preference (by over 50% of the respondents with clear margin with respect to the other possible choices) is given to basic (mostly non-derived) verbs referring to a simple action without additional specification of the manner of action, namely: *тичам* ('run'), *бягам* ('run'); *копая* ('dig') (*рина* ('shovel')), *светя* ('shine') (*осветявам* ('shine on')), *прегръщам се* ('hug') (*гушкам се* ('cuddle')), *седна* ('sit') (*поседна* ('sit')), *подстриже* ('cut (hair)') (*оформя* ('form'), *подрежа* ('cut')), *наля* ('pour'), *сипя* ('pour'), *горя* ('burn') (*изгарям* ('burn out'))).

4.2.1.1 Correlation scores of the experiment results

The correlation scores between the best and the second best answer (in %) are negative (as the children systematically have given a single answer) and statistically significant (Spearman's rho coefficient of the answers of 7- to 10-year-olds is $r_s = -0.74772$, and of the answers of 10- to 14-year-olds is $r_s = -0.82805$). The correlation scores between the best answers of the two age groups show strong positive correlation and are statistically significant ($r_s = 0.84242$) – see Figure 5.

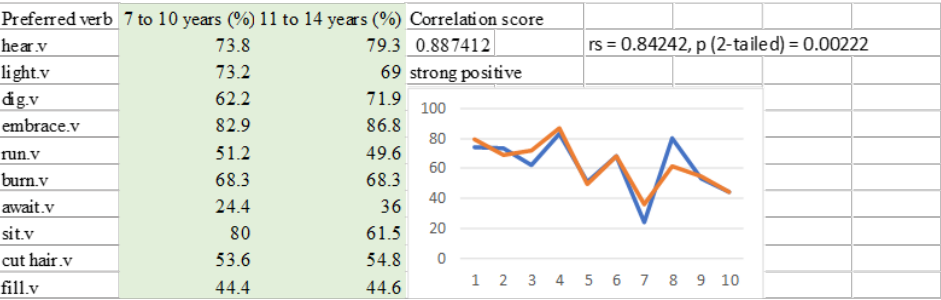


Figure 5: Correlation score between best answers of the two age groups (blue line – 7- to 10-year-olds; orange line – 10- to 14-year-olds).

7 Basic verb vocabulary: An empirical approach

4.2.2 Target verbs evoked by situation stimuli

In the second type of tasks, the picture stimulus also represents a possible realisation of a core frame element but respondents have to choose among ten verbs from different synsets, which may be analysed by different semantic frames. Five of the ten verbs are expected to be chosen, three are completely inappropriate, while the rest two are appropriate to a certain degree.

For example, the picture stimulus showing objects related to eating and drinking is associated with the verbs *закусвам* ('to eat breakfast') (20.9%⁸ among 7- to 10-year-olds; and 21.9% among 11- to 14-year-olds), *ханвам* ('snack') (17.4% and 16.7%, respectively), *ям* ('eat') (17.4% and 16%, respectively), *поханвам* ('snack') (15.1% and 16%, respectively), and *пия* ('drink') (15.1% and 15%, respectively). All five verbs may be analysed by the frame Ingestion, with the picture stimulus corresponding to the core frame element INGESTIBLES. The preference for the manner verb *закусвам* ('eat breakfast') may be related to the time constraint in the sentence stimulus (*Сутрин* ('in the morning') which illustrates the non-core frame element TIME). The picture stimulus – arranged objects related to food – also determines the next most frequent choices – *готвя* ('cook') (4.7% and 8.2%, respectively) and *подреждам* ('arrange') (3.5% and 3.9%, respectively).

Here, the disagreement among the focus groups is much lower probably due to the larger set of possible answers. The correlation scores between the highest probable and the highest less probable answer show weak positive correlation, but are not statistically significant (Spearman's rho coefficient of the answers of 7- to 10-year-olds is $r_s = -0.74772$, and of the answers of 10- to 14-year-olds is $r_s = -0.82805$).

4.2.2.1 Correlation scores of the experiment results

The correlation scores between the average of the target answers between the two age groups again show strong positive correlation and are statistically significant ($r_s = 0.81818$) – see Figure 6.

This basically confirms the conclusion from the thematically related tasks that there is correlation between basic verbs acquired by the two age groups. However, verbs' senses do not associate with each other.

Most of the verbs in both tasks evoke semantic frames that are linked to the Event top frame, as follows.

⁸All answers are equal to 100 percent, and the ratio is calculated accordingly.

