Ontology Supported Frame Classification

Svetla Koeva
Institute for Bulgarian Language, BAS
svetla@dcl.bas.bg

Emil Doychev
Plovdiv University Paisiy Hilendarski
e.doychev@uni-plovdiv.bg

Abstract

We present **BulFrame** – a web-based system designed for creating, editing, validating and viewing conceptual frames. A unified theoretical model for the formal presentation of Conceptual frames is offered, which predetermines the architecture of the system with which the data is processed. A **Conceptual frame** defines a unique set of syntagmatic relations between verb synsets representing the frame and noun synsets expressing the frame elements. Thereby, the notion of Conceptual frame combines semantic knowledge presented in WordNet and FrameNet and builds upon it. The main difference with FrameNet semantic frames is the definition of the sets of nouns that can be combined with a given verb. This is achieved by an ontological representation of noun semantic classes. The framework is built and evaluated with Conceptual frames for Bulgarian verbs.

Keywords: Conceptual frames, ontology of noun semantic classes, verb semantics

1 Introduction

There are many rich semantic resources (mainly for English but also for other languages) that include different types of semantic information: WordNet (Miller et al., 1990), FrameNet (Baker et al., 1998), VerbNet (Kipper et al., 2007), PropBank (Palmer et al., 2005), Ontonotes (Weischedel et al., 2011), Pattern Dictionary of English Verbs (Hanks, 2004), Yago (Suchanek et al., 2007), BabelNet (Navigli and Ponzetto, 2012), VerbAtlas (Di Fabio et al., 2019), SynSemClass (Uresova et al., 2020), among others.

BulFrame[1] is a web-based system designed for creating, editing, validating and viewing Conceptual frames. A unified theoretical model for the formal presentation of conceptual frames is offered, which predetermines the architecture of the system with which the data is processed. In this regard, several fundamental theoretical models focused on verb semantics have been taken into account – among the most famous research in this field are Charles Fillmore's theory of frame semantics (Fillmore, 1982), the description of verb classes and possible alternations by Beth Levin (Levin, 1993), the concept of representation of verb frames in FrameNet (Baker et al., 1998; Fillmore and Baker, 2001) and others.

Some of the main advantages of both resources (WordNet and FrameNet) with regard to the conceptual description of the predicate – argument structure are complemented and upgraded to expand WordNet with Conceptual frames that represent verb predicate-argument syntagmatic relations. The main advantages of WordNet for semantic analysis focused on introducing Conceptual frames are: a) the large number of concepts organized in a semantic network and b) the grouping of concepts in semantic classes according to their general meaning. The main advantages of FrameNet for implementing Conceptual frames are: a) the extensive description of semantic knowledge about an event type and its participants and b) the linking of semantic frames with semantic relations (Koeva, 2021).

The paper is organized as follows: we begin with a brief introduction to the notion of **Conceptual frame** in Section 2. In Section 3 we present the design of the **BulFrame** system. Section 4 is dedicated to the linguistic interpretation of Conceptual frames with a special focus on the ontology of semantic classes of nouns. Finally, related work (section 5), conclusions and future directions of our work (section 6) are presented.

Our main contributions are: (a) identification of verbs that evoke a particular FrameNet semantic frame; (b) detailed ontological representation

[1] https:\dcl.bas.bg/bulframe/

of semantic classes of noun synsets; (c) specification of frame elements relevant to the expression of syntagmatic relations; (d) assigning the frame elements with noun semantic classes or a combination of classes ensuring the words' compatibility in Bulgarian; (e) definition of Conceptual frames depicting semantics of Bulgarian verbs.

2 The Notion of Conceptual Frames

Conceptual frames are abstract structures that define the semantic and syntactic compatibility between verb predicates and noun arguments. A particular Conceptual frame is: associated with a semantic class that expresses its general semantic properties; represented by a set of verbs organized in the WordNet synonym sets, and described by a set of frame elements. The verbs in the same frame can be one or several: linked between each other with lexical relations (synonymy, antonymy) and/or hierarchical relations (hypernymy, troponymy, entailment). The Conceptual frame elements roughly correspond to the FrameNet core elements; however, there is no one-to-one correspondence between FrameNet Semantic frames and Conceptual frames (because of some differences in conceptualization in different languages and because of differences between the two theoretical representations) (Koeva, 2020, 2021).

Each Conceptual frame element is associated with a set of nouns that are compatible with the verb predicate. Again, the set could contain a single noun or several nouns linked between each other with lexical relations (synonymy, antonymy) and/or hierarchical relations (hypernymy, hyponymy). The association between the frame (verb synsets) and its elements (noun synsets) can be explicitly introduced in WordNet by means of syntagmatic relations. If more than one noun synset can express the frame element (which is the usual case), the syntagmatic relation links the verb synset with the top-most noun synset of the hierarchy grouping nouns with the same semantic properties (semantic class). The diversity in the compatibilities between representatives of verb classes and noun classes drives the necessity for a detailed **Ontology of semantic classes of nouns**.

We can generalize that a **Conceptual frame** defines a unique set of syntagmatic relations between:

- verb synsets representing the frame, and
- noun synsets expressing the frame elements.

