EXTENDED ONTOLOGY OF NOUN SEMANTIC CLASSES

Svetla Koeva

Institute for Bulgarian Language „Prof. Lyubomir Andreychin“ – Bulgarian Academy of Sciences

svetla@dcl.bas.bg

Abstract. The primary aim of this study is to propose an effective approach for characterising the frame elements of conceptual frames within the Bulgarian FrameNet using noun classes. This approach will facilitate the prediction of comprehensive and semantically valid combinations between verbal lexical units that evoke conceptual frames and appropriate nouns. The study provides a concise overview of the semantic classifications of nouns in WordNet, Corpus Pattern Analysis, FrameNet, and VerbAtlas, emphasising their significance in predicting verb-noun compatibility. The structure of the Bulgarian FrameNet (BulFrame) is presented, containing components adapted from FrameNet for Bulgarian along with a substantial amount of lexical, morphological, syntactic, and semantic information specific to the Bulgarian language. One distinctive feature of the Bulgarian FrameNet is the specification of noun classes, indicating appropriate nouns for the lexical realisation of frame elements. By aligning synonym sets from WordNet with the semantic types of Corpus Pattern Analysis and FrameNet, as well as with the selective preferences of VerbAtlas, the foundational structure of the Extended Ontology of Noun Semantic Classes is established. The Ontology concepts are linked, not exclusively, to synonym sets from WordNet and thereby to sets of nouns suitable for combination with verbal lexical units that evoke the corresponding conceptual frames. The contribution of this development lies in detailing the steps for selecting semantic classes (concepts) and constructing the structure of the Extended Ontology of Noun Semantic Classes.

Keywords: *semantic class, semantic type, selectional preferences, conceptual frame, frame element*

* Svetla Koeva. Extended ontology of noun semantic classes. – In: Koeva, Sv. (Ed.) *Semantic studies. Syntactic and semantic description of Bulgarian verbs*. Sofia: BAS Press „Prof. Marin Drinov“, pp. 12 – 45. https://doi.org/10.7546/SemanticStudies2024.02

1. Introduction

There are different (ontological) representations of word classes, each with its own set of concepts and degree of complexity. For example, nouns and verbs in WordNet are categorised into semantic classes (Miller 1990/1993: 16 – 17), nouns in Corpus Pattern Analysis are classified into semantic types (Hanks 2012: 57 – 17), frame elements in FrameNet are often specified according to the semantic types of nouns (Ruppenhofer et al. 2016: 86 – 87), and semantic roles in VerbAtlas are supplied with selectional preferences (Di Fabio et al., 2019: 630).

The objective of this study is to propose an effective approach for characterising the conceptual frame elements in the Bulgarian FrameNet using noun classes. This method is designed to streamline the prediction of comprehensive and semantically sound pairings between verbal lexical units, which evoke conceptual frames, and appropriate nouns.

The study offers a succinct overview of the semantic classifications of nouns in WordNet, Corpus Pattern Analysis, FrameNet, and VerbAtlas, underscoring their pivotal role in anticipating verb-noun compatibility. Furthermore, as part of a broader investigation, it sheds light on the inadequacies of existing noun classifications, which are often unsuitable for effectively illustrating syntagmatic verb-noun combinations.

The structure of the Bulgarian FrameNet (BulFrame) is introduced, comprising components tailored within FrameNet, complemented by a significant array of lexical, morphological, syntactic, and semantic information unique to the Bulgarian language. A notable aspect of the Bulgarian FrameNet is its delineation of noun classes, providing predictions on the lexical realisation of frame elements.

Drawing from existing language resources, the study offers a comprehensive approach that combines appropriate ontological representations of noun classes and extends them with additional classifications. By aligning synonym sets from WordNet with the semantic types of Corpus Pattern Analysis and FrameNet, as well as with the selective preferences of VerbAtlas, the foundational structure of the Extended Ontology of Noun Semantic Classes is established. The Ontology concepts are linked, not exclusively, to synonym sets from WordNet and thereby to sets of nouns suitable for combination with verbal lexical units that evoke the corresponding conceptual frames. The contribution of this development lies in detailing the steps for selecting semantic classes (concepts) and constructing the structure of the Extended Ontology of Noun Semantic Classes.

The paper is structured as follows: the subsequent four sections offer an overview of the semantic classifications of nouns in WordNet, Corpus Pattern Analysis, FrameNet, and Bulgarian FrameNet, focusing on their relevance to verb-noun compatibility. Section six provides a brief discussion of appropriate resources presenting groupings of nouns that can co-occur with specific verbs and outlines our approach to identifying noun classes with appropriate candidates for particular verb-noun compatibilities. The initial development of the Extended

Ontology of Semantic Classes of Nouns entails mapping relevant ontologies to the WordNet noun hierarchy, facilitating the identification of concepts indicating appropriate noun fillers for specific frame elements.

The contribution of this investigation lies in the detailed and systematic steps employed to establish a suitable set of semantic classes organised within the Extended Ontology of Semantic Classes of Nouns. Additionally, the mapping with Corpus Pattern Analysis semantic types, VerbAtlas selectional preferences, and FrameNet semantic types highlights the importance of high-rated concepts. The alignment of these resources emphasises the relevance of certain semantic classes. Furthermore, it illustrates how the Ontology of Semantic Classes of Nouns remains open to augmentation.

2. FrameNet, Semantic Frames and Semantic types

FrameNet is based on the theory of Frame Semantics (Fillmore 1976, 1982; Fillmore, Baker 2010). The core concept of Frame Semantics revolves around describing word meanings in relation to semantic frames, which are seen as schematic representations of conceptual structures within a speech community (Fillmore et al. 2003: 235). Semantic frames, more succinctly, represent speakers' understanding of the underlying situations or states of affairs shaping the meanings of lexical items (Fillmore 2007: 130). When a lexical unit evokes a frame, it invokes the associated conceptual structure, while a valency description of a specific lexical unit outlines how the semantic valences are expressed within sentences constructed around the frame-bearing unit (Fillmore 2007: 131).

FrameNet functions as a repository of semantic frames and valency patterns associated with lexical units, offering conceptual-semantic and syntactic descriptions derived from annotated examples. Frame Semantics aligns target lexical units with linguistic and conceptual details. Linguistic facets encompass definitions, semantic classes, and syntactic patterns for frame elements, while conceptual aspects involve situational descriptions, participant delineations, and inter-frame relations (Sikos, Pado 2018: 2).

2.1 FrameNet semantic frames

In FrameNet, the **semantic frame** comprises several key components: the frame name, an informal **definition** of the represented situation, an optional specification for the semantic type, a set of associated **frame elements** (including core, core-unexpressed, and non-core elements like peripheral and extrathematic), specifications for **relations between frame elements** and **frame-to-frame relations** if applicable, and the **lexical units** evoking the frame. Information on frame elements includes their names, informal definitions, optional semantic types, and illustrative examples. Details for lexical units encompass definitions, optional semantic types, examples supplied with **annotations covering frame elements, grammatical categories, and functions**.

While certain **frame element names** may align with semantic roles, these names primarily function as mnemonic aids (Fillmore et al. 2003: 237). **Frame element definitions** are statements that convey a frame element's semantics with regard to the target lexical unit (and, in some cases, to other frame elements as well).

Two criteria are employed in formulating semantic frames (Ruppenhofer et al. 2016: 11 – 17): a checklist of features and additional principles such as paraphrases and alternative answers to a question. The checklist of features encompasses:

- Consistent number and type of frame elements across all lexical units.
- Uniform set of stages and transitions (subevents) shared by lexical units. For example, while *decapitate* implies death, *shoot* merely implies firing and hitting.
- Uniform participant perspective, such as the buyer's or seller's viewpoint.
- Consistent interrelations between frame elements for all lexical units.
- Similar presuppositions, expectations, and concomitants of the target lexical units.
- Similar basic denotation among lexical units.
- Consistent pre-specifications given to frame elements by frame-evoking lexical units. For example, verbs like *crowd*, *flock*, *pour*, and *stream* are part of the **Mass motion** frame but not the **Self motion** frame, as they imply movement by a mass theme.

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 substituted for another while evoking the same frame and the same configuration of frame elements ((Ruppenhofer et al. 2016: 15 – 17). One and the same semantic frame can be evoked by synonyms, near synonyms, antonyms, derivationally related lexical units, hypernyms, or hyponyms (Koeva 2021: 183).

In FrameNet, frame elements are categorised based on their centrality within a given frame, delineating three levels: core, peripheral, and extrathematic. A **core frame element** is indispensable to the frame's central meaning (Fillmore 2007: 133) and embodies a conceptually essential component that distinguishes the frame from others (Ruppenhofer et al. 2016: 23). **Peripheral frame elements** denote notions like Time, Place, Manner, Means, and Degree, devoid of distinctiveness among frames and applicable to any semantically appropriate frame (Ruppenhofer et al. 2016: 24). **Extrathematic frame elements**, while present, aren't conceptually inherent to the frame; they relate to other abstract frames and contextualise events against the backdrop of another event (Fillmore 2007: 133). **Core-unexpressed frame elements** function as core elements but remain unexpressed in descendants of the frame; they're absorbed by lexical units in child frames, lacking individual representation (Ruppenhofer et al. 2016: 25).

Observations indicate that frame elements may not operate independently; certain groups behave akin to sets, termed **Core Sets**, wherein the presence of any member suffices to fulfil the predicator's semantic valence (Ruppenhofer et al. 2016: 25). For example, in motion frames, **Source**, **Path**, and **Goal** core frame elements constitute a **Core Set**, allowing for the appearance of one or two (rarely all three) frame elements in a sentence without violating the semantic structure. The relation **Requires** is applied when the presence of one core frame element necessitates the occurrence of another core frame element. Conversely, the relation **Excludes** emerges when one frame element from a set of conceptually related elements is present, precluding the presence of any other element from that group (Ruppenhofer et al. 2016: 26). For instance, in the **Attaching** frame, the frame elements **Goal** and **Item** are mutually required to complement each other, while excluding the presence of the frame element **Items**.