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

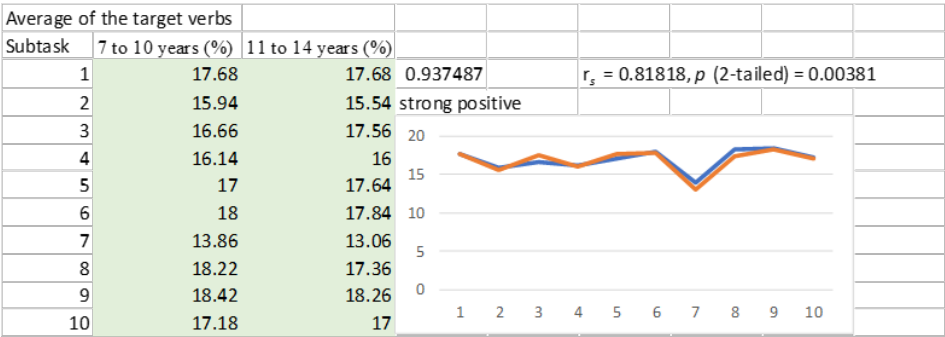


Figure 6: Correlation score between the recognised target verbs between the two age groups (blue line – 7- to 10-year-olds; orange line – 10- to 14-year-olds).

The frame Motion is linked via inheritance to Self_motion, Motion_noise, and Motion_scenario via usage to Departing, Bringing, Removing, Emanating, and via subframe to Halt.

The frame Intentionally_affect is linked to Cutting, Education_teaching, Grooming, Arranging, as well as to Communication, which, in its turn, is linked via inheritance to Communication_manner and Communication_noise and via usage to Questioning.

The frame Intentionally_create is linked to Cooking_creation, Create_physical_artwork, which is linked to Create_representation.

The frame Intentionally_act is linked to Change_posture, Perception_active, and Manipulation, which is further linked via inheritance to Ingestion.

The picture stimuli most often activate core frame elements such as: [AGENT | COMMUNICATOR | SPEAKER | TEACHER | CREATOR | PROTAGONIST | SELF_MOVER (Sentient)] (where [PROTAGONIST and SELF_MOVER can also be represented by an animal); [THEME | VEHICLE (Physical object)], as well as GOAL, BODY_PART, PATIENT, SOURCE, INGESTIBLES, MEDIUM, MESSAGE. The non-core frame elements include: INSTRUMENT, MEANS, as well as SOURCE, GOAL, BODY_PART.

The most frequent non-core frame element, evoked by the picture stimuli, is INSTRUMENT, while the most frequent core frame element is THEME. The picture stimulus may activate the preference to one verb with a core frame element instead of another with a non-core frame element – for example, a picture of an ear is most often associated with the verb *чувам* ('hear') (74% and 79.3% in the two focus groups, respectively), which may be analysed by the frame Perception_experience with the core frame element BODY_PART; and less often with the verb *слушам* ('listen'), which may be analysed by the frame Perception_active

7 Basic verb vocabulary: An empirical approach

where the BODY_PART frame element is non-core (and unexpressed).

Hesitancy among respondents – demonstrated when a verb is chosen by a number of respondents below the mean – is observed if the picture stimulus does not entirely meet the characteristics of the frame element. For example, the picture of the dog standing, mentioned above, is associated with the verbs *чака* ('wait') by 36% among the 11- to 14-year-olds (but by 24.40% among the 11- to 14-year-olds, which is below the mean), while for the prefixed verb *изчаква* ('wait') the ratio is 37.80% among the 7- to 10-year-olds against only 8% of the 11- to 14-year-olds, while the preference to *дебне* ('lurk') is the opposite – 17.80% among the 7- to 10-year-olds (which is below the mean) and 36% among the 11- to 14-year-olds. The frame element PROTAGONIST is specified as (Sentient), while the picture is of an animal. This is also the only task of the ten from the first type with such heterogeneous preference patterns (in five tasks, the preference is given to only one of the verbs; in two – to two verbs, while in the rest the two focus groups are unanimous about one verb, but do not agree on the second possible choice).

Table 5 illustrates the distribution of the respondents' choices in one of the tasks where the verbs may activate different semantic frames. The respondents have associated the meaning of the verbs stimulated by the meaning of a core frame element – for example, the verb *плува* ('swim') is analysed more often by the frame Self_motion and sentient AGENT instead of the semantic frame MOTION and THEME frame elements.

4.2.3 Prevailing semantic frame preferences in the respondents' selection

The results of the experiment on evoking verbs via associative stimuli allow us to summarise the most frequently selected frames of the experiment target verbs. The description of the frame includes a list of Bulgarian verbs, their semantic frame definition, their frame elements, information for the selectional specifics, and syntactic representation of the frame elements.