Thereby, the notion of Conceptual frame combines semantic knowledge presented in WordNet and FrameNet and builds upon it.

The main difference between Conceptual frames and the FrameNet Semantic frames (Ruppenhofer et al., 2016) is that Conceptual frames are explicitly linked with the noun synsets representing the words with which the verb predicate can be combined (to the extent this is possible due to WordNet structure and content and metaphoric language use). For example, a Conceptual frame that roughly corresponds to the FrameNet semantic frame **Experiencer focused emotion**[2] is represented by the verb synsets: **dislike** 'have or feel a dislike or distaste for'; **hate, detest** 'dislike intensely; feel antipathy or aversion towards'; **like** 'find enjoyable or agreeable'; **love** 'have a great affection or liking for'. The Conceptual frame elements are Experiencer and Content (if we keep the names of the FrameNet core elements). The semantic classes of nouns that they could be expressed with are **[Human]**, **[Animal]**, **[Physical entity]**, and **[Abstraction]** and the combinations are the following:

- **Experiencer**: person — **Content**: physical entity and abstraction
- **Experiencer**: animal — **Content**: physical entity.

3 BulFrame Design

BulFrame is a system whose functionality is designed for the definition and description of Conceptual frames. The functionality is divided into three main modules: (a) definition of the abstract structure; (b) description of particular Conceptual frames based on the defined structure; and (c) public access to the Conceptual frames, with a read-only restriction.

3.1 Abstract structure

The abstract structure of the system provides a complete set of components and operations for setting up any hierarchical structure. Moreover, it can be changed over time taking into account the risk of information loss after certain operations.

3.1.1 Objects

The abstract structure has only one object type, which is defined by attributes related to the object with system internal relations. Thus, the difference between the object and the attribute is that the object does not have a parent, or in other words, it is

[2] https://framenet.icsi.berkeley.edu/fndrupal/frameIndex

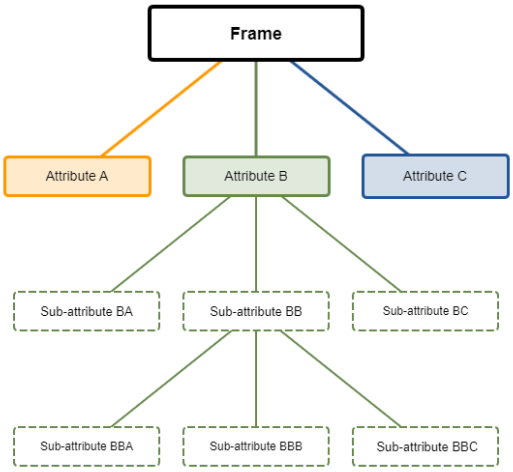


Figure 1: Abstract Structure of the Object FRAME

always a root, while the attribute is always related either directly to the object or to the other attribute (the attribute can never be a root). In **BulFrame** implementation of the abstract structure the object has two instances: **WORD** and **FRAME**. The object instances are linked with specially defined relations by means of the following formal properties: equivalence, reflexivity, transitivity. The relations might be: FRAME to FRAME, WORD to WORD and FRAME to WORD.

The abstract structure is represented as a strictly hierarchical structure. This is illustrated at Figure 1, in which the root is the object FRAME and the other nodes represent the frame attributes. The hierarchical organization of the abstract structure is achieved in two ways: by nesting of object attributes and by encoding taxonomy relations between objects.

3.1.2 Attributes

The attributes have a uniform structure represented by the pair **key : value**, where the key is the name of the attribute and the value determines the way the value is defined: directly by a value or by a sub-attribute. Defining the attribute value as a sub-attribute forms complex (nested) attributes.

Each attribute, as a separate element of the hierarchical structure, is defined in tables from the database as shown in Figure 2.

The value types that are supported by the framework are: text, number, relation and sub-attribute (for complex attributes). In addition to defining the value type of the attributes, the framework also provides the opportunity to define the type of the visual component with which the corresponding value has to be represented. The supported components are:

- for the value **text**: a text box, a text box with autocomplete function based on the existing values for the same attribute, a drop-down with a single select option based on a predefined list of values and a drop-down with multiple select options based on a predefined list of values;
- for the value **number**: a numeric box;
- for the value **relation**: a combo-box based on the predefined FRAMES and WORDS.

Table 1 contains the general information about the attributes (the combinations of possible values and their interpretation). In addition to the name and the value of the attributes, there are some restrictions (minimal occurrences, maximal occurrences) that determine whether the attribute is mandatory and how many times it can be repeated in the frame description.

Attribute	**MIN**	**MAX**
Value	Meaning	Meaning
Null	Not allowed	Unlimited
Digit X	At least X	Not more than X

Table 1: Definition of attributes, where MIN states for minimum value occurrences, and MAX – for maximum.

Additional elements describing the attributes are:

- **position** – associates the attributes and their parents in the user interface;
- **import/export** – the name of the XML / JSON element that is responsible for the data import/export;
- **code** – a system element enabling the implementation of functionality linked with a specific attribute;
- **parent object type** – a system element ensuring the hierarchical structure (parent_obj_type_id).