As it has been pointed out, the semantic and syntactic descriptions in FrameNet differ from other lexical resources in several ways (Fillmore 2007: 129), including: (1) relying on corpus evidence; (2) basing the semantic layer of valency on an understanding of the cognitive frames that motivate and underlie the meanings of each lexical unit; (3) recognising various kinds of discrepancies between lexical units on the semantic (functional) level and patterns of syntactic form; and (4) providing the means of assigning partial interpretations to valents that are conceptually present but are syntactically unexpressed.

2.2 FrameNet Semantic Types

Frames, frame elements, and lexical units are categorised by ontological semantic types. On frames, the semantic type indicates that every lexical unit of the frame could be labeled with an equally or more specific type. The **Clothing** frame, for example, has the semantic type [Artefact]. As a result, all of its lexical units (boot.n, cape.n, dress.n, and so on) denote artefacts (Loenneker-Rodman, Baker 2009: 422 – 423).

Semantic types for frame elements classify the type of filler that is expected to appear as the frame element (Ruppenhofer et al. 2016: 86). Not all frame elements (and frames) come with a specified semantic type, and in general, the semantic types tend to be too broad, lacking precision in conveying actual restrictions for 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 (Koeva 2021: 187).

The semantic types in FrameNet provide a high level of abstraction to express the multiple entities that can fill a frame element. A common semantic type in

FrameNet is [Sentient], which is associated with frame elements that typically represent humans or sentient entities. For example, in the semantic frame **Telling** (with a definition 'A **Speaker** addresses an **Addressee** with a **Message**, which can be indirectly referred to as a **Topic**. Instead of (or in addition to) a **Speaker**, a **Medium** may also be mentioned.'), both frame elements **Speaker** and **Addressee** have the semantic type [Sentient], although only nouns referring to humans can be the appropriate fillers when used non-metaphorically.

Most semantic (ontological) types in FrameNet can be directly mapped to WordNet synset nodes and ontologies. They are organised hierarchically, with inheritance relations (or "is-a" relations) between parent nodes and child nodes (Ruppenhofer et al. 2016: 86). At the top level, there are five distinct types: [Attribute], [Physical entity], [Event], [Group], and [Relations]. The remaining forty ontological semantic types are subsumed under one of these five top-level types. For example, [Speed] is a type of [Attribute]; [Human] is a type of [Sentient], which is a type of [Animate being], which is a type of [Living thing], which is a type of [Physical object], which is a type of [Physical entity]; [Activity], [Accomplishment] and [Achievement] are types of [Event], which is a type of [State of Affairs]; [Organisation] is a type of [Group]; [Source]. [Path] and [Goal] are types of [Locative relation] which is a type of [Relations], and so forth.

As evident, FrameNet semantic types offer a high level of abstraction and, in many instances, prove challenging to employ for predicting diverse context realisations.

3. WordNet, Sentence Frames and Semantic Classes (Primitives)

WordNet is a semantic network whose nodes host synonyms denoting different concepts and whose arcs, connecting the nodes, encode different types of relations (semantic: genus-kind, part-whole, cause-effect, etc.; morpho-semantic: agent-predicate, predicate-instrument, predicate-state, etc.; derivational; extralinguistic, i.e., membership to a thematic domain; inter-language, i.e., translation equivalents). The idea for organising the lexicon of a given language into a (lexico-)semantic network was first executed in the Princeton WordNet[1] (Miller et al. 1990/1993). Some of the fundamental ideas on which the WordNet is based encompass: a) the use of a semantic network that embraces taxonomies, meronomies, and non-hierarchical relations with clearly defined properties that allow for quick and easy automatic processing; b) a different organisation of the lexicon in comparison to traditional dictionaries where words are ordered alphabetically and the links among semantically related words (such as between sister hyponyms, between a whole and its parts, etc.) are not explicitly presented (Miller 1986).

---

1 https://wordnet.princeton.edu

3.1. WordNet sentence frames

WordNet presents one or more sentence frames for each verb synset to clarify the basic syntactic properties of verbs – the subcategorization features of verbs by indicating the types of sentences they can occur in (Fellbaum 1990/1993: 55). These sentence frames detail semantically (and syntactically) obligatory constituents and constraints on combinability. For example, the synset {read} 'interpret something that is written or printed' is associated with three sentence frames:

Somebody ----s
Somebody ----s something
Somebody ----s that CLAUSE
while synset {assure, tell} 'inform positively and with certainty and confidence' – with two:
Somebody ----s somebody of something
Somebody ----s that CLAUSE

Apart from the number of the syntactically obligatory constituents, minimal details concerning the specific combinability restrictions – that is, whether a particular sentence frame element can be realised as a human noun or not – and minimal syntactic details – that is, whether the element is realised as a noun, prepositional phrase, or clause – are provided.

3.2. WordNet semantic classes

WordNet categorises concepts and arranges them in a hierarchical structure using a series of semantic primitives (semantic classes) for nouns and verbs (Miller 1990/1993: 16; Fellbaum 1990/1993: 41). These semantic classes serve as the basic building blocks for capturing essential semantic distinctions and categorising words based on their meanings.

Nouns are categorised into twenty-five semantic classes: {**act**; action; activity} {**animal**; fauna}, {**artifact**}, {**attribute;** property}, {**body;** corpus}, {**cognition;** knowledge}, {**communication**}, {**event;** happening}, {**feeling**; emotion}, {**food**}, {**group**; collection}, {**location**; place}, {**motive**}, {**natural object**}, {**natural phenomenon**}, {**person**; human being}, {**plant**; flora}, {**possession**}, {**process**}, {**quantity**; amount}, {**relation**}, {**shape**}, {**state**; condition}, {**substance**}, {**time**} (Miller 1990/1993: 16). Verbs are classified into fifteen semantic classes: fourteen classes for events or actions (verbs of bodily care and functions, change, cognition, communications, competition, consumption, contact, creation, emotion, motion, perception, possession, social interaction, and weather verbs), and one class for verbs denoting states (Fellbaum 1990/1993: 57 – 61).

Nouns are categorised according to semantic classes, with a (relatively small) number of basic concepts selected and each treated as the unique starting point of a separate hierarchy. These hierarchies correspond to relatively independent semantic fields, each with its own vocabulary (Miller 1990/1993: 16). Noun-hypernym subtrees can be linked in a hierarchical structure to the most abstract root {entity} 'that which is perceived or known or inferred to have its own distinct existence (living or nonliving)'. There are over 550 verb subtrees in WordNet, and the same semantic class can occur in many subtrees. The verb subtrees can be artificially combined into an abstract root *eventuality*, not present in WordNet and comprising events, processes, states, changes, etc.

Although (at least for nouns) the idea is that the semantic classes in WordNet should be differentiated in order to develop multiple hierarchies organising words with different semantic classes, the WordNet structure includes nouns and verbs of different semantic classes within a single hypernymy tree. For example, for nouns, the synset {act; human action; human activity} 'something that people do or cause to happen' has the semantic class noun.act, while its hyponym {communication; communicating} 'the activity of communicating; the activity of conveying information' has the semantic class noun.communication. Among the verbs, {interact} 'act together or towards others or with others' has the semantic class verb.social, while its troponym {communicate; intercommunicate} 'transmit thoughts or feelings' has the semantic class verb.communication and the next level troponym {grimace; make a face; pull a face} 'contort the face to indicate a certain mental or emotional state' has the semantic class verb.body.

Further, the semantic class noun.communication applies to noun synsets, including {communication; communicating}, {language; linguistic communication}, {visual communication}, {email; e-mail; electronic mail}, {body language} and so on. The semantic class verb.communication specifies verb synsets such as {communicate; pass on; pass; pass along; put across}, {articulate; enunciate; vocalise}, {sign; signal; signalise}, {broadcast; air; transmit; beam; send}, etc. It is intuitively obvious that nouns and verbs referring to verbal and non-verbal communication belong to different semantic subclasses; the same applies to written communication, communication by signalling, communication by broadcasting, and so on. From this, it can be concluded that there is still much potential for additional specification of WordNet's semantic classes in order to capture the semantic compatibility of verbs and nouns.

Not all verbs marked with the semantic class verb.communication can be combined with the nouns of the class nun.communication as objects. Based on an analysis of the superordinates (hypernyms), it has already been observed that the nodes in the WordNet hierarchy do not (always) represent semantic classes, nor do those classes occupy specific slots in the verb argument structure (Hanks, Pustejovsky 2005: 66).

WordNet has been used as an ontology (lightweight ontology) in some applications based on the fact that the hypernymy hierarchies represent subsumption between concepts (Basile 2015). In addition, there are attempts (OntoWordNet) to convert WordNet into an ontology in two phases: automatic phase, in which WordNet glosses are parsed and an approximate definition of WordNet concepts is generated, in which generic associations (A-links) are established between the concept and other concepts that appear in its gloss; and partially automatic phase, in which the foundational top ontology DOLCE (in its DOLCE-Lite+ version with about 300 concepts) (Borgo et al. 2022) is used to interpret A-links in the form of axiomatized conceptual relations (Gangemi et al. 2003). WordNet has also been mapped several times to different ontologies: SUMO (Niles, Pease 2003); KYOTO (Laparra et al. 2012), etc.

The fact that WordNet can be mapped to pre-existing ontologies or that it can be transformed into an ontology shows that hypernymy inheritance between concepts has the character of an ontological representation, although some formal requirements for ontologies are not fulfilled. The two observations that the WordNet semantic classes can be subdivided into subclasses and that the taxonomic structure of a WordNet is close to an ontological representation lead to the view that, on the one hand, hierarchies of nouns in a WordNet are appropriate for selecting sets of nouns suitable for pairing with verbs in their valency slots, and that general semantic classes of nouns can be subdivided into appropriate subclasses appropriate for noun fillers' specification.