1. Semantic Frame: Cutting

Target verbs: *кълцам* ('chop'); *режа* 'cut.v' (*изрязвам, отрязвам, нарязвам, разрязвам* are language-specific verbs derived from the verb *режа* ('cut'))

Frame Definition: 'An AGENT cuts an ITEM into PIECES using an INSTRUMENT (which may or may not be expressed).'

Frame Elements: AGENT // ITEM // PIECES

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

Table 5: Results of Association task – a picture of a ship (Activated frame element is underlined)

Verb	7-10 yrs	11-14 yrs	Frame	Frame elements
плава ‘sail.v’	22.7%	20.3%	Motion	AREA; DIRECTION; DISTANCE; GOAL; PATH; SOURCE; <u>THEME</u> (PhysObj)
плува ‘swim.v’	8%	7.5%	Self_ motion	AREA; DIRECTION; GOAL; PATH; SELF_ MOVER (Sent); SOURCE
акостира ‘shore.v’	12.5%	11.4%	Vehicle_ landing	GOAL; <u>VEHICLE</u> (PhysObj)
потегля ‘de- part.v’	17%	18.3%	Departing	SOURCE (Loc); <u>THEME</u> (PhysObj)
спира ‘halt.v’	9.1%	7.8%	Halt	THEME (Physical object)
превозва ‘carry.v’	19.3%	19.9%	Bringing	AGENT (Sent); AREA; <u>CARRIER</u> ; GOAL; PATH; SOURCE; <u>THEME</u> (PhysObj)
износва ‘carry off.v’	6.8%	5.9%	Bringing	AGENT (Sent); AREA; <u>CARRIER</u> ; GOAL; PATH; SOURCE; <u>THEME</u> (PhysObj)
продължава ‘keep on.v’	4.5%	7.25%	Activity_ ongoing	AGENT (Sent); ACTIVITY; DURATION
писука ‘chirp.v’	0%	0.7%	N/A	N/A
мълчи ‘be quiet.v’	0%	1%	N/A	N/A

7 Basic verb vocabulary: An empirical approach

Semantic and selectional specifics: NP AGENT {person} *cut.v* NP ITEM {artifact}, PP PIECES {piece}

2. Semantic Frame: Ingestion

Target verbs: *ям* ('eat'), *поханвам* ('snack'), *ханвам* ('eat'), *закусвам* ('have breakfast'), *гълтам* ('swallow'), *гриза* ('nibble'), *пия* ('drink')

Frame Definition: 'An INGESTOR consumes food or drink INGESTIBLES, which entails putting the INGESTIBLES in the mouth for delivery to the digestive system.'

Frame Elements: INGESTOR // INGESTIBLES

Semantic and selectional specifics: NP INGESTOR {person} | {animal} *eat.v*
NP INGESTIBLES {food} | {nutrient} | {meat} | {fish} | {vegetable} | {fruit}
or

NP INGESTOR {person} | {animal} *drink.v* NP INGESTIBLES {beverage} | {alcoholic drink} | {water}

3. Semantic Frame: Make_noise

Target verbs: *тананикам* ('hum'), *свиря* ('play'), *крещи* ('scream'), *писука* ('chirp')

Frame Definition: 'A physical entity, construed as a point – SOUND_SOURCE, emits a SOUND. This includes animals and people making noise with their vocal tracts.'

Frame Elements: SOUND // SOUND_SOURCE // NOISY_EVENT

Semantic and selectional specifics: *chirp.v* NP SOUND {sound}, PP SOUND_SOURCE {mouth}, PP NOISY_EVENT {occurrence}

4. Semantic Frame: Motion

Target verbs: *пътува* ('travel'), *лети* ('fly'), *духа* ('blow'), *плава* ('float'), *плува* ('drift') (English verbs for *плава* ('sail') and *плува* ('swim') are recognised as Self_Motion verbs in FrameNet.)

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

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

Frame Elements: AREA // DIRECTION // DISTANCE // GOAL // PATH // SOURCE
// THEME

Semantic and selectional specifics: PP AREA, DIRECTION, DISTANCE, GOAL, PATH, SOURCE can be in some case or another – {location} | {path} | {way} | {longness} | {land} | {area}, NP [THEME] {physical object1}

5. Semantic Frame: Self_motion

Target verbs: *бягам* ('run'), *тичам* ('run'), *скачам* ('jump'), *подскачам* ('jump'), *подривам* ('caper')

Frame Definition: '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.'