For the value **text**, the component drop-down selection with single/multiple select options requires the definition of the list of possible values. For the value **relation**, the specific relations have to be defined. The reference to a relation is defined by the **relation type** (relation_type_id), which is constituted by several components:

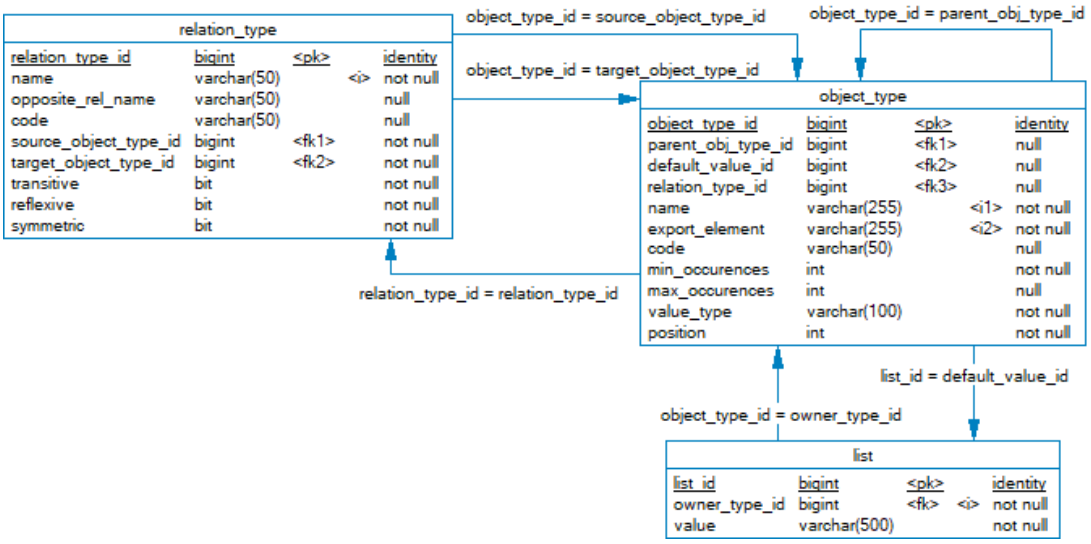


Figure 2: **BulFrame** database

- **name** – the name of the relation;
- **reverse relation name** – the name of the reverse relation (reverse_rel_name), if any;
- **source / target object** – the type of objects that are linked by the relation: either WORD or FRAME (source / target_object_type_id);
- **relation properties** – transitive, reflexive, symmetric.

Figure 3 in the Appendix A shows the definition of the Conceptual frame structure within the **BulFrame** system.

The **BulFrame** allows the following actions: import verbs and Conceptual frames, edit existing entries and delete verbs with no associated Conceptual frames and Conceptual frames with no associated verbs.

4 Linguistic interpretation of Conceptual frames

Conceptual frames represent the lexical meaning and morphological features of (Bulgarian) verbs which actually predict the syntactic realization and semantic combinability of their arguments (core frame elements), which are also the subject of description (Koeva, 2021). The structure of a Conceptual frame consists of the following sections: Lexical, Morphological, and Frame (Syntactic and Semantic) section.

4.1 Lexical section

The **Lexical section** embraces the information about the verb lemmas (object WORD). The unique interpretation of a verb is ensured by its WordNet ILI (Inter-Lingual-Index) number, WordNet sense number and definition. The WordNet ILI has two purposes: it links the synonyms in a synset and shows the mapping to the respective synset (concept) from the Princeton WordNet (Vossen et al., 1998).

The **Lexical section** includes: verb lemma (literal), whether the verb is a multiword expression or not, part of speech, WordNet ILI to which the verb belongs, sense number, sense definition, synset semantic class, stylistic or usage note, and relations with literals from other synsets.

The verb multiverb expressions can be classified mostly as non-fixed lexicalized expressions: **reflexiva tantum se**: smeya se ‘laugh’, izpravyam se ‘stand up’; **reflexiva tantum si**: vaobrazyabvam si ’imagine’; **reciproca tantum se**: sastezavam se ‘compete’; **reciproca tantum si**: govorya si ‘talk to oneself’; **accusativa tantum**: marzi me ‘feel lazy’; **dativa tantum**: hrumna mi ‘it occurred to me’; **reflexiva dativa tantum** gadi mi se ‘feel sick’; **with obligatory preposition(al phrase)**: privezhdam v sila ‘enforce’; **with obligatory noun (phrase)**: podavam zhalba ‘file a complaint’, davam si smetka ‘realise’; **with obligatory adverb(ial phrase)**; stoya nastrana ‘stand aside’.

Perfective and imperfective verbs in Bulgarian express different meanings, although the verb aspect pairs are closely related, for example the

verbs (rodya) 'give birth' and (razhdam) 'am giving birth'. The definition should describe the meaning in a way that uniquely distinguishes a verb from other senses of the same word; the definition also reflects the morphological features of the verbs (for example, the limited person paradigm as third-person, impersonal and plural personal) and the lexical-grammatical category aspect.

The verbs in the (Princeton) WordNet are organized into semantic classes (primitives): generic concepts, perceived as unique roots (beginners) of separate hierarchies, and the verbs belonging to the hierarchies are subsumed under the common **semantic class**: bodily care and functions, change, cognition, communication, competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction, weather verbs, state (Fellbaum, 1990). One and the same semantic class might be assigned to many verb roots and as close to the root the concept is as abstract or general its meaning is.