4. Corpus Pattern Analysis, Verb Patterns and Semantic Types

Corpus Pattern Analysis (CPA) is a technique for linking word meaning and usage (Hanks, Pustejovsky 2005: 64). The key idea is that while words have multifaceted potential to contribute to the meaning in a context, they have no specific meaning when used alone, and depending on the situation, different aspects of this meaning potential are realised. Evidence from corpuses demonstrates that the contextual patterns of word use are very regular.

4.1. CPA verb patterns

The Corpus Pattern Analysis is used to create a Pattern Dictionary of English Verbs (PDEV). A PDEV verb entry consists of a list of numbered patterns (frames) that are linked to implicatures – explanations of the meaning of the patterns (Hanks 2004: 88). For example, there are eleven registered patterns of the verb *grasp*, three of which are present here:

Human | Animal **grasp** Physical_Object
[[Human | Animal]] seizes [[Physical_Object]] and holds it firmly

Human 1 **grasp** Human 2 | Body_Part | Garment
[[Human 1]] seizes [[Body_Part | Garment]] of [[Human 2]] for some purpose

Human **grasp** at | for Physical_Object
[[Human]] attempts to seize [[Physical_Object]]

4.2. CPA semantic types

Collocation analysis enriches understanding beyond valency and captures nuanced semantic relationships between words (Hanks 2012: 58-60). By identifying statistically significant collocates, they can be organised into lexical sets sharing semantic features, such as [Human]. Each verb pattern's slot is represented by a lexical set of nouns, ranging from single words to extensive collections (Hanks 2012: 62).

The patterns are comprised of structured sentence roles, typically filled with nouns sharing aspects of their meaning, categorised as semantic types (Cinkova, Hanks 2010: 4; Hanks 2012: 66). A semantic type serves as a classification to which words can be assigned; for example, *Peter* or *the old man* are classified under the semantic type [Person] (Hanks, Pustejovsky 2005: 64). Essentially, semantic types such as [Human], [Animal], or [Part], generalise properties expressed by frequently encountered words in specific pattern positions (Hanks 2012: 57 – 59).

The initial iteration of the CPA ontology comprises a shallow ontology featuring 65 semantic types chosen for their prevalence in a manually identified selection of context patterns (Pustejovsky et al. 2004). The ontology of CPA semantic types is expandable, accommodating the addition of new types as they surface within emerging verb patterns, currently numbering 253 types. While some verb patterns demonstrate broad preferences, such as [Anything], others delineate preferences for a restricted set of words grouped into semantic types. These semantic types express the semantic preferences dictating the array of nouns and noun phrases commonly found in a given sentence position.

The verb *ask*, for example, is linked to twelve verb patterns, in most of which the semantic types are [Human], [Institution], and [Anything]:

Human 1 **ask** Human 2 QUOTE WH+

[[Human 1]] says {QUOTE} to ([[Human 2]]) in the form of a question, for example because [[Human 1]] wants to find out

Human 1 | Institution 1 **ask** Human 2 | Institution 2 question about Anything

[[Human 1 | Institution 1]] puts a {question} to ([[Human 2 | Institution 2]]) in order to find out ({about [[Anything]]})

Some verb patterns accept only a small selection of lexical units (in some cases, a word) as noun collocates, and no semantic type is defined; instead, the lexical units are listed in the verb pattern. For example, the word *permission* fits the following pattern for the verb *ask*:

Human 1 **ask** Human 2 permission to+INF
[[Human 1]] formally requests permission from [[Human 2]] to be allowed {to/INF [V]}

Semantic types, in contrast to WordNet semantic classes, stem from corpus-driven generalisations regarding noun groupings based on their collocations with specific verbs. CPA semantic types outnumber WordNet semantic classes approximately tenfold and can be further extended. While semantic types embody cognitive concepts grounded in corpus evidence, they retain abstract status without associations with collections of nouns belonging to their respective classes. Initially, the set of 65 semantic types was linked with around 20,000 nouns (Pustejovsky et al. 2004), but the details are not available. Hence, the task of relinking remains an ongoing endeavour.

5. Bulgarian FrameNet, Conceptual Frames and Semantic Classes

The efforts towards establishing the Bulgarian FrameNet span approximately 20 years, with its origins tracing back to predecessors like the Valency Dictionary for Bulgarian and the Semantic-Syntactic Dictionary for Bulgarian (Koeva et al. 2003). Initially, the focus of resources on frame-like semantic and syntactic descriptions was exclusively on Bulgarian, without establishing connections with FrameNet.

In subsequent phases, appropriate semantic frames were manually selected, and language-independent information was extracted from these frames. This data was then enriched with Bulgarian lexical units evoking corresponding frames, along with annotated examples (Koeva et al. 2008; Koeva 2010). However, this work encountered challenges, including the lack of appropriate means to maintain correspondence with semantic frames while reconstructing them to adequately represent certain Bulgarian lexical units. Additionally, challenges arose in encoding translation equivalence between Bulgarian and English lexical units and ensuring annotation consistency regarding Bulgarian grammatical structure.

At its current stage, the Bulgarian FrameNet incorporates two abstract semantic structures: a superframe and a conceptual frame. It includes lexical units accompanied by comprehensive lexical, semantic, and grammatical information that evoke conceptual frames, along with valency patterns derived from authentic examples.

The primary aim of introducing superframes and conceptual frames is to incorporate language-specific information while ensuring alignment with relevant semantic frames. Superframes establish abstract mappings between semantic frames in FrameNet and their Bulgarian counterparts, serving as a bridge between semantic resources. Conceptual frames, linked with a specific superframe, encode relevant information for Bulgarian, which may fully or partially overlap with their English counterparts (Figure 1).

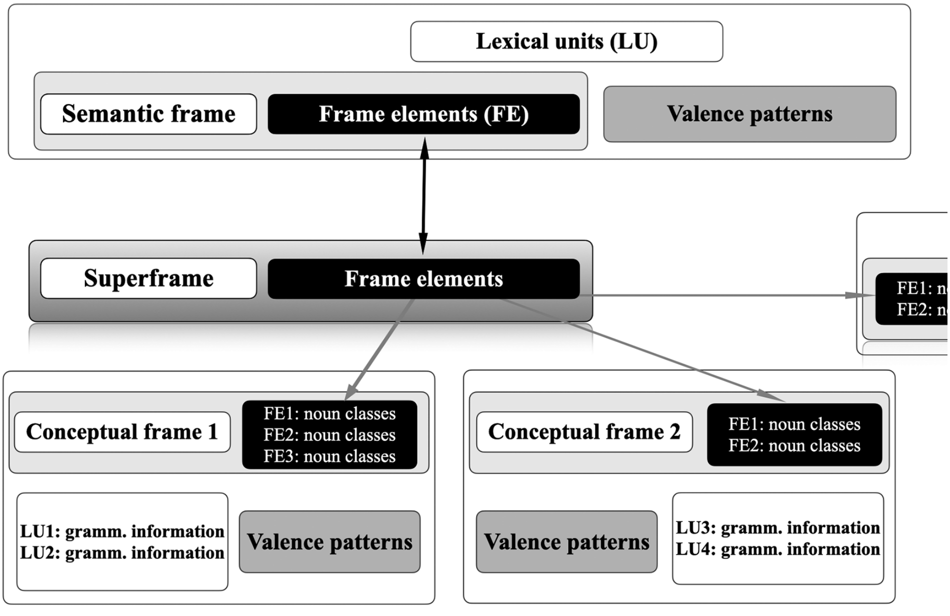

Figure 1. Organisation of data in the Bulgarian FrameNet

5.1 Superframes and conceptual frames in BulFrame

By introducing an intermediary abstract layer like superframes, the appropriate components in the Bulgarian FrameNet can align with FrameNet semantic frames while preserving specificity where needed. **Superframes** are crafted by stripping away all language-specific details for English, such as lexical units evoking the frames, their parts of speech and lexical types, and English sentences illustrating the frame and frame element definitions. Instead, they retain only non-language-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 details like frame and frame element names.

This approach aims to establish a seamless connection with FrameNet while facilitating the identification and outlining of language-specific conceptualisations unique to Bulgarian. It also allows for the possibility of splitting one semantic frame into multiple conceptual frames, each characterised by different levels of reconstruction. An equivalence relation is established between the language-independent information in a semantic frame and its counterpart in a superframe.

Conceptual frames serve to introduce script-like descriptions relevant to Bulgarian, which may be entirely or partially analogous to the information for English or unique (in rare cases), providing conceptual descriptions specific to Bulgarian. Consequently, one superframe may be associated with one or more conceptual frames. However, there can be at most one conceptual frame whose language-independent components are related to the superframe and, consequently, to the semantic frame through an equivalence relation. The remaining conceptual frames are linked to the superframe by partial equivalence relations, allowing for tracing to determine which components align with those in the superframe and which do not.

A **conceptual frame** can (similarly to the semantic frame) be defined as an abstract structure that describes a certain type of situation or event, along with its actors and properties (Koeva 2020: 7). The **conceptual frame** is characterised by frame elements, relations between frame elements; and it is supplemented with sets of nouns that are compatible with the lexical units evoking the frame. A given conceptual frame is evoked by a set of lexical units, that (as of 2024) are exclusively defined for verbs.

Conceptual frames include a frame name, definition, semantic type, frame elements, and relations between frames. **Frame elements** possess a name, definition, semantic type, core status, and relations to each other, including **Core sets**, **Requires**, and **Excludes.** This information is inherited from the semantic frames via superframes, provided they are already defined, and then validated for Bulgarian through annotation.

