

Chapter 1

Universality of semantic frames versus 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. 2025. Universality of semantic frames versus specificity of conceptual frames. In Svetla Koeva (ed.), *Universality of semantic frames and language-specific Bulgarian data*, 1–64. Berlin: Language Science Press. DOI: [10.5281/zenodo.15682184](https://doi.org/10.5281/zenodo.15682184)



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 (Baker 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.

In the following sections, we present the structure of FrameNet, the creation of FrameNets for other languages, the motivation for the introduction of superframes and conceptual frames, 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) and in 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, 1976, 1985, 2006, 2008, Fillmore & Baker 2009), going beyond the general semantic roles of Case Grammar (Fillmore 1968).

¹<https://dcl.bas.bg/bulframe/>

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 (Baker 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 versus 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 such 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. A 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).

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 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;
- receives a definite interpretation if it is omitted (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 (Baker 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,

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

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*.

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	

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). The following example illustrates this: 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_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 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: 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.

1 Universality of semantic frames versus specificity of conceptual frames

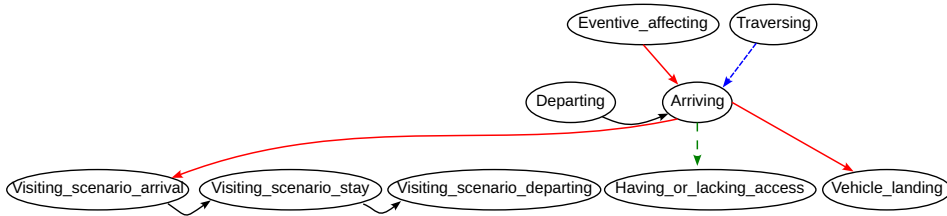


Figure 1: The immediate frame-to-frame relations of the semantic frame Arriving. Red arrows *Inheritance*, black – *Precedes*, green – *Using*, blue – *Subframe*, the direction is parent-child, the dashed lines show inverse relations.

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* 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?

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 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, da Silva Matos, 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 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).

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 by Boas (2009a): degree of overlapping cross-lingual polysemy, differences in lexicalisation patterns, measurement of para-

phrase 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, Salomão, Campos, 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 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 super-frame.

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. Superframes 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-

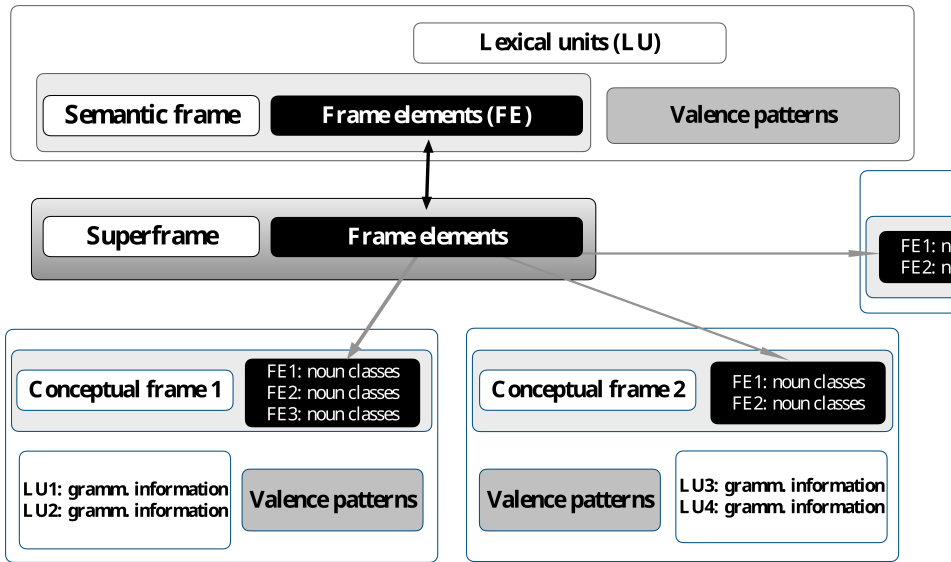


Figure 2: The correspondence between Berkeley semantic frames, superframes and conceptual frames for Bulgarian.

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.

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.

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.’

Another example is 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 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 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

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.

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 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 realisation 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

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

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) [Бацана]_{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 *ce* (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 as *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 to 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.