The note to the literal can express: the belonging to **non-standard lexis** – a dialectal word, a folk word, a word with an undesired use; use in a **specific functional style** – a colloquial word; a poetic word; a literary word; term; the **historical period of use** – an obsolete word; a new word; the **expressive properties** of the literals – a word with pejorative meaning; the **frequency of use** of the literals – a rare word; the **nuances in the use** of the literals – a figuratively used word. It should be pointed out, however, that stylistic marking usually excludes words from the core vocabulary, so although the information is intended to be transferred from WordNet, it is not expected that there will be many such cases, so far among the 639 verbs described with Conceptual frames (as of May 2022), only 4 have been marked as belonging to colloquial speech.

4.1.1 Selection of Verbs

The Bulgarian verbs included in the database of Conceptual frames are selected according to several criteria.

• Presence in **Age of Acquisition** (AoA) test – the school level at which a word (meaning of a word) must be studied or mastered. The resource includes a list of 44,000 entries (31,000 words and compound words; not only verbs) compiled by Dale and O'Rourke's Living Word Vocabulary (Dale and O'Rourke, 1981) and supplemented by estimates from other authors (Goodman et al., 2008; Morrison et al., 1997). For example, the Age of Acquisition ratings are a self-esteem given by adults (mostly students) about the age at which they learned a word, which is also further adjusted by other assessments and experiments (Kuperman et al., 2012).

• **Root distance** – the distance of the synset to the root of the local tree (the hierarchical substructure in Wordnet in which the corresponding synonym set is included). The distance is represented by the number of nodes (synsets) between the node in which the corresponding verb is located and the respective root, a node with an abstract meaning in WordNet.

• Presence in **Base concepts** – targeting maximum overlap and compatibility across wordnets of different languages (Vossen et al., 1998). 1,024 Base Concepts are identified on the basis of English, Dutch, Spanish and Italian along the following criteria: high position in the semantic hierarchy and maximum number of relations with other concepts in the WordNet. New Base Concepts have been added of second and third batch on the basis of data from Bulgarian, Greek, Romanian, Serbian and Turkish (Tufiş et al., 2004) and the first batch of Base Concepts has been expanded to 4,689. The following additional criteria were used to identify the main concepts of the second and third batch: a) the most common words in large representative corpora and b) the hyperonyms of already selected synsets to the root of the corresponding local tree.

• Relative **frequency** – represents a) frequency of verbs in the Bulgarian National Corpus[3] (in the whole corpus, in fiction texts and in news); and b) frequency of verbs in Bulgarian textbooks from 1st to 4th grade. The frequency is calculated at the level of lemma; however, some noise is left due to lack of sense disambiguation.

The presented measures were evaluated by experts in order to select a proper set of basic verb vocabulary for Bulgarian:

– If the following criteria are fulfilled: the verb is part of the AoA, the distance to the root is 0 or 1 and the verb is a member of the Base concepts (batch 1 or batch 2), the verb, accompanied with its sense number, ILI record and definition, is selected.

– If the verb is present in the AoA, but the other two criteria are not met, the expert judges according to the frequency of use and his/her personal intuition.

[3] https://search.dcl.bas.bg

– If the verb is not present in the AoA, the other criteria are used in the following order: root distance, member of the Base concept lists, frequency of use.

– The following principles have been also adopted during the selection: if a perfect verb is selected, the corresponding imperfect verb is also included; secondary imperfect verbs are not selected (at the moment).

4.1.2 Semantic Relations

The semantic relations are inherited from WordNet and inserted in the database. Taxonomic relations are: inverse and transitive (**is hypernym of** and **has a hypernym**, **has a troponym** and **is a troponym of**); meronymic relations are also inverse and transitive (**has subevent** and **is subevent of**). Non-hierarchical relations are: symmetric, reflexive, transitive and Euclidean (**synonymy**), symmetric, irreflexive and non-transitive (**antonymy**); symmetric, irreflexive and Euclidean (**also see, verb group**). The relations in WordNet are defined between synsets. As the basic unit in our system is the verb (WORD) and in the Bulgarian wordnet verbs with a different lexical aspect are grouped in one and the same synset, the following rules are implemented while inheriting the WordNet relations (Table 2).

Verb	Hypernym	VerbG	Antonym
1: Imperfect	1,2	All	1,2
2: ImperfT	1,2	All	1,2
3: Perfect	3,4	All	3,4
4: PerfT	3,4	All	3,4

Table 2: Verb to verb semantic relations, ImperfT stands for Imperfectiva tantum, PerfT – for Perfectiva tantum, VerbG – for Verb group.

4.2 Morphological section

A morphological classification of a target verb is necessary because the grammatical and morphosyntactic features determine in some cases the syntactic structures associated with a given word. We have distinguished four groups of grammatical subclasses of Bulgarian verbs depending on the subject: **personal**, **impersonal**: zazoryava se 'it dawns', **third personal singular and plural**: rekata se vliva v moreto 'the river flows into the sea', and **plural personal**: sabirame se okolo masata 'we gather around the table'.