Our rationale for employing conceptual frames and superframes is founded on the following arguments:

- **Not all lexical units evoking a given semantic frame manifest the same semantic structure, leading to varied syntactic behaviours.**

In FrameNet, lexical units are clustered based on sharing the same frame semantics, disregarding similarities in syntactic behaviour, unlike Levin's verb classes (Levin 1993), and within a FrameNet frame, there could be sets of verbs with related senses but distinct syntactic properties. Within the context of FrameNet's comprehensive approach to conceptual description, we strive to distinguish sets of lexical units with equivalent semantic properties. Therefore, we adhere to the principle that the semantic description of lexical units associated with a particular conceptual frame is achieved through the utilisation of the same number and type of frame elements. This approach maintains the structure of semantic frames, as multiple conceptual frames may correspond to one superframe, and through it to one semantic frame. Furthermore, a direct one-to-one correspondence between a FrameNet semantic frame and a conceptual frame is often absent due to differences in conceptualization across languages. Through the abstract superframe, conceptual frames describing a scene (either wholly or partially) are interconnected, both with each other and with the corresponding FrameNet semantic frame. For example, the Bulgarian verbs *вдявам* 'I am threading' and *вдяна* 'thread', with the definition in WordNet 'pass, pierce a thread or floor through the eye of a needle' evoke the **Placing** frame. This frame encompasses core elements such as **Agent**, **Theme**, and **Goal**. The **Agent** is part of a **Core set** with the frame element **Cause**, and either of them controls the **Theme**, positioning it at a location, the **Goal**. When describing the Bulgarian verbs *вдявам* and *вдя-*

на, only the **Agent** frame element holds relevance within the conceptual frame **Placing**, while the **Cause** frame element is omitted.

1.a. [*Жената*]$_{\text{Agent}}$ *ВДЯНА* [*конеца*]$_{\text{Theme}}$ [*в ухото на иглата*]$_{\text{Goal}}$.
The woman inserted the thread into the eye of the needle.

1.b. *[*Вятърът*]$_{\text{Cause}}$ *ВДЯНА* [*конеца*]$_{\text{Theme}}$ [*в ухото на иглата*]$_{\text{Goal}}$.
The wind inserted the thread into the eye of the needle.

● **Unlike English and other languages, a large part of diatheses in Bulgarian involve a lexical and/or morphological alteration of the base verb.**

In FrameNet, there is no specific list of verbal diatheses that a semantic frame encompasses. However, certain instructions in the annotation process imply that diatheses linked with a particular predicate are considered part of the frame associated with the basic diathesis. For instance, an additional frame is not created to accommodate uses like *Those boots sell well*, which depersonalises and generalises one or more of the prominent participants, such as the **Seller** (Ruppenhofer et al. 2016: 12). A similar rationale applies to passives. On the other hand, the systematic non-inheritance relationships between stative frames and the inchoative and causative frames referring to them are delineated using the frame-to-frame relations **Causative of** and **Inchoative of** (Ruppenhofer et al. 2016: 85).

In Bulgarian, several verbal diatheses exist, including se passives, impersonal participle passives, impersonal se passives, middles, anticausatives, and lexical reciprocals (Koeva 2022: 153). These diatheses can be classified into two categories: structure-preserving and structure-modifying. In structure-preserving diatheses, the number of frame elements remains constant, but at least one of the frame elements is changed (i.e., lexical reciprocals).

2.a. [*Момчето*]$_{\text{Author}}$ *ПИШЕ* [*на едно китайче*]$_{\text{Addressee}}$.
The boy is writing to a Chinese.

2.b. [*Момчето*]Author$_1$/Addressee$_1$ *СИ ПИШЕ* [*с едно китайче*]Author$_2$/Addressee$_2$.
The boy is corresponding with a Chinese.

Conversely, in structure-modifying diatheses, the number of frame elements differs across the alternating diatheses (impersonal passives), and these may be accompanied by a frame element's alternation (middles and anticausatives).

3.a. [*Съседката*]$_{\text{Grinder}}$ *ТРОШИ* [*лед*]$_{\text{Patient}}$ *в кухнята*.
The neighbour is crushing ice in the kitchen.

3.b. *Забранено е да* *СЕ ТРОШИ*.
Crashing is prohibited.

3.c. [*Ледът*]$_{\text{Theme}}$ *СЕ ТРОШИ лесно*.
Ice crushes easily. 'The ice has the tendency to easily break.'

3.d. [*Ледът*]$_{\text{Theme}}$ *СЕ ТРОШИ*.
Ice crushes. 'The ice has the tendency to break.'

Hence, the verbs *пиша* (‘to compose a text in writing addressed to somebody’) and *пиша си* (‘to correspond with somebody’) will each be represented in distinct conceptual frames within the superframe **Text creation**. Similarly, the verbs *троша* (‘to break into small pieces’), *троши се* (‘it breaks’), *троши се* (‘to have the tendency to easily break’), and *троши се* (‘to have the tendency to break’) will be allocated to four separate conceptual frames within the superframe **Grinding**. This approach preserves the structure of the semantic frames without modification, while effectively reflecting the differences in both the semantic and syntactic structures of the verbs they represent.

Various types of information are encoded within the Bulgarian FrameNet. This includes lexical-semantic details such as lemma, part of speech, lexical type, semantic class, stylistic labels, and semantic relations. Additionally, grammatical information like verb aspect, transitivity, and the range of grammatical subjects is provided. Frame information encompasses frame definitions, frame-to-frame relations, frame elements, their definitions, core status, semantic types and relations, and ontological semantic classes of nouns suitable for pairing with a given target lexical unit. Furthermore, syntactic details for valency patterns are offered through the annotation of examples, including grammatical categories, grammatical functions, and the implicitness of frame elements. The sources for the information presented in Bulgarian FrameNet are outlined in Table 1.

Type of information	**Components**	**FrameNet**	**BulFrame**
Lexical information	Lemma	FrameNet	WordNet, BulFrame
	Part of speech	FrameNet	WordNet
	Definition	FrameNet, OUP*	WordNet, BulFrame
	Semantic class	No	WordNet
	Stylistic note	No	WordNet
	Lexical type	FrameNet	BulFrame
	Semantic relations	No	WordNet
Grammatical information	Verb aspect	No	WordNet
	Transitiviry	No	BulFrame
	Personality	No	BulFrame

* OUP – Oxford University Press

Frame information	Frame definition	FrameNet	FrameNet, BulFrame
	Frame-to-frame relations	FrameNet	FrameNet
	Frame elements	FrameNet	FrameNet, BulFrame
	Frame elements core status	FrameNet	FrameNet, BulFrame
	Frame elements definition	FrameNet	FrameNet, BulFrame
	Frame elements semantic type	FrameNet	FrameNet, BulFrame
	Frame elements relations	FrameNet	FrameNet, BulFrame
	Verb-to-Noun compatibility	No	BulFrame
Syntactic information	Grammatical category	FrameNet	BulFrame
	Grammatical function	FrameNet	BulFrame
	Implicitness	FrameNet	BulFrame

Table 1. Sources of information in Bulgarian FrameNet

5.2 Noun frame elements fillers

Conceptual frames differ from semantic frames by linking frame elements to a set of lexical units for potential realisation. Each core frame element within a conceptual frame is associated with a set of nouns that are compatible with the verbs that evoke the frame (Koeva 2020: 17; Koeva 2021: 184 — 185). This set can consist of one, several or numerous nouns linked by semantic relations at the lexical level such as synonymy and antonymy or by hierarchical conceptual relations such as hypernymy and hyponymy. For example, the verb *варя* 'boil' with the definition 'cook food in very hot or boiling water' evokes the semantic frame **Apply heat**, which is described by four core frame elements: **Cook**, **Food**, **Container**, and **Heating instrument**. Each frame element is linked to synsets (one or more) from the Bulgarian WordNet, which unite (as roots of hypernym subtrees) the sets of appropriate nouns for collocations with the target verb.

Cook: eng-30-00007846-n: {person}

Food: eng-30-07555863-n: {food}; eng-30-07649854-n: {meat}; eng-30-07775375-n: {fish}; eng-30-07707451-n: {vegetable}

Container: eng-30-03990474-n: {pot}

Heating instrument: eng-30-03343560-n: {fire}; eng-30-03543254-n: {stove}; eng-30-08581699-n: {hearth}

6. Towards an Extended Ontology of Semantic Classes of Nouns

The development of the Extended Ontology of Semantic Classes of Nouns is motivated by the development of the conceptual frames: abstract semantic descriptions, part of the Bulgarian FrameNet, in which frame elements are associated with semantic classes of nouns encompassing sets of nouns appropriate for collocations with target verbal lexical units evoking conceptual frames. The Extended Ontology of Semantic Classes of Nouns is based on WordNet, implying that each concept in the ontology is linked to a synonym set from WordNet whenever possible (although not mandatory). It is evident that some classes serve to ensure compatibility among multiple lexical units. However, there are also classes introduced solely to describe restrictions on several, or even just one, lexical unit.

The accepted approach involves selecting either the highest-ranked concept or a combination of concepts (part of the Ontology), which means the highest-ranked synset or a combination of synsets from the (Bulgarian) WordNet that encompasses all suitable noun synsets for the fillers of the frame elements (Koeva 2020: 17; Koeva 2021: 184 –185).

This approach streamlines the development of a robust training dataset for automatically labelling nouns and their semantic classes as frame element instances, thereby facilitating the annotation of valency patterns and the assignment of noun fillers to frame elements. In contrast, the manual annotation process in FrameNet, which entails identifying valency patterns and potentially extracting sets of noun fillers from annotated examples, is more labor-intensive. Furthermore, it lacks automatic classification of noun classes due to the limited size of sense-annotated corpora (with disambiguated senses), such as the Bulgarian sense-annotated corpus (Koeva 2012).

As previously noted, WordNet categorises nouns into broad semantic classes, which may not sufficiently reflect the semantic preferences of a diverse range of verbs. Moreover, multiple hypernymy in WordNet result from consolidating diverse taxonomic relations into a singular hypernymy. To address these challenges, we propose the following: a) linking WordNet synsets with more detailed ontological representations of noun semantic classes to enhance the selection of noun fillers; b) resolving multiple hypernymy within the WordNet structure; c) introducing additional semantic classes to specific synsets within hypernymy subtrees to facilitate precise selection.