Frame Elements: AREA; DIRECTION; GOAL; PATH; SELF_MOVER; SOURCE

Semantic and selectional specifics: NP SELF_MOVER {person} *run.v* PP AREA, DIRECTION, GOAL, [PATH], SOURCE can be in some case or another – {location} | {path} | {way} | {land} | {area}

4.3 Analysis of contextually related verbs

The contextual competency of respondents was tested in two types of tasks with different level of difficulty – sentence usage of thematically related verbs and textual usage of verbs. The results of the two approaches are split. Respondents were able to handle thematically related verbs placed in a specific sentence context, but showed considerable difficulty in selecting verbs in a connected text.

4.3.1 Target verbs in a sentence context

The selected target verbs and the corresponding sentences denote situations and actions in respondents' everyday life. The results of the experiment showed that over 90% of the participants recognised the meaning of the verbs and use them correctly in the context. In this type of tasks we use the syntactic realization of core frame elements of a semantic frame into the sentences as stimuli for the selection of target verbs. The sentences illustrating the verb's context represent situations evoked by semantic frames. The illustrative sentence for the verb *ям* ('eat') is analysed by the semantic frame Ingestion, *поръся* ('sprinkle') is analysed by the semantic frame Filling, *наляя* ('pour') is analysed by the frame Container_focused_placing.

7 Basic verb vocabulary: An empirical approach

All sentences have core frame element Agent in the subject position with a null instantiation. The other core frame elements are: INGESTIBLES (*delicious and healthy food*); THEME (*sprinkle salt*) GOAL (*on the toast*); THEME (*juice*); GOAL (*into a large glass*).

In other sentences, however, the frames Absorb_heat *сваря* ('boil'), Apply_heat *нпенека* ('toast'), Grinding *настържа* ('grate') remain with null instantiations of the frame elements. For example, in the sentence *I boiled an egg* the HEAT_SOURCE is not evoked, while in the sentence *Gonna grate some cheese*, the core frame element STATIC_OBJECT or TOPIC – the surfaces that rub against each other – *пенде* ('grater') is not instantiated. However, there are enough elements that are explicit, including for non-core frame elements, e.g., *I squeeze oranges for my favourite juice* [GOAL] or *I will drink with relish* [MANNER].

In the examples below we demonstrate the application of semantic-syntactic frames as stimuli in the concrete sentences from the tasks:

- (1) *Сутрин обичам да* [pro]_{ING} **ЯМ** [Ingestion] [*вкусна и здравословна храна*]_{INGIBLES}.

'In the morning I like to eat delicious and healthy food.'

- (2) *Ето сега* [pro]_{AGE} **ИЗЦЕЖДАМ** [Manipulation] [*портюкали*]_{ENT} [*за любимия сок*]_{RESULT} (*non-core*).

'Right now I am squeezing oranges for my favorite juice.'

* Unexpressed frame elements [BODYPART_OF_AGENT] – the part of the AGENT's body being used to manipulate the ENTITY.

- (3) *Преди това* [pro]_{AGE} **СВАРИХ** [Absorb_heat] [*едно яйце*]_{ENT} [*в тенджера*]_{CONTAINER}.

'Previously, I boiled an egg in a pot.'

* The source of heat treatment [HEAT_SOURCE] is not expressed.

- (4) *В момента* [pro]_{AGE} **РЕЖА** [Cutting] [*домата*]_{ITEM} [*на парчета*]_{PIECES}.

'Currently, I am slicing the tomato into pieces.'

- (5) *Взех филийки хляб, за да* [zu]_{FOOD} [pro]_{COOK} **ПРЕПЕКА** [Apply_heat] [*в тостера*]_{HEATING_INSTRUMENT}.

'I took slices of bread to toast them in the toaster.'

* Null instantiation frame elements are: [CONTAINER] – the object where food is stored and to which heat is applied, [HEATING_INSTRUMENT] –

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

the object that emits heat, [TEMPERATURE_SETTING] – the temperature at which the food is processed.

- (6) *Когато филийките са готови*, [pro]_{AGE} *ще* [zu]_{GOAL} **НАМАЖА** [Filling] [с масло]_{THM (PhysObj)}.

‘When the slices are ready, I will butter them.’

- (7) *След това* [pro]_{AGE} *ще* **ПОРЪСЯ** [Filling] [сол]_{THM (PhysObj)} [върху филийката]_{GOAL}.

‘Then I will sprinkle salt on the slice.’

- (8) [Върху филийката]_{PLACE (non-core)} *ще* **НАСТЪРЖА** [Grinding] *и* [малко кашикавал]_{PAT}.

‘On top of the slice, I will grate some cheese.’

- (9) *Накрая* [pro]_{AGE} *ще* **НАЛЕЯ** [Container_focused_placing] [портокалов сок]_{THM (PhysObj)} [в голяма чаша]_{GOAL...}

‘Finally I will pour orange juice into a large glass...’