The different meaning of **verb aspect pairs** is reflected at both the morphological and the syntactic levels: the paradigms of the perfective and imperfective verbs are different – perfective verbs do not have the so-called independent present tense, and they do not form either present participles (agentive and adverbial) or negative imperative forms; the derivational potential of the perfective and imperfective verbs is different – perfective verbs do not form some types of deverbal nouns and some nouns denoting professions; perfectivity is directly related to the syntactic realization of obligatory complements – direct objects of perfective verbs cannot remain implicit and perfective verbs cannot be complements of phase predicates; perfectivity is also directly related to the possibility for different types of verb diathesis: perfective verbs do not form middles, optatives or impersonals (Koeva, 2010).

In the Bulgarian WordNet verb aspect pairs are included in one and the same synsets, although the perfective and imperfective members of a pair are not cognitive synonymous, and as a consequence only some of the literals are translation equivalents to the respective synonyms in English. For the differentiation of verbs of different aspect, a literal note is attached to each verb indicating its aspect: **perfective verb**: (zapeya) 'start singing'; **imperfective verb**: (zapyavam) 'sing off'; a **simultaneously perfective and imperfective verb**: (pensioniram) 'retire'; an **imperfective verb with no perfective equivalent**: (vali) 'it rains'; a **perfective verb with no imperfective equivalent**: (povyarvam) 'get to believe'. The values of the category verb aspect are transferred directly from WordNet.

Verbs are also classified according to their transitivity.

4.3 Frame section

One part of the elements in the **Frame section** are inherited from the Berkley FrameNet, another part is constructed in compliance with the FrameNet organisation and yet another one is specific for the organisation of the Conceptual frames. The FrameNet related parts are: frame name, frame definition, frame-to-frame relations, and frame elements with their names, status (core, non-core and extra-thematic) and definition.

Several types of frame-to-frame relations are defined, of which for the definition of the Conceptual

frames the important ones are: **Inheritance** (an is-a relation, the child frame is a subtype of the parent frame), **Using** (the child frame presupposes the parent frame as background); **Inchoative of** and **Causative of** (Ruppenhofer et al., 2016). Inheritance is the strongest relation between frames corresponding to an **is-a** relation in many ontologies. The basic idea of the inheritance relation is that each semantic fact about the parent must correspond to an equally specific or more specific fact about the child (Ruppenhofer et al., 2016). The origin of the information is marked: inherited form FrameNet; from FrameNet with modifications; completely new information.

So far, 104 different semantic frames were used as basic structures for defining Conceptual frames. 105 unique frame elements were used, among which the most frequently selected are: **Agent** – 175 instances, **Experiencer** – 81 instances, **Cause** – 66 instances, **Stimulus** – 57 instances and so on. Together with the frame elements that can be encountered in different semantic frames, there are 30 cases of rare use of a particular frame element – 1 or 2 times. For example, frame elements **Intervention**, **Medical condition**, **Medical professional** and **Result** are so far selected only once.

4.3.1 Frame element Syntactic Structure

The phrases that express the frame elements may be **obligatorily explicit** (in rare cases in Bulgarian) or **non-explicit**, which means that the potential for a syntactic realization of the phrase exists, but its explicitness is not mandatory because it is understood from the context in a broader sense (verb morphology, preceding text, extralinguistic information, etc.), a special case is **pronominal drop** in the subject position.

The **syntactical phrases** that can be candidates for arguments in Bulgarian are: **NP** (noun phrase), **PP** (preposition phrase), **AdvP** (adverb phrase), **S** (sentence), **SC** (small clause), **ACCCL** (obligatory accusative clitic), **DATCL** (obligatory dative clitic). For a single verb with a unique meaning, there might be more than one combination of obligatory environments. Each personal verb incorporates an argument – a noun phrase (NP) or a sentence (S) that are realized as the subject in the sentence. The subject may not be explicitly stated – with personal verbs the information for person and number of the omitted pronoun subject is expressed by the verb inflexion.

The frame elements related to the subject of Bulgarian verbs can be characterized as follows: (a) with an explicitly or implicitly expressed subject with a full paradigm of the category of person; (b) with an explicitly or implicitly expressed third-person subject; (c) with no subjective argument. The frame elements related to the complements of Bulgarian verbs can be classified as follows: (a) with a single NP complement; (b) with an NP complement and an S complement; (c) with an NP complement and PP complements, regardless of their number; (d) with an NP complement, PP complements, regardless of their number, and an S complement; (e) with PP complements, regardless of their number; (f) with PP complements, regardless of their number, and an S complement; (g) with an S complement; (h) with an AdvP predicate modifier; (i) With an SC (small clause) NP argument; (j) with an SC (small clause) PP argument; (k) with an SC (small clause) AP argument; (l) with no complements.

The **syntactic functions** (names of syntactic positions taken from traditional grammar) are subject, direct object, indirect object, adverbial, subject clause, object clause, adverbial clause and small clause. The syntactic structure is described by information about the phrases: explicitness (check box), syntactic category (check box) and syntactic function (check box).