6.1. Other Representations of Noun Fillers

Each of the mentioned resources (WordNet, CPA, and FrameNet) follows its own unique methodology. However, none of them aims to explicitly delineate the permissible and potential combinations of verbs and nouns realised in context, despite the assignment of semantic types regarding the noun fillers in PDEV and FrameNet.

Another manually crafted resource, VerbAtlas, offers comprehensive coverage of English verbs, defining prototypical argument structures for each cluster of WordNet synsets that build a semantically coherent frame (Di Fabio et al. 2019: 627). It also offers a limited set of explicit semantic roles, selectional preferences for the arguments in frames, and links to WordNet and BabelNet (Navigli et al. 2021). To address data sparsity concerns, VerbAtlas adopts VerbNet's roles (Kipper et al. 2008), reducing them from 39 to 27 (in contrast to FrameNet's practically unlimited number of frame elements). Selectional preferences in VerbAtlas are manually labeled from a set of 122 "macro-concepts" defined by WordNet synsets, whose hyponyms are expected to be probable candidates for the corresponding argument slot (Di Fabio et al., 2019: 630), employing a strategy akin to a previous algorithm-based approach (Agirre, Martinez, 2001).

The comparison in the interpretation of the verb *phone* in FrameNet and VerbAtlas shows some differences. In FrameNet, it evokes the semantic frame **Contacting** which is defined as "A **Communicator** (whose **Location** may be indicated) directs a **Communication** to an **Addressee** at a particular **Address**" together with the verbs: *cable.v*, *call in.v*, *call up.v*, *call.v*, *contact.v*, *e-mail.v*, *fax.v*, *get in touch.v*, *get through.v*, *mail.v*, *page.v*, *phone in.v*, *phone.v*, *radio.v*, *reach.v*, *ring up.v*, *ring.v*, *telegraph.v*, *telephone.v*, *telex.v*, *write in.v*, *write.v* (some nouns are also present). On the other hand, VerbAtlas frames it under **Communicate • Contact** which is defined as "Communicate with a place or person; establish communication with, as if by telephone" and embraces the following verbs: ***phone***[2], ***telephone***, ***call up***, ***contact***, *get through, get hold of; drop a line,* ***write***; *get in touch, connect, touch base, correspond, get, commune;* ***call***, ***call in***, *grunt-hoot, telepathise, telepathize; telecommunicate, pant-hoot, ping, rich out; network, cell phone, reticulate,* ***e-mail***, *email, netmail, call back, raise* (explicit synonymy between verbs in VerbAtlas is borrowed from WordNet; other semantic relations are retrievable through WordNet).

In VerbAtlas, the semantic roles describing the frame **Communicate • Contact** are **Agent**, **Patient**, **Topic**, **Recipient,** and **Instrument**, as opposed to the five core frame elements within the semantic frame **Contacting** in FrameNet: **Communicator** ("The person that receives the message from the **Communicator**"), **Communication** ("The information that the **Communicator** wishes to impart to the **Addressee**, often to get a particular response"), **Topic** ("This is the subject matter to which the message pertains. It is typically expressed as a PP Complement headed by *about*"), **Address** "This frame element is used for a (metaphorical) place in a system of communication where **Communicators** and **Addressees** can access the system (*call* ***at 555885***)", and **Addressee** ("The person that receives the message from the **Communicator**") and ten non-core. Some of the verbs in VerbAtlas such as *phone, telephone*, *call up*; *contact*, *get through*, *get hold of* are only defined with the semantic roles **Agent**, **Patient**, **Topic**, and **Recipient**, while some other verbs,

---

2 The lexical units that appear in both resources are underlined.

such as *telepathise*, *telepathize*; *telecommunicate*; *cell phone*; *email*, are additionally defined with the semantic role **Instrument.** In FrameNet, the frame elements {**Communication**, **Purpose** (a peripheral frame element), **Topic**} and {**Address**, **Addressee**} form **Core sets**, which means that at least one member of the group or all members can appear.

It seems that the frame elements **Communicator** and **Addressee** correspond to the semantic roles **Agent** and **Patient**, since **Patient** is characterised as [individual]. The lack of definitions for semantic roles makes it difficult to determine which of the frame elements, **Communication** and **Topic** (or both), corresponds to the semantic role **Topic** despite the coincidence of the names. The semantic role **Instrument**, described as **Implicit** and **Shadow**, possibly corresponds to the peripheral frame element **Medium** ("The physical or abstract setting in which the message is conveyed"). And the semantic role **Recipient** remains unclear in relation to the role **Patient.**

A comparison between the selectional preferences and the semantic types shows that in VerbAtlas, the **Agent** and **Patient** are specified as [individual], [social group] and [facility] respectively, while in FrameNet, the corresponding frame elements **Communicator** and **Addressee** are classified as [Sentient]. Similarly, VerbAtlas specifies **Topic** and **Recipient** as [entity], while FrameNet assigns the semantic type [Communication] to the frame element **Communication** and does not specify the frame element **[Topic]**. The implicit semantic role **Instrument** is characterised by a single concept rather than selectional preferences defining a set. For example, for the lexical unit *call back*, the implicit **Instrument** is a *phone*, while for the lexical unit *cell phone*, it is a *mobile phone*, and so forth.

For both resources, a prediction of verb-noun combinations presents a challenge, regardless of whether manual or corpus-observing methods are used. This is due to the abstract level of semantic types and selectional preferences and the inherent difficulties in reconciling figurative but acceptable usage. Although the number of verbs described in VerbAtlas still exceeds the number of verbs in FrameNet, the semantic and syntactic information (by means of frame elements, definitions of semantic frames and frame elements, relations between frames and frame elements, valence patterns, and annotation of examples) in FrameNet is much more extensive and comprehensive.

In the Brazilian FrameNet, each core frame element is analysed based on the aspect of the scene it represents, resulting in the assignment of one or more frames to the frame element (Torrent et al. 2022: 4 – 5). Only frames representing events, states, attributes, and relations are eligible for frame element-to-frame relations. The information provided by the core frame element definition or semantic type is used to determine the type of concept it refers to (e.g., *people*, *location*, *event*) and the top-level frame that represents it. By linking the conceptual structures that build it, such as semantic frames and frame elements, the FrameNet gains extra semantic information. The technique looks to be similar to the establishment of morphosemantic relations between verb and noun synsets in WordNet; however, this extension is not applicable to Bulgarian FrameNet because nouns have yet to be introduced in it.

Based on the review of the presented resources, several conclusions can be drawn. The definition of noun fillers for frame elements relies on ontological representations of abstract entities, some sourced from WordNet or linked to it. However, due to the relatively high level of abstraction (with only 253 semantic types used to classify nouns in CPA), accurately predicting noun fillers for frame elements (including semantic types for verb patterns' slots and selectional preferences for semantic roles) faces significant challenges.

The generalised concepts in the ontological representation of semantic classes of nouns to encompass a wide range of words, some of which are semantically incompatible with the target lexical units (apart from the fact that it is not technically possible to retrieve the members of the semantic classes unless they are linked to an extended ontology or to WordNet). Moreover, the WordNet noun hierarchies are constructed for a different purpose, and a hypernymy-hyponymy subtree may contain synsets belonging to different semantic classes or to different ontological classes, i.e., concrete and abstract nouns.

6.2. Mapping existing ontological representations to WordNet noun hierarchy

The extension of WordNet's 25 semantic classes includes the linking of WordNet concepts to different hierarchies (ontologies): CPA semantic types, FrameNet semantic types, and VerbAtlas selectional preferences.

Such a sub-classification has already been achieved by a manual mapping in which the semantic types of CPA were matched with the corresponding WordNet synsets (Koeva et al. 2018a). As a result, the synsets are categorised into 253 semantic CPA types in addition to the WordNet semantic classes, with hyponym noun synsets inheriting both the semantic class and the semantic type of their parent. The taxonomic organisation of WordNet facilitates inheritance between semantic classes and semantic types along the hierarchy, ensuring a more precise delineation of verb-noun compatibility.

For example, the synset eng-30-07881800-n: {beverage; drink} is marked both with the WordNet semantic class noun.food and CPA semantic type [Beverage], which is inherited by its hyponyms {smoothie}, {cider; cyder}, {wine}, etc.; the synset {wine} is additionally marked with the semantic type [Wine] complementary to semantic class noun.food and semantic type [Beverage], which is inherited by its hyponyms {dessert wine}, {Burgundy; Burgundy wine}, etc. As the organisation is taxonomic, the semantic type [Wine] shows that the entity is also [Beverage] and noun.food; however, the specification is more narrow and excludes solid food and nonalcoholic beverages as well as other types of alcoholic drinks.

Initially, 199 semantic types were mapped to one WordNet concept, 39 semantic types mapped to two WordNet concepts, 12 semantic types mapped to three concepts, 2 semantic types mapped to four concepts, and 1 semantic type mapped to five concepts (Koeva et al. 2018a: 75 – 76). For instance, the semantic type [Physical object] was mapped to synsets eng-30-00001930-n: {physical entity}, eng-30-00002684-n: {object; physical object}, and eng-30-00003553-n: {whole; unit}, while its child concept, the semantic type [Animate], was mapped to two synsets: eng-30-00004258-n: {living thing; animate thing} and eng-30-00004475-n: {organism; being}. Similarly, the next-level semantic type [Animal] was also mapped to two synsets: eng-30-00015388-n: {animal; animate being} and eng-30-01861778-n: {mammal; mammalian}. However, such one-to-many mappings violate ontological representation.

The ambiguity surrounding the initial decisions was resolved during the development of Bulgarian FrameNet by enriching conceptual frames with appropriate semantic classes that describe the realisations of frame elements. Further elaboration on the mapping revealed the exact mappings, for example: semantic type [Physical object] to synset eng-30-00002684-n: {object; physical object}, semantic type [Animate] to eng-30-00004258-n: {living thing; animate thing}, and semantic type [Animal] to eng-30-00015388-n: {animal; animate being} (Figure 2).