- 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 (с 'in'), eng-30-03990474-n: *pot*
- HEATING INSTRUMENT: optional, PP, object (с '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*, *стпрошава се*, *стпроши 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.'

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:

- 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, *захърквам* (am snoring-IPFV, ‘start snoring’); *захъркам* (snore-PFV, ‘start snoring’); *затърквам* (am looking for-IPFV, ‘start looking for’); *затърся* (look for-PFV, ‘start looking for’).

- (20) [Тоѝ]_{SOUND_SOURCE} *моментално* **ЗАХЪРКА** *в хотела.*
He instantly start_snore-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

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 (Baker 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 wordnet;⁴ 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/>

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

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

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 bucket’).

In this context, we can distinguish different types of personal verbs whose lemma is formed with a reflexive-in-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.

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* (Baker 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 (Baker 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 have-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 to evoke 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

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.

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 1990: 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 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 *is a subevent of*). 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 versus 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 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>вода</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' |

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 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

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 given by 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.

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.

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

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

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

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.’

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.'

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 massage
 [с ръце]_{BODYPART_OF_AGENT}
 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 message-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)’).

- (44) *ХРУМНА* [*ми*]_{EXP} [*една идея*]_{CONTENT}.
 Occur I-DAT an idea.
 ‘An idea occurred to me.’
- (45) *ХРУМНА* [*му*]_{EXP} [*да излезе от стаята*]_{CONTENT}.
 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 element 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 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)

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

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 is: *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 is 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

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 versus specificity of conceptual frames

Apply_heat	
Attribute	Value
Frame (saps)	<input checked="" type="checkbox"/> Included + -
Lemma	sapa (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

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;

- 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]_{TOPIC}.*
'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'.

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.

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.

☰ Syntactic obligatoriness	prodrop
▼ ☰ Syntactic category	<input checked="" type="checkbox"/> Included +
▼ ☰ NP	<input checked="" type="checkbox"/> Included
☰ Syntactic function	subject
○ Selective preferences	човек (eng-30-00007846-n)

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.

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

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. *Osnovna balgarska gramatika [Basic Bulgarian grammar]*. Sofia: Hemus Publishing House.

- 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., Charles J. Fillmore & John B. Lowe. 2003. Background to Framenet. *International Journal of Lexicography* 16(3). 235–250. DOI: 10.3115/980845.980860.
- 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*, 2nd edn., 267–292. Boca Raton: CRC Press.
- Boas, Hans C. (ed.). 2009a. *Multilingual FrameNets in computational lexicography: Methods and applications*. Berlin: 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: 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 Padó & Manfred Pinkal. 2009. Using FrameNet for the semantic analysis of German: Annotation and representation and automation. In Hans C. Boas (ed.), *Multilingual FrameNets in computational lexicography: Methods and applications*, vol. 200 (Trends in Linguistics. Studies and Monographs), 209–244. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110212976.3.209.
- 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 2014)*, 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. 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.
- 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.
- 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. 1990. English verbs as a semantic net. *International Journal of Lexicography* 3. 278–301. DOI: 10.1093/ijl/3.4.278.
- 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. 1976. Frame semantics and the nature of language. In Stevan R. Harnad, Horst D. Steklis & Jane Lancaster (eds.), *Origins and evolution of language and speech* (Annals of the New York Academy of Sciences 280), 20–32. New York: New York Academy of Sciences. DOI: 10.1111/j.1749-6632.1976.tb25467.x.
- Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin Publishing. http://brenocon.com/Fillmore%201982_2up.pdf.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6. 222–254. <https://www.cs.toronto.edu/~gpenn/csc2519/fillmore85.pdf>.
- 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. DOI: 10.3765/bls.v20i1.1478.