- (10) ... *и* [pro]_{ING} *ще* **ИЗПИЯ** [Ingestion] [с наслада]_{MANNER (non-core)} [вкусната напитка]_{INGIBLES}.

‘... and I will drink with delight the delicious drink.’

4.3.2 Target verbs in a text

The tasks aimed at textual usage of words gather information about the respondents’ ability to acquire knowledge and to research, evaluate, and control this knowledge. Thus, they are the most difficult ones and combine a complex of stimuli. The tasks imply that the participants have to take into account the lexical, grammatical, and morpho-semantic specificity of the verbs studied within a task (verbs with concrete and abstract meanings from all semantic classes, i.e., cognitive verbs, verbs of emotions, stative verbs, motion verbs, etc.).

Verbs that do not fit the specific usage in the text are also embedded in the sentences. They are selected on the following principle – a phonological competitor (a paronym) of the correct verb see (Examples 11d, 12b); a verb similar in meaning, but with a syntactic realisation incompatible with the context (Examples 12b, 13a) or which does not meet the lexico-grammatical requirements for the verb form – transitivity, reflexiveness etc. (Example 12c, 13c). These principles are illustrated in the short text part – adapted fragment of “Alice in Wonderland” used in one of the tasks.

7 Basic verb vocabulary: An empirical approach

- (11) Алиса **СКУЧАЕШЕ** (страдаше (a), доскучаваше (b), нуждаеше (c)) и си **МИСЛЕШЕ** (приспиваше (d), успиваше (e), колебаеше (f)) дали да **НАБЕРЕ** (прибере (g), отнесе (h), обере (i)) един букет от маргаритки в тежката следобедна горещина.

‘Alice was beginning to get very bored (suffer (a), beginning to suffer (b), need (c))⁹ and she was **considering** (dozing off (d), starting to sleep (e), hesitating (f)) in her own mind whether to **pick** (take (g), bring (h), steal (i)) a branch of daisies in the hot afternoon.’

- (12) През това време един Бял Заек със светлочервени очи **ПОДСКОЧИ** (посочи (a), поклати (b), изсмя (c)) край нея.

‘At the same time a White Rabbit with pink eyes **ran** (pointed (a), shook (b), laughed (c)) close by her.’

- (13) Това не се **СТОРИ** (оказа (a), престори (b), помисли (c)) необикновено на Алиса и тя не **СЕ ИЗНЕНАДА** (изстрада (d), измисли (e), сметна (f)) дори когато **ЧУ** (слуша (g), попита (h), нахлу (i)) как Заекът **СИГОВО-РИ** (въобразява (a), внушава (b), спори (c)) „О, божичко, божичко!“.

‘There **was** (did (a), pretended (b), thought (c)) nothing so very **remarkable** in that; nor **did** Alice **think** (suffer (d), invent (e), count (f)) it so very much out of the way to **hear** (listen (g), ask (h), invade (i)) the Rabbit **say** (imagine (a), suggest (b), argue (c)) to itself, “Oh dear! Oh dear!”’

- (14) По-късно, като **РАЗМИСЛИ** (замисли (a), измисли (b), сметна (c)), **РЕШИ** (разреши (d), представи (e), каза (f)), че това е доста необичайно.

‘When she **thought** (considered (a), imagined (b), reckoned (c)) it over afterwards, it **occurred** (allowed (d), presented (e), said (f)) to her that she ought to have wondered at this.’¹⁰

In addition, the task relies on respondents’ knowledge of semantic and causative relations between sequences of verbs and uses those contextual relations as stimuli:

- semantic relation between target verbs as a contextual stimuli. For example: *скучаеше и си мислеше* (was bored and thought about); *размисли* (thought) → *реши* (decided) in Example 14;

⁹The translation of the alternative verbs in the tasks is literal and does not follow the above mentioned criteria – paronym, synonym etc.

¹⁰The original Bulgarian translation of the text is adapted and simplified.

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

- grammatical selection of a target verbs in the main clause and the target verb in the subordinate clause as contextual stimuli *скучаеше и си мислеше* (was bored and thought about) → *дали да набере* (whether to pick) in Example 11; *чу* (heard) → *как си говори* (her talk) in Example 12;
- grammatical combinability between target verbs prepositions and conjunctions (Examples 12, 13).

Another criterion of difficulty is the need of consideration of the information in the whole context used as last level stimuli of selection. All verb choices are presented shuffled below the text. Although in some sentences an alternative choice is possible (for example, Алиса мислеше и си каза дали да посочи... (Alice thought and said to herself whether she should indicate...) in composing the overall text, the alternatives are not acceptable.