4.3.2 Frame element Semantic Structure

FrameNet allows for the characterization of 'role fillers' by semantic types of frame elements, which ought to be broadly constant across uses (Ruppenhofer et al., 2016). However, not all frame elements are supplied with a semantic type or the semantic types are too general, and in some cases, they do not show the actual restrictions for lexical combinations. For example, the following frame elements of the semantic frame **Experiencer focused emotion** are equipped with 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].

We call selectivity the **semantic restrictions** to a given argument in a certain context selectivity. Due to the fact that selective restrictions act between a concrete predicate and the arguments that belong to it, they can be different for each separate case. The

most general semantic classification distinguishes among **abstract** and **concrete** nouns. On their part, concrete nouns can be **animate** or **inanimate**. Animate nouns may be classified as **persons** and **non-persons**, **persons** as **agents** or **experiencers**. This classification tree is convenient but too shallow to represent the selective restrictions that act with verbs and nouns. Besides the general cases, there may also be cases where concrete restrictions are required, as for example **liquid, food**, etc. That is why we include the link to the top most synset (or the conjunction/disjunction of top most synsets) taken from the Bulgarian WordNet. The top most synset should dominate all appropriate synsets for a given syntactic slot, i.e., **liquid is a hypernym of water, milk, liquor**, etc.

The semantic classes of nouns in WordNet might be subdivided into a set of semantic subclasses. For example, within the semantic class **[Food]** we can introduce the sub-class of **[Beverage]** for nouns associated with verbs like **stir, sip, drink, lap**, etc. Such representation aims to specify the organization of concepts into an **ontological structure** which allows inheritance between the semantic classes down the hierarchy and ensures more precise specification of verb – noun compatibility.

One potential to extend the repository of WordNet semantic classes is to map the WordNet synsets to an existing hierarchy of semantic types, such as the CPA types (Hanks, 2004). The semantic types (e.g. **[Human], [Animal], [Part]**, etc.) refer to properties which can be expressed by words regularly found to participate in particular verb pattern positions (Hanks 2012: 57–59). In other words, the semantic types state the semantic preferences of verbs that determine the sets of nouns and noun phrases that are normally found in a particular clause role depending on a verb predicate.

Some verb patterns may contain very general preferences, i.e., the semantic type **[Anything]**, while others impose preferences for a limited set of lexical units grouped into more particular semantic types. For example, some verbs are associated with nouns characterised as **[Body part]**. However, the verb **shampoo** is associated with a more particular semantic type **[Hair]**; the same is referred to the verb **nod**, which is associated with the type **[Head]**, etc. Some verb patterns require a very small set of lexical units for a particular slot and in this case, a semantic type is not formulated; instead, the concrete lexical units are listed in the verb pattern. The expansion of WordNet semantic classes with CPA semantic types is performed manually by matching the CPA semantic types with WordNet synsets and choosing the most appropriate ones (Koeva et al., 2018).

The 253 CPA semantic types are manually mapped onto the respective WordNet concepts (synsets) as follows: 199 semantic types are mapped directly to one concept, i.e., **[Permission]** is mapped to **permission** 'approval to do something, semantic class noun communication'; **[Dispute]** is mapped to **disagreement** 'the speech act of disagreeing or arguing or disputing', semantic class noun communication; 39 semantic types are mapped to two WordNet concepts, i.e., **[Route]** is mapped to **road; route** 'an open way (generally public) for travel or transportation, semantic class noun artefact, and **path; route** 'an established line of travel or access', and semantic class noun location, and so on. Automatic mapping of hyponym synsets to the inherited semantic types was performed. In the cases where a semantic type and its ancestor were both mapped to the same synset, the ancestor was removed. 82,114 WordNet noun synsets were mapped to the 253 semantic types of the CPA ontology, resulting in 172,991 mappings. As there are multiple hypernymy relations in WordNet, some of the inheritances are not correct; furthermore, the inheritance by multiple hypernyms will be manually evaluated, and if necessary, adjusted (Koeva et al., 2018).

Some of the initially selected classes were not chosen as dominant classes for nouns compatible with particular verbs, for example the class **[Plant]** (eng-30-00017222-n), the class **[Abstract object]** (eng-30-00019128-n), and so on. This obviously is a consequence of the selections of the verbs. On the other hand, 84 unique selective restrictions were used identifiable by a representative noun and its ILI number. Some new classes were introduced (28 altogether, which constitutes 35,7 % of the total number of classes used so far. For example, new classes are: **[Text]** (eng-30-06387980-n), **[Examination]** (eng-30-07197021-n), **[Fire]** (eng-30-07302836-n), and so on. Still, the abstract notions show more instances in the dataset: **[Person]** — selected 850 times, **[Entity** — selected 249 times, **[Object]** — selected 175 times, **[Physical object]** — selected 170 times and so on.

The concrete prepositions for a given frame element expressed with a prepositional phrase are to

be selected from a list box. The same holds for frame elements that express the obligatory noun (phrase) or adverb (phrase). The types of subordinate clauses depend on the method of linking – interrogative pronouns or conjunctions, thus the respective linking phrase or complementizers are to be selected (more than one choice is permissible).