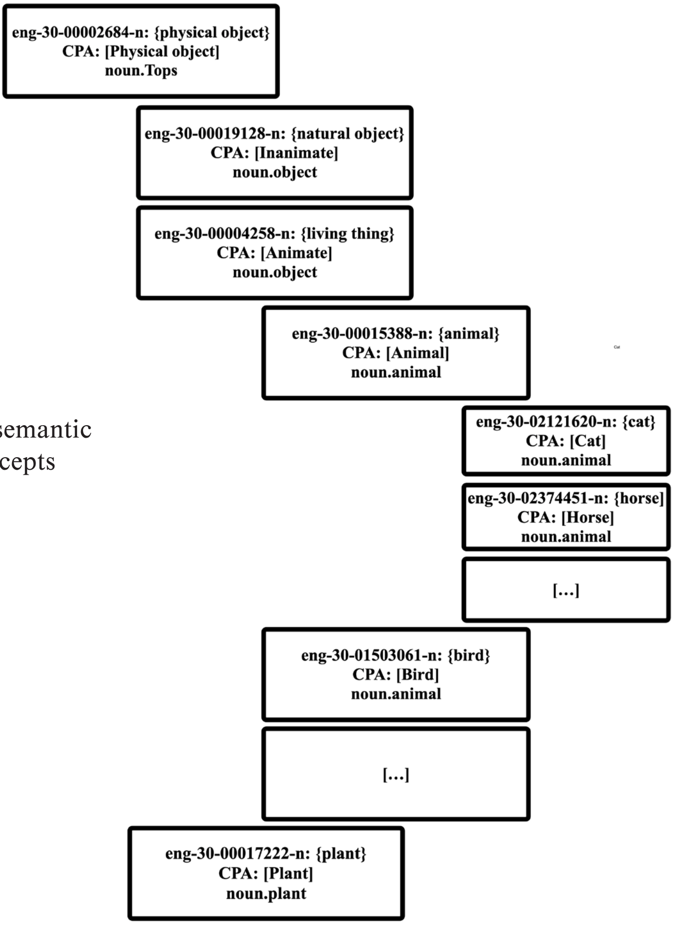

Figure 2. Mapping CPA semantic types to WordNet concepts

The small excerpt presented in Figure 2 also illustrates the sub-classification within the WordNet semantic classes, where the semantic class noun.animal was further categorised into semantic types such as [Cat], {Horse], [Bird], and others.

Other ontological representations mapped to WordNet synsets include FrameNet semantic types and VerbAtlas selectional preferences. This forms the basic structure of the Extended Ontology of Semantic Classes of Nouns.

The hierarchy of semantic types in FrameNet (Ruppenhofer et al. 2016: 86) aligns with WordNet concepts and, consequently, with CPA semantic types (and VerbAtlas selectional preferences). Due to the abstract nature of semantic types in FrameNet and their alignment with WordNet, only four additional types are included in the Extended Ontology of Semantic Classes of Nouns, different from the CPA semantic types. For example, the semantic type [Line] is linked to the synset eng-30-08593262-n: {line} with the semantic class noun.shape. Throughout the mapping process, any terminological differences used to denote the same concepts were standardised.

The mapping process between the VerbAtlas selectional preferences and WordNet concepts was facilitated by VerbAtlas's explicit referencing to BabelNet synsets, which in turn are linked to WordNet synsets. A comparison between the semantic types in the CPA and the selectional preferences in VerbAtlas reveals extensive overlap, with only a few unique selectional preferences found in VerbAtlas but not in the CPA. For example, the selectional preference [Liquid] exists in both the CPA and the VerbAtlas, and has already been mapped to the WordNet synset eng-30-14940100-n: {liquid} with the semantic class noun.substance. Only seven selectional preferences are not present in the CPA, and they were additionally added in the Extended Ontology of Semantic Classes of Nouns. For example, the selectional preference [Shot] is mapped to the synset eng-30-00565302-n: {stroke, shot}.

The mapping of the three ontological representations to WordNet concepts is illustrated in Figure 3 with the same example. Despite the overlap between four concepts at a very abstract level, they provide a stable foundation for the Extended Ontology of Semantic Classes of Nouns. Three and two overlaps also offer a certain level of confidence.

However, some levels within the hierarchy of the Ontology remain unpopulated, as they will be filled in during the process of defining concrete lexical units, conceptual frames, and frame elements in the Bulgarian FrameNet.

To conclude, the initial development of the Extended Ontology of Semantic Classes of Nouns entails mapping the relevant ontologies to the WordNet noun hierarchy, allowing for the identification of nodes indicating suitable noun fillers for specific frame elements. Through this mapping process, the original 25 semantic classes in WordNet were initially enriched with an additional 253 semantic types from the CPA. The ambiguity in the initial mapping of CPA semantic types was resolved by relying on evidence from the Bulgarian FrameNet and further validated by mapping it to the selectional preferences of the VerbAtlas and the semantic types of the FrameNet. Further, the mapping of CPA semantic

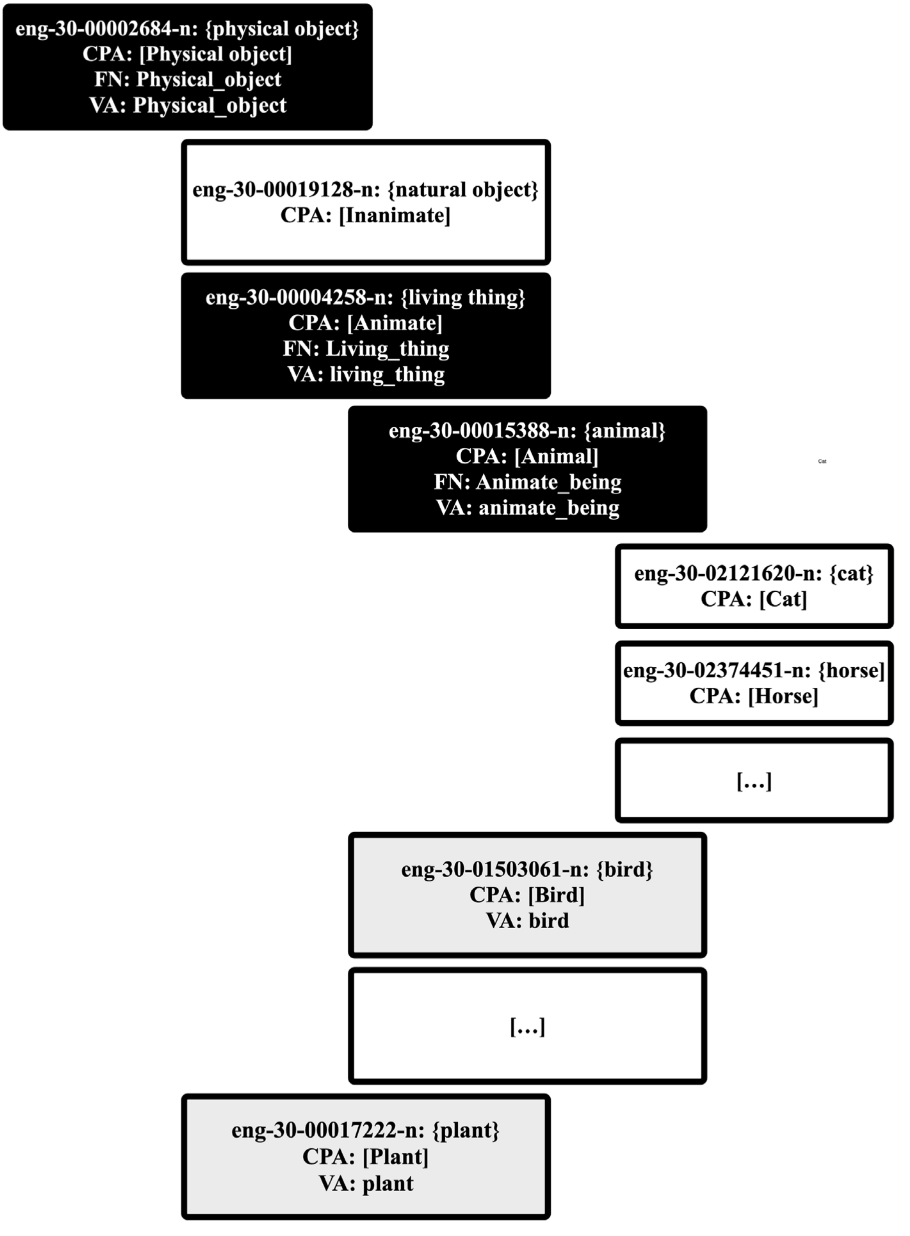

Figure 3. Mapping CPA semantic types, FrameNet semantic types, and VerbAtlas selectional preferences to WordNet concepts

types is augmented by incorporating an additional 7 selectional preferences from VerbAtlas and 4 semantic types from FrameNet.

The Extended Ontology of Semantic Classes of Nouns can be further expanded with new, more specific classes if the number of annotated Bulgarian verbs and elaborated frame elements increases, necessitating additional characterisation.

6.3. Resolving multiple hypernymy

In WordNet, the hypernymy relation may encompass various sub-relations, leading to hierarchies where nouns with vastly different characteristics coexist. For example, abstract and concrete nouns may appear together in a WordNet noun subtree, implying compatibility with verbs whose frame elements accommodate a broad range of nouns, such as *I think of* [entity: {*gesture*}; {*idea*}]. Some other verbs require nouns from specific classes, as seen in *I see* [physical object: {*gesture*}] *[abstraction: {*idea*}], and the inheritance of classes from multiple hypernyms introduces ambiguity. Therefore, to ensure the unambiguous inheritance of noun semantic classes, it is necessary to eliminate instances of multiple hypernymy and restrict class inheritance solely to the **is-a** relation or true hypernymy.

The structure of nouns in WordNet forms a directed connected graph comprising various semantic relations, with the taxonomic relation **is-a** (hypernymy) and its inverse relation (hyponymy) being most important for the semantic organisation. Both relations form a tree where noun synsets are interconnected through unique paths, with hypernyms having multiple hyponyms and each hyponym linked to exactly one hypernym.