As seen from Table 6, the largest difference in the responses between the two age groups – namely 7- to 10-year-olds and 11- to 14-year-olds – was observed at 35% and 65%, respectively, which can be explained by the complexity of the task. A total of 260 respondents filled in at least one verb position, with only 5% of respondents correctly filling in all positions in the task given, while 68% of responses were incomplete or incorrect, or possibly arbitrary.

Most errors were made with polysemous verbs, verbs of perception and verbs of cognition, such as *мисля* ('think'), *изглеждам* ('look'), or verbs like – *измъкна* ('pull'), *свия* ('shrink'), *вися* ('hang'), as well as verbs with low frequency of use, such as *здрача се* ('dusk'), *тъмнея* ('darken'), some of which are on the periphery of the basic vocabulary list.

The results for the four pre-established semantic frames activating the meanings of the verbs that were hypothesised to belong to the core lexicon and whose frame uniquely determines their position in the text, are tentative and correlated with the number of complete responses. This is probably also due to the unequal distribution of the number of responses obtained for the different options and in-between the different age groups.

5 Conclusion

In this article, we discussed the assumptions and subsequent results of a pilot survey aiming to explore whether the formulated language tasks can be used to test the respondents' acquisition of the semantics of a set of high frequency verbs, which are assumed to be part of the basic vocabulary. The language tasks activate

7 Basic verb vocabulary: An empirical approach

Table 6: Results of Contextual task – the adapted text part of *Alice in Wonderland*

Verb	Semantic class	7-10 yrs	11-14 yrs	Frame
чета 'read.v'	verb. cognition	32%	0.3%	Reading_perception: The READER attends to a TEXT to process its INFORMATION. [<i>книгата</i>] _{TEXT} , <i>която</i> [<i>сестра ѝ</i>] _{RDR} <i>четеше</i> ('The book which her sister was reading')
набера 'pick.v'	verb. contact	38%	41.9%	Food_gathering: A GATHERER removes CROP ripe and ready to an accepted degree. [<i>Алиса</i>] _{GATHERER} ... <i>да набере</i> [<i>един букет от маргаритки</i>] _{CROP} (Example 11)
подскоча 'jump.v'	verb. motion	11.2%	70%	Self_motion: The SELF_MOVER, a living being, moves under its own direction along a PATH. <i>През това време</i> [<i>един Бял Заек</i>] _{SMOV} ... <i>подскочи</i> [<i>край нея</i>] _{PATH} (Example 12)
говоря 'speak.v'	verb. communi- cation	1.7%	26%	Statement: the act of a SPEAKER to address a MESSAGE to some ADDRESSEE using language. [<i>Заекът</i>] _{SPKR} <i>да си говори</i> [<i>„О, божичко!“</i>] _{MSG} (Example 13)
свия се 'shrug.v'	verb. contact	11.9 %	27.6 %	Posture: An AGENT supports their BODY in a particular LOCATION. ... [<i>неговите уши</i>] _{BODP} <i>се свиха</i> [<i>назад</i>] _{LOC} ('his ears fell back'); [<i>Зениците му</i>] _{BODP} <i>се свиха</i> ('His pupils shrunk')
покажа се 'appear.v'	verb. change	6.7 %	1.8 %	Cause_to_perceive: An AGENT, causes a PHENOMENON to be perceived by a PERCEIVER ... [<i>в раззиналата му уста</i>] _{LOC} <i>се показа</i> [<i>белите му, остри зъби</i>] _{PHEN} ('in his gaping mouth showed his white, sharp teeth')

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

the verbs’ semantic frames and selectional preferences through different stimuli to explore the respondents’ knowledge of specific semantic features – thematic verb groups, argument selection, syntagmatic usage of correct verb form.

Based on the results, we may conclude that the selectional preferences of the verbs we explored are mainly based on associations, and the choice of the respondents depends on the stimuli – whether pictorial, or textual. In addition, verbs that are considered part of the common lexis associated with a particular thematic area, are intuitively linked to a set of participants that are also part of the area.

The experiment helps us confirm or reject the hypothesis about the affiliation of the investigated verbs to the basic vocabulary. Respondents demonstrated a good understanding of the studied verbs’ meanings related to nature, states, and actions from everyday life, as well as to material culture. They have internalised the usage of the verbs and their association with a semantic class.

The observations on the format of the experiment revealed that, in choosing the answers, respondents follow a ”strategy” influenced by the uncontrolled on-line environment. The elective nature of the language games employed leads to their perception as a set of tasks that will ultimately be evaluated. Respondents tend to search for a single ”most correct” answer, influenced by the presence of images and the arrangement of the verbs in the tasks.