- 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: De Gruyter Mouton. DOI: 10.1515/9783110198775.1.129.
- Fillmore, Charles J. 2008. The merging of “frames”. In Remi Rossini Favretti (ed.), Bologna: Bononia University Press. 1–12. http://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. DOI: 10.1093/oxfordhb/9780199544004.013.0013.
- 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. https://www.euralex.org/elx_proceedings/Euralex2004/009_2004_V1_Patrick%20HANKS_Corpus%20pattern%20analysis.pdf.
- Hauck, Paulo, Fernanda Campos, Tiago T. Torrent, Ely Matos & José Maria David. 2015. FrameNet-based automatic suggestion of translation equivalents. *Proceedings of the Brazilian Seminar on Ontologies (ONTOBRAS 2015)* 1442.
- 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. Balgarskiyat FreymNet. Semantiko-sintaktichen rechnik na balgarskiya ezik – kontseptualen model [Bulgarian FrameNet. Semantic-syntactic dictionary of Bulgarian – conceptual model]. In *Balgarskiyat FreymNet. Semantiko-sintaktichen rechnik na balgarskiya ezik [Bulgarian FrameNet: Semantic-syntactic dictionary of Bulgarian]*. Svetla Koeva (ed.). Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin. 5–51.
- Koeva, Svetla. 2010a. *Balgarskiyat FreymNet [Bulgarian FrameNet]*. 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. DOI: 10.11649/cs.2012.012.
- Koeva, Svetla. 2020. Semantic relations and conceptual frames. In *Towards a semantic network enriched with a variety of semantic relations*. Svetla Koeva (ed.). Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin. 7–20.
- Koeva, Svetla. 2021. The Bulgarian WordNet: Structure and specific features. *Papers of the Bulgarian Academy of Sciences. Humanities and Social Sciences* 8(1). 47–70. https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf.
- Koeva, Svetla. 2022. The system of diatheses in Bulgarian and stative predicates. In Svetla Koeva, Elena Ivanova, Yovka Tisheva & Anton Zimmerling (eds.), *Ontologiya na situatsiite za sastoyanie – lingvistichno modelirane. Sapostavitelno izsledvane za balgarski i ruski [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. DOI: 10.7546/STONTBgRu2022.05.
- Koeva, Svetla (ed.). 2024. *Semantichni studii. Sintaktichno i semantichno opisanie na balgarski glagoli [Semantic Studies. Syntactic and semantic description of Bulgarian verbs]*. 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 Francis Bond, Takayuki Kuribayashi, Christiane Fellbaum & Piek Vossen (eds.), *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 Fifth 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 Mari-

- ani, Hélène Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, 6988–6994. Marseille, France: European Language Resources Association (ELRA). <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. *Teoretichna gramatika na balgarskiya ezik. morfologiya [Theoretical grammar of Bulgarian. Morphology]*. Plovdiv: Publishing House of Plovdiv University.
- Lenci, Alessandro, Martina Johnson & Gabriella Lapesa. 2010. Building an Italian FrameNet through semi-automatic corpus analysis. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *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/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.
- Maslov, Yuriy. 1982. *Gramatika na balgarskiya ezik [Grammar of the Bulgarian language]*. Sofia: Nauka i izkustvo.
- Materna, Jiří & Karel Pala. 2010. Using ontologies for semi-automatic linking VerbaLex with FrameNet. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *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. *Quarterly Journal of Experimental Psychology, Section A* 50(3). 528–559. DOI: 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. *Balgarska gramatika. Morfologiya [Bulgarian grammar. Morphology]*. Sofia: Publishing House of Sofia University.
- Obara, 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.
- Obara, Kyoko, Seiko Yamaguchi Fujii, Toshio Ohori, Ryoko Suzuki, Hiroaki Saito & Shun Ishizaki. 2004. *The Japanese FrameNet project: An introduction*. <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.
- 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 Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, H  l  ne Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (eds.), *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-Corr  a, Simone, Alexandre Diniz, Meire Lara, Ely Matos & Tiago T. Torrent. 2016. FrameNet-based automatic suggestion of translation equivalents. In Jo  o Silva, Ricardo Ribeiro, Paulo Quaresma, Andr   Adami & Ant  nio Branco (eds.), *Computational processing of the Portuguese language*, 347–352. Cham: Springer International Publishing.
- R  ggeberg, 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. *Glagolniiyat vid v balgarskiya knizhoven ezik [Verb aspect in the Bulgarian literary language]*. 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: 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., 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. *Frontiers in Psychology* 12. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.838441/full>.
- Torrent, Tiago T., Maria Margarida M. Salomão, 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>.
- 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. DOI: 10.1075/cf.6.1.05tor.
- 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*. Amsterdam. <https://research.vu.nl/ws/portalfiles/portal/74100710/D017>.
- You, Liping & Kaiying Liu. 2005. Building Chinese FrameNet database. In *International Conference on Natural Language Processing and Knowledge Engineering*, 301–306. DOI: 10.1109/NLPKE.2005.1598752.