Multiple hypernyms are classified into three types: exclusive, conjunctive, or non-exclusive, as outlined by EAGLES (1999). Exclusive multiple hypernyms (*albino* is either an *animal* or a *human*), often associated with polysemy, assign different hypernyms to distinct meanings of the same word. In contrast, conjunctive multiple hypernyms combine various semantic relations (*spoon* is both *cutlery* and *container*). Non-exclusive hypernyms accommodate both disjunctive and conjunctive relations (*knife* can be *cutlery*, a *weapon* or both), but the last ones are generally avoided in WordNet since different concepts should not be encoded within the same synset (node).

The current version of WordNet includes 1,421 synsets featuring multiple hypernyms, thereby undermining its taxonomic structure and the corresponding ontological representation of noun semantic classes. Figure 4 provides an illustration of three concepts characterised by multiple hypernyms. Moreover, the synsets depicted in Figure 4 (in grey) serve as examples of the alterations in noun semantic classes within certain noun hypernym sub-trees.

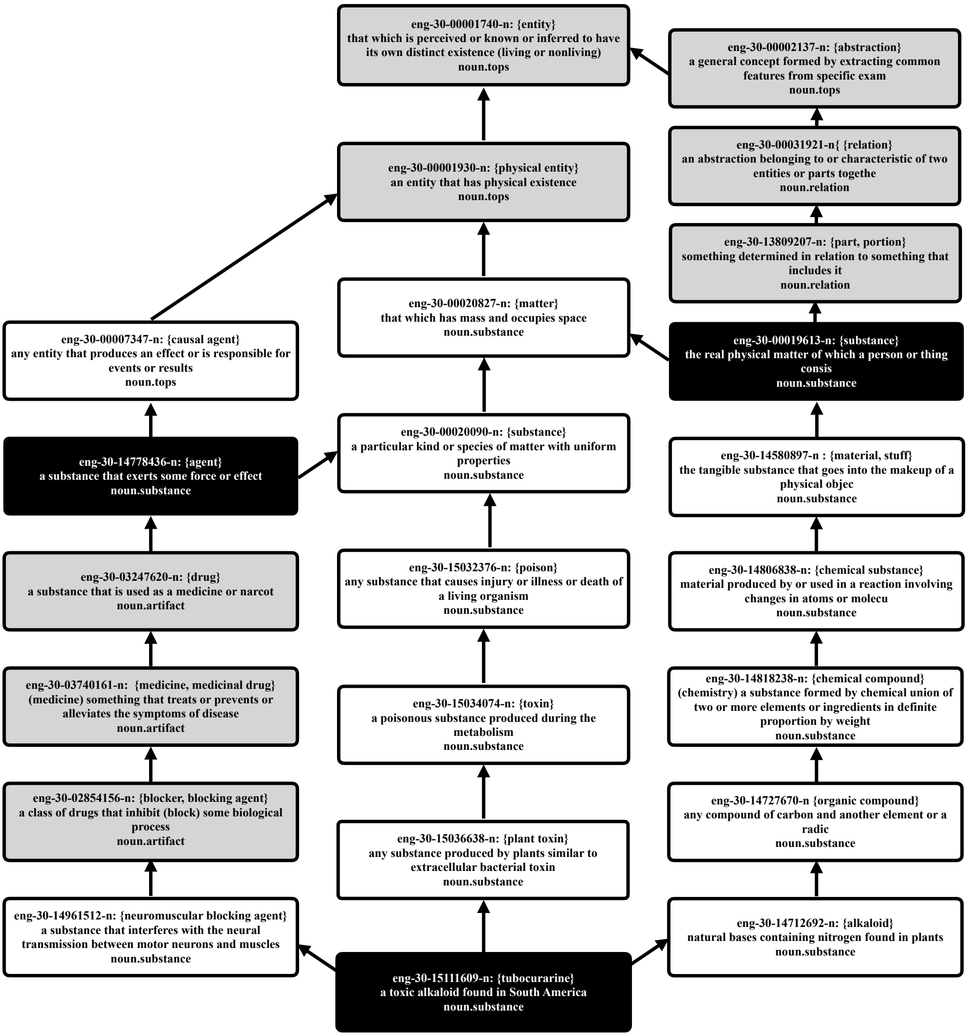

Figure 4. Multiple hypernyms in WordNet

Drawing on the principle that a synset ought to be linked to a solitary hypernym, the issue of multiple hypernymy was addressed through one of the following approaches (Koeva, Hristov 2023: 349) (Figure 5):

- Converting a multiple hyperonymy relation to one of nine alternative relatiions: **origin**, **form**, **function**, **characteristic**, **purpose**, **use**, **member**, **part**, or **portion**, the first three of which have already been proposed (Koeva et al. 2018b) and the last four of which are used in WordNet;
- Removal of a hypernymy relation when it lacks appropriate connectivity (rarely);

Figure 5 presents the decisions on resolving multiple hypernymy.

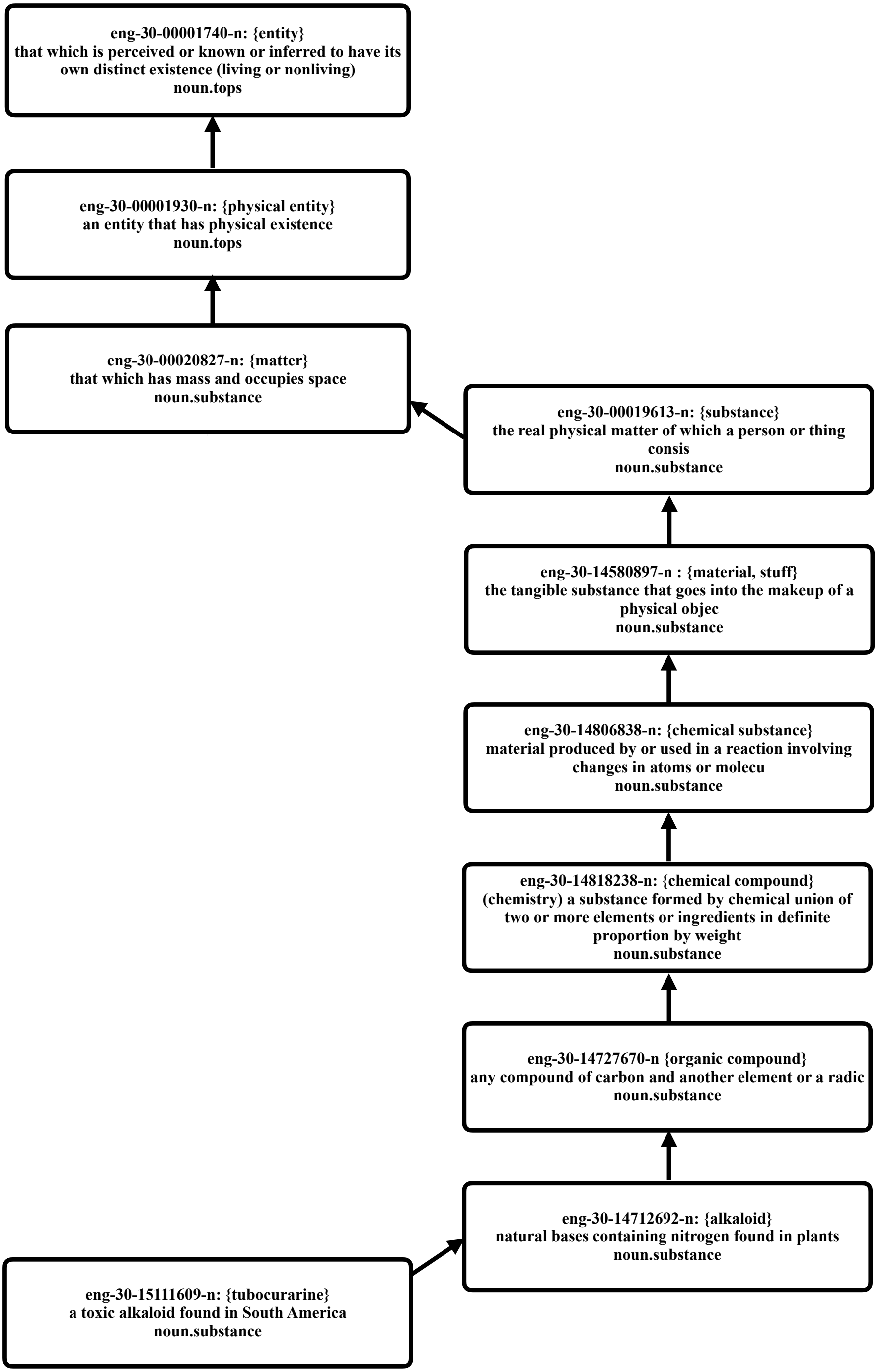

Figure 5. Resolving multiple hypernymy

- Introduction of a new hypernymy relation if none of the currently linked hypernyms are deemed suitable (rarely).

For example, the synset eng-30-15111609-n: {tubocurarine} ‘a toxic alkaloid found in South America’ was originally associated with three hypernyms: eng-30-14961512-n: {neuromuscular blocking agent} ’a substance that interferes with the neural transmission between motor neurons and skeletal muscles’; eng-30-15036638-n {plant toxin, phytotoxin} ‘any substance produced by plants that is similar in its properties to extracellular bacterial toxin’, and eng-30-14712692-n: {alkaloid} ‘natural bases containing nitrogen found in plants’. Following the analysis, the relation with {neuromuscular blocking agent} was reclassified as **function**, while the relation with eng-30-15036638-n: {plant toxin, phytotoxin} was designated as **purpose**.

Similarly, the synset eng-30-00019613-n: {substance} ‘the real physical matter of which a person or thing consists’ initially is linked to two hypernyms: eng-30-13809207-n: {part; portion; component} ‘something determined in relation to something that includes it’ and eng-30-00020827-n: {matter} ‘that which has mass and occupies space’. In this case, the relation to {part; portion; component} was removed.