The substantial difference in the results between the types of tasks is indicative in several respects such as: the difficulties in solving complex tasks, the role of the visual stimuli, the importance of knowing the selectional features of verbs such as reflexivity, transitivity, etc.

Despite the text being selected from a textbook for a lower grade, most of the respondents aged 7 to 10 faced considerable challenges with the last task. This raises several questions related to their reading skills, as well as the readability of the text for respondents of a certain age, and the presentation method of the task in the survey.

Abbreviations

Age	AGENT	Loc	LOCATION
BodP	BODY_PART	Msg	MESSAGE
Ent	ENTITY	Pat	PATIENT
HeatIns	HEAT_INSTRUMENT	Phen	PHENOMENON
Ing	INGESTOR	Rdr	READER
Ingbles	INGESTIBLES	SMov	SELF_MOVER

7 Basic verb vocabulary: An empirical approach

Spkr SPEAKER
Thm THEME

Acknowledgements

This research is carried out as part of the project *Enriching Semantic Network WordNet with Conceptual frames* funded by the Bulgarian National Science Fund, Grant Agreement No. KP-06-H50/1 from 2020.

References

- Andonova, Elena. 2021. VERBA: Adaptation for Bulgarian of questionnaires on communication development (in Bulgarian). In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2021)*, 218–224.
- Arnaud, Pierre & Henri Bejoint. 1992. *Vocabulary and applied linguistic*. Basingstoke: Macmillan.
- Blanco-Carrión, Olga. 2006. Framenet as a corpus tool for the learning of second languages and for the lexical awareness of one's first language. *Porta Linguarum* 6. 67–76. DOI: <http://hdl.handle.net/10481/30661>.
- Boyadzhiev, Todor. 2002. *Bulgarian lexicology (in Bulgarian)*. Sofia: Anubis.
- Brysbaert, Marc & Andrew Biemiller. 2017. Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Method* 49. 1520–1523. DOI: 10.3758/s13428-016-0811-4.
- Burov, Stoyan. 2000. *Bulgarian language for foreigners: Basic course in lexicon and grammar (in Bulgarian)*, vol. 1. Veliko Tarnovo: Faber.
- Carter, Ronald. 1998. *Vocabulary: Applied linguistic perspectives*. London: Routledge.
- Cohen, Andrew D. 1990. *Language learning*. New York: Newbury House.
- Dalpanagioti, Thomai. 2021. A frame-inspired task-based approach to metaphor teaching. *Lexis – Journal in English Lexicology* 18. 1–22.
- Dimchev, Kiril. 2005. *Relations between concepts first language and second language in the modern research discourse (in Bulgarian)*. Veliko Tarnovo: Linguistic Contributions in honor of corr. member prof. Mihail Videnov.
- Dolgova, Natalia & Andrea Tyler. 2019. Applications of usage-based approaches to language teaching. In Xuesong Gao (ed.), *Second handbook of English language teaching*, 1–23. Cham: Springer. DOI: 10.1007/978-3-319-58542-0_49-1.

Valentina Stefanova, Maria Todorova & Tsvetana Dimitrova

- Duabeitia, Jon Andoni, Davide Crepaldi, Antje S. Meyer, Boris New, Christos Pliatsikas, Eva Smolka & Marc Brysbaert. 2018. MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology* 71. 808–816.
- Fellbaum, Christiane. 1990. English verbs as a semantic net. *International Journal of Lexicography* 3. 278–301. DOI: 10.1093/ijl/3.4.278.
- Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Publishing.
- Fillmore, Charles J. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st meeting of the North American chapter of the association for computational linguistics*. <https://aclanthology.org/A00-2008>.
- Georgiev, Vladimir & Ivan Duridanov. 1959. *Linguistics (in Bulgarian)*. Sofia: Nauka i izkustvo.
- Glanzer, Murray. 1972. Storage mechanisms in recall. *Psychology of Learning and Motivation* 5. 129–193. DOI: 10.1016/S0079-7421(08)60441-7.
- Jódar-Sánchez, José Antonio. 2018. FrameNet as a resource to teach Spanish as a foreign language. In María Luisa Carrió-Pastor (ed.), *Teaching language and teaching literature in virtual environments*, 121–149. Singapore: Springer. DOI: 10.1007/978-981-13-1358-5_7.
- Karmiloff, Kyra & Annette Karmiloff-Smith. 2001. *Pathways to language: From fetus to adolescent*. Cambridge, MA: Harvard University Press.
- Koeva, Svetla. 2021a. The Bulgarian WordNet: Structure and specific features. *Papers of BAS Humanities and Social Sciences* 8, No. 1. 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla. 2021b. Towards Expanding WordNet with Conceptual Frames. In *Proceedings of the 11th Global Wordnet Conference*, 182–191. University of South Africa (UNISA): Global Wordnet Association. <https://aclanthology.org/2021.gwc-1.21>.
- Koeva, Svetla & Emil Doychev. 2022. Ontology supported frame classification. In *Proceedings of the fifth international conference computational linguistics in bulgaria*, 203–214. <https://aclanthology.org/2022.clib-1.23>.
- Koeva, Svetla, Ivelina Stoyanova, Svetlozara Leseva, Tsvetana Dimitrova, Rositsa Dekova & Ekaterina Tarpomanova. 2012. The Bulgarian national corpus: Theory and practice in corpus design. *Journal of Language Modelling* 0. 65–110. DOI: 10.15398/jlm.v0i1.33.
- Meara, Paul. 1980. Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics Abstracts* 13. 221–246.