5 Related work

FrameNet is the most famous language resource that contains lexical and conceptual knowledge (Ruppenhofer et al., 2016). FrameNet can be viewed as a semantic network (or a set of small semantic nets), whose nodes indicate the semantic frames and whose arcs represent semantic relations between frames. For the purposes of the presented research, the following information is employed: the sets of verb lexical units related to semantic frames, the inheritance relation between semantic frames, and the description of core and peripheral frame elements and their semantic types. The FramNet annotation is mostly used for automatic role labelling while we offer the definition of noun sets compatible with verbs from a particular Conceptual frame (and such approach offers much more training data for automatic processing). FrameNets for languages other than English are being developed, including for Bulgarian.

VerbAtlas is a relatively new, hand-crafted lexical-semantic resource, whose goal is to bring together WordNet verbal synsets into semantically-coherent frames (Di Fabio et al., 2019). The frames define a common, prototypical argument structure, while at the same time provide new concept-specific information. VerbAtlas also offers an explicit, cross-frame set of semantic roles linked to selectional preferences expressed in terms of WordNet synsets, and is the first resource enriched with semantic information about implicit, shadow and default arguments. The main difference between the VerbAtlas and the presented framework is that the VerbAtlas selectional preferences are too general, similarly to the semantic types of core elements in FrameNet, in comparison to the extensive semantic information provided within the BulFrame to ensure accurate noun-to-verb compatibility.

Some efforts to describe Bulgarian frame lexicon were also shown, and we believe our work draws on the best approaches in the field.

6 Conclusion and Future work

The presentation of Conceptual frames of Bulgarian verbs provides opportunities for the enrichment of already existing resources (Wordnet and Framenet) with new semantic information (in the direction of completeness and structural expansion), developing a detailed ontology of the semantic classes of nouns and linking it to the hierarchical structure of WordNet and the frame elements of FrameNet.

The main characteristic of the approach we have taken is the manner of connecting FrameNet and WordNet – not by assigning frames to synsets, i.e., not in the usual way, but by showing which WordNet subtrees are suitable to take one or another syntactic position in which a frame element is realized. The morpho-syntactic features that are specific for Bulgarian are shown in detail; selective restrictions are specified so that the resource can be used for automatic prediction of semantic roles in any text.

As future work, we plan to take full advantage of the semantic features available in BulFrame, such as wide-coverage selectional preferences and verb level grammatical information, by employing them in semantic role labelling tasks.

Acknowledgments

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

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley FrameNet project. In *Proceedings of the 17th International Conference on Computational Linguistics*, volume 1, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics.

Edgar Dale and Joseph O'Rourke. 1981. *The Living Word Vocabulary: A National Vocabulary Inventory*. World Book-Childcraft International, Chicago.

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

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

Charles J. Fillmore. 1982. *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Charles J. Fillmore and Collin F. Baker. 2001. Frame Semantics for Text Understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop*, Pittsburgh. NAACL, NAACL.

Judith C. Goodman, Philip S. Dale, and Ping Li. 2008. Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35(3):515–531.

Patrick Hanks. 2004. Corpus pattern analysis. In *Proceedings of the 11th EURALEX International Congress*, pages 87–97, Lorient, France. Université de Bretagne-Sud, Faculté des lettres et des sciences humaines.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2007. A large-scale classification of English verbs. *Language Resources and Evaluation*.

Svetla Koeva. 2010. *Bulgarian FrameNet*. Prof. M. Drinov Academic Publishing House, Sofia.

Svetla Koeva. 2020. Semantic Relations and Conceptual Frames. In Svetla Koeva, editor, *Towards a Semantic Network Enriched with a Variety of Semantic Relations*, pages 7–20. Sofia: Professor Marin Drinov Publishing House of BAS.

Svetla Koeva. 2021. Towards Expanding WordNet with Conceptual Frames. In *Proceedings of the 11th Global Wordnet Conference*, pages 182–191, University of South Africa (UNISA). Global Wordnet Association.

Svetla Koeva, Cvetana Dimitrova, Valentina Stefanova, and Dimitar Hristov. 2018. Mapping WordNet concepts with CPA ontology. In *Proceedings of the 9th Global Wordnet Conference*, pages 69–76, Nanyang Technological University (NTU), Singapore. Global Wordnet Association.

Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, 44(4):978–990.

Beth Levin. 1993. *English Verb Classes and Alternations A Preliminary Investigation*. University of Chicago Press, Chicago and London.

George A. Miller, Richard Beckwith, Christiane. Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International journal of lexicography*, 3(4):235–244.

Catriona M. Morrison, Tameron D. Chappell, and Andrew W. Ellis. 1997. Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology Section A*, 50(3):528–559.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193(0):217 – 250.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

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

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A Core of Semantic Knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 697–706, New York, NY, USA. ACM.

Dan Tufiş, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science and Technology*, 7(1-2):9–43.

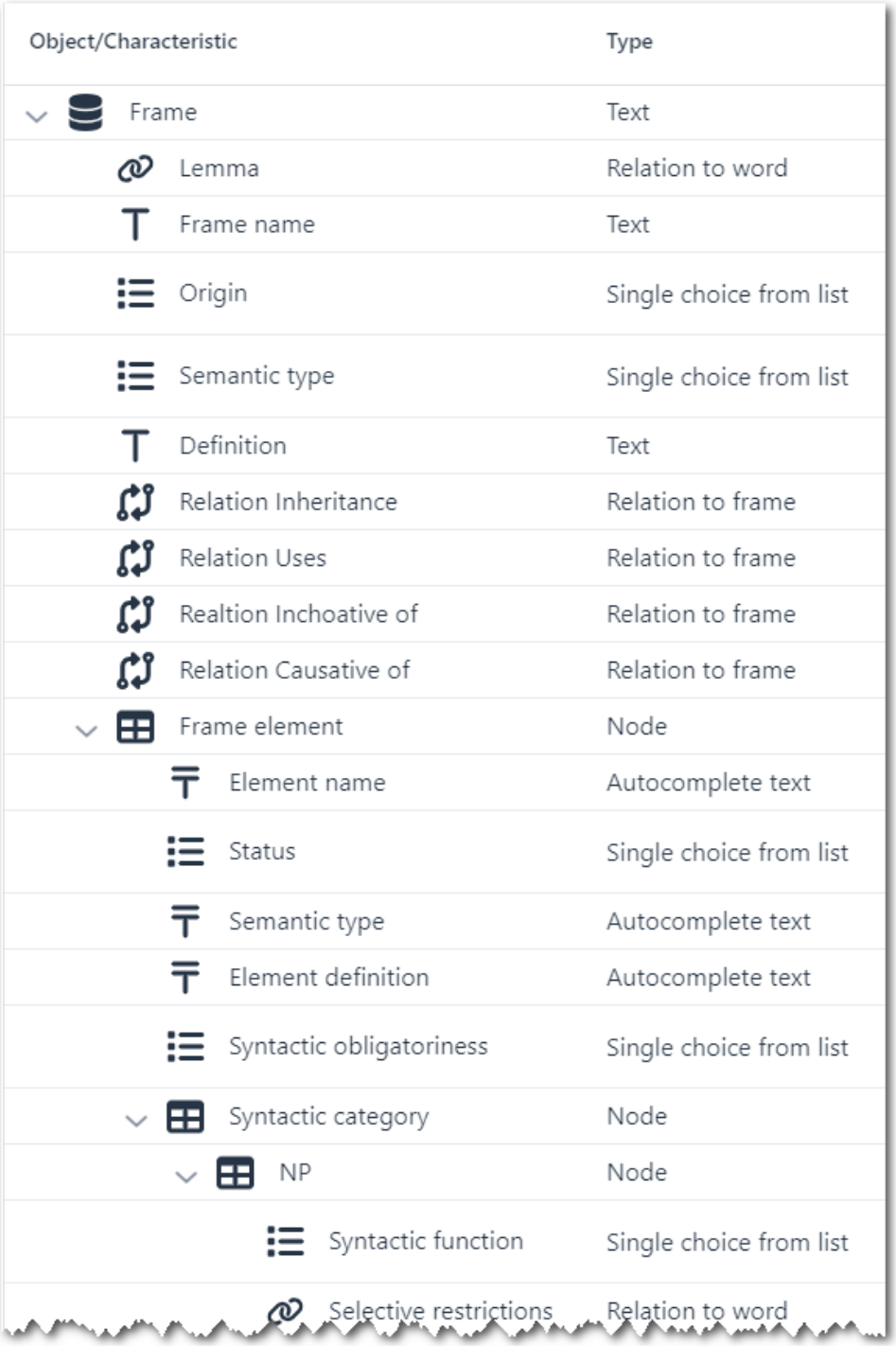
Zdenka Uresova, Eva Fucikova, Eva Hajicova, and Jan Hajic. 2020. SynSemClass linked lexicon: Mapping synonymy between languages. In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, pages 10–19, Marseille, France. European Language Resources Association.

Piek Vossen, Laura Bloksma, Horacio Rodriguez, Salvador Climent, Nicoletta Calzolari, Adriana Roventini, Francesca Bertagna, Antonietta Alonge, and Wim Peters. 1998. The EuroWordNet base concepts and top-ontology. Technical report, GPLN - Grup de Processament del Llenguatge Natural and TALP - Centre de Tecnologies i Aplicacions del Llenguatge i la Parla.

Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin Sameer Pradhan Lance Ramshaw, and Nianwen Xue. 2011. Ontonotes: A large training corpus for enhanced processing. *Joseph Olive, Caitlin Christianson, andJohn McCary, editors, Handbook of Natural LanguageProcessing and Machine Translation: DARPA GlobalAutonomous Language Exploitation*.

Appendix A Conceptual frame structure (some parts of the structure are not presented)

.



Object/Characteristic	Type
Frame	Text
Lemma	Relation to word
Frame name	Text
Origin	Single choice from list
Semantic type	Single choice from list
Definition	Text
Relation Inheritance	Relation to frame
Relation Uses	Relation to frame
Realtion Inchoative of	Relation to frame
Relation Causative of	Relation to frame
Frame element	Node
Element name	Autocomplete text
Status	Single choice from list
Semantic type	Autocomplete text
Element definition	Autocomplete text
Syntactic obligatoriness	Single choice from list
Syntactic category	Node
NP	Node
Syntactic function	Single choice from list
Selective restrictions	Relation to word

Figure 3: Conceptual frame structure