This approach allows the redefinition of certain hypernymy relations and resolves instances of multiple hypernyms. By delineating distinct semantic relations, we can utilise solely the **is-a** inheritance relation to categorise the semantic classes of noun synsets into more refined groups and anticipate the compatibility of verbs and nouns.

6.4. Identification of noun fillers based on corpus analysis

For the set of verbs evoking a particular conceptual frame, suitable examples are sought in the Bulgarian National Corpus[3] or in other sources. By annotating examples that show the syntactic realisation of the frame elements, the valence patterns for a given lexical unit are constructed. The collection of examples also serves to observe which nouns are suitable for the syntactic realisation of the frame elements (as a single noun, noun phrase, or prepositional phrase).

This task is conducted by annotators, who select WordNet nodes that could potentially represent the set of candidate fillers and verify whether the concept is already included in the Extended Ontology of Semantic Classes of Nouns. Subsequently, they assess whether the subordinate noun synsets are suitable for pairing with the target verb as the syntactic realisation of its frame elements and make a decision on whether to utilise a single concept from the Ontology, a set of concepts, or introduce a new concept for defining the noun fillers.

To illustrate the proposed approach, the noun fillers for the frame elements of the **Statement** frame evoked by the verbs *обяснявам* and *обясня* (explain) ‘make plain and comprehensible’ will be presented. The core frame elements are

---

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

defined in FrameNet as follows: "The **Speaker** is the sentient entity that produces the **Message** (whether spoken or written); The **Message** is the frame element that identifies the content of what the **Speaker** is communicating to the **Addressee**. It can be expressed as a clause or as a noun phrase; The **Topic** is the subject matter to which the **Message** pertains. It is normally expressed as a PP Complement headed by *about*, but in some cases it can appear as a direct object". For example, *Evelyn spoke candidly* ***about her past***; "**Medium** is the physical entity or channel used by the **Speaker** to transmit the statement". For example, ***Reports*** *say Iran is working on laser enrichment technologies*. There are several non-core frame elements, among which **Addressee** is defined as: "The **Addressee** is the person to whom the **Message** is communicated. When this frame element is expressed, it often appears in a prepositional phrase introduced by *to*, or as a direct object".

The noun 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., denoting organisations responsible for certain functions, policies, or services. In this context, such nouns embody abstract concepts of administrative authorities, policy formulation, regulatory oversight, etc., which refer not to physical, tangible entities but to the collective functions and responsibilities associated with human activities. The concepts in the Ontology are defined by: a) the WordNet synset eng-30-02472293-n: {human, human being}, which matches the semantic type [Human] in CPA and FrameNet (although the frame element is specified as [Sentient]) and the selectional preference [human] in VerbAtlas; b) the WordNet synset eng-30-08008335-n {organisation}, which matches the semantic type [Institution] in CPA and [Organisation] in FrameNet and the selectional preference [social group] in VerbAtlas. This selectional preference corresponds to the synset eng-30-07950920-n: {social group} 'people sharing some social relation', which, however, is overly abstract. Among its direct hyponyms are synsets such as: eng-30-07966140-n: {society} 'an extended social group having a distinctive cultural and economic organisation'; eng-30-07966719-n: {sector} 'a social group that forms part of the society or the economy'; etc. Therefore, the most appropriate direct hyponym {organisation} was chosen as part of the Ontology. The same criterion was applied to assign the Ontology semantic classes to the frame element **Addressee**.

The primary consideration regarding the frame element **Message** fillers is that they ought to be nouns categorised as either 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 the nouns that cannot be colocated with the verb *explain* as objects.

The synset eng-30-00033020-n: {communication}, defined as 'something that is communicated by or to, or between people or groups', stands at the root of the hierarchy of nouns classified under the semantic class noun.communication. Nevertheless, this concept is too abstract to serve as a filler for the **Message** frame element.

Although the hyponyms of this synset (direct and full) are appropriate in some cases, others introduce inappropriate nouns to be combined with the target verbs. For example: eng-30-06520222-n: {receipt} 'an acknowledgment (usually tangible) that payment has been made'; eng-30-06275634-n: {mail} 'the bags of letters and packages that are transported by the postal service'; and eng-30-01102436-n: {publication} 'the communication of something to the public; making information generally known', among others.

Some of the non-combinable nouns within the subtree are concrete nouns {receipt} and {mail}, while others are labelled with the semantic class noun.act as {publication}. In such cases, an appropriate approach is to narrow down the set to nouns classified under the semantic class noun.communication and introduce another layer of classification: abstract and concrete nouns.

The synset at the top of the hierarchy, designated with the semantic class noun.cognition, is eng-30-00023271-n: {cognition; knowledge}, defined as 'the psychological result of perception, learning, and reasoning'. However, not all of its hyponyms are suitable as fillers for the **Message** frame element. To address this issue, two more specific concepts are chosen: eng-30-05816287-n {information} 'knowledge acquired through study or experience or instruction' and eng-30-05833840-n {idea; thought} 'the content of cognition; the main thing you are thinking about'. Both of these concepts are also defined within the selectional preferences in VerbAtlas, and the former is included in the CPA semantic types as well.

For the frame element **Topic**, the concept represented by the synset eng-30-00002137-n {abstraction; abstract entity}, defined as 'a general concept formed by extracting common features from specific examples', is chosen.

Medium can be conveyed through hyponyms of several concepts presented by the synsets: eng-30-06722453-n: {writing; written material}; eng-30-06722453-n: {statement}; eng-30-06263369-n: {press; public press}; eng-30-06277280-n: {television; telecasting}; and eng-30-06619428-n: {broadcast; program}, all falling under the semantic class noun.communication. The corresponding semantic types in CPA are [Television program] and [Document].

It can be concluded that the accurate determination of the appropriate noun classes to fill the positions of the frame elements of a given verb necessitates the combination of two approaches:

Selecting the most suitable concept or combination of concepts from the Extended Ontology of Semantic Classes of Nouns (presented by the WordNet noun synsets that dominate appropriate nouns).

Introducing additional elementary semantic types and classifying WordNet noun synsets based on these types to provide correct generalisations. These types may encompass *collective, abstract, concrete*, *agentive* and so on.

The development's contribution lies in detailing the systematically conceived and executed steps to establish a suitable set of semantic classes organised in the Extended Ontology of Semantic Classes of Nouns. The creation of the Ontology of Semantic Classes of Nouns was guided by the following principles:

a) Utilising available ontological representations of semantic noun classes, comparing and incorporating both overlapping and unique concepts based on real examples used to develop conceptual frames for Bulgarian FrameNet.

b) Aligning ontology concepts with synonym sets from WordNet to leverage WordNet's taxonomic organisation and associate ontology concepts with sets of nouns that are their hyponyms in WordNet.

c) Validation and refining the taxonomic organisation of nouns in WordNet to resolve instances of multiple hypernymy.

d) Augmenting the taxonomic organisation of nouns in WordNet with additional noun classifications, including divisions into abstract and concrete, animate and inanimate, human and non-human, agentive and non-agentive, where distinctions are not evident from existing semantic classes.

The first principle draws upon the ontology of semantic types in CPA and FrameNet, as well as the ontological representation of selectional preferences in VerbAtlas. While theoretically, these ontological representations can be supplemented with new concepts, they are in practice fixed structures. For example, the CPA ontology allows for the inclusion of new concepts based on new verb patterns; however, the resource has remained stagnant for many years. Moreover, the number of verb lexical units in FrameNet (over 5000), VerbAtlas (over 13 000), and CPA (approximately 1700 words with more than 5000 verb patterns) underscores the need for a comprehensive approach to model the compatibility of varied meanings of lexical units in context.

The second principle is followed to some extent in VerbAtlas and to a lesser extent in FrameNet, but both resources provide only a high-level representation of semantic combinations between verbs and nouns, akin to dictionaries.

Regarding the third and fourth principles, there have been no known attempts to address multiple hypernymy in WordNet or supplement its concepts with new semantic classes to distinguish between concrete and abstract nouns, animate and inanimate, etc.

7. Conclusions

The advancement of modern technologies and the emergence of powerful large language models have greatly enhanced the ability to predict the next word in a given context, particularly in English but increasingly in other languages as well. This development prompts the question of whether there is a need for classifying noun classes based on their compatibility with verbs. According to the authors of this study, the necessity for such classification is justified by the following factors:

Firstly, the proposed approach for classification is primarily useful for in-depth analysis and study of a particular language, whether it be one's native tongue or a foreign language. The classification of nouns not only illustrates word compatibility but also organises nouns based on shared semantic properties, offering insights into their ontological and independent class groupings.

Secondly, such classification can enhance the fine-tuning and performance of large language models by enabling them to accurately predict the compatibility of verbs and nouns not only in common contexts but also in rare and unconventional ones. This refinement can lead to more nuanced and contextually appropriate language generation.

The ultimate goal is to identify classification models that effectively describe noun classes, whether achieved by experts, automatically, or through a combination of both methods. This approach aims to enhance our understanding of language structure and improve the capabilities of language processing systems.

Acknowledgments: This study is carried out as part of the scientific programme under the project *Enriching the Semantic Network WordNet with Conceptual Frames* funded by the Bulgarian National Science Fund (Grant Agreement No. KP-06-N50/1 of 2020).

References

Agirre, Martinez 2001: Agirre, E., D. Martinez. Learning class-to-class selectional preferences. – In: *Proceedings of the ACL 2001 Workshop on Computational Natural Language Learning (ConLL)*. <https://aclanthology.org/W01-0703> [20.01.2024]

Basile 2015: Basile, V. WordNet as an Ontology for Generation. – In: *WebNLG 2015 1st International Workshop on Natural Language Generation from the Semantic Web, Jun 2015, Nancy, France*. <https://inria.hal.science/hal-01195793/document> [20.01.2024]

Borgo