7 *Basic verb vocabulary: An empirical approach*

- Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross & Katherine J. Miller. 1990. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography* 3(4). 235–244.
- Nisheva, Bozhana. 2013. Lexical units with no equivalents in the textbooks of Bulgarian language for foreigners. *OPERA SLAVICA* 23. 295–302.
- Popova, Velka. 2020. *Children's dictionary of verbal associations (in Bulgarian)*. Shumen: Shumen University Press.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. Petruck, Christopher R. Johnson, Collin F. Baker & Jan Scheffczyk. 2016. *FrameNet II: Extended theory and practice*. Berkeley, California: International Computer Science Institute.
- Schmitt, Norbert. 2007. Current perspectives on vocabulary teaching and learning. In Jim Cummins & Chris Davison (eds.), *International handbook of English language teaching* (Springer International Handbooks of Education 15). Springer, Boston, MA. DOI: 10.1007/978-0-387-46301-8_55.
- Stenger, Irina & Tanya Avgustinova. 2021. Multilingual learnability and reaction time in online Slavic intercomprehension experiments. In *Proceedings of the Annual International Conference of the Institute for Bulgarian Language (Sofia, 2021)*, vol. 2, 191–201.
- Stoianova, Juliana. 2014. Littera et lingua. Series, dissertations. In <http://slav.unisofia.bg/naum/liliseries/diss/2014/6>.
- Torrent, Tiago T., Ely Edison da Silva Matos, Frederico Belcavello, Marcelo Viridiano, Maucha Andrade Gamonal, Alexandre Diniz da Costa & Mateus Coutinho Marim. 2022. Representing context in FrameNet: A multidimensional, multimodal approach. *Front Psychol* 13. 1–20. DOI: 10.3389/fpsyg.2022.838441.
- Vulchanova, Mila, David Saldaña & Giosué Baggio. 2020. Word structure and word processing in developmental disorders. In Vito Pirrelli, Ingo Plag & Wolfgang Dressler (eds.), *Word knowledge and word usage*, 680–708. Berlin: De Gruyter Moton.
- Vulchanova, Mila & Valentin Vulchanov. 2021. Language development in typically developing children and children with developmental deficits: A concise qualitative review. *Papers of the Institute for Bulgarian Language Prof. Lyubomir Andreychin* XXXIV. 257–273. DOI: 10.7546/PIBL.XXXIV.21.10.

Universality of semantic frames and language specific Bulgarian data

The book *Universality of semantic frames and language specific Bulgarian data* is devoted to the principles of data organisation in the Bulgarian FrameNet, which has been in development for more than 20 years and has gone through various phases. Originally it was developed as an independent resource, but for about fifteen years it has been correlated with the Berkeley FrameNet, observing the following basic principles: The information in the FrameNet that is relevant for the description of Bulgarian is considered language-independent (e.g. definition of frames and relations between them, definitions of frames and elements and relations between them, etc.) and is automatically transferred into a structure called a superframe. For each superframe, there may be one or more Bulgarian frames in which the language-independent information is restructured, if necessary, so that it corresponds exactly to the description for the Bulgarian language. The Bulgarian verbs of communication, change, movement, contact and emotion are described in more detail, their subclasses are delineated and the similarities and differences in the semantic and syntactic description for Bulgarian and English are compared and discussed. It is shown how the semantic frames can be used in an experiment to assess children's mastery of semantic conceptualisation and syntactic use of verbs from their basic vocabulary. These and a number of other applications: automatic assignment of semantic roles, automatic recognition of events in news, automatic recognition of scenes in images and videos are some of the applications in which the Bulgarian FrameNet can be used. In addition, the semantic and syntactic information in the Bulgarian FrameNet can be used for theoretical considerations, including comparative studies focussing on the modern state of the Bulgarian language and other languages for which a FrameNet has been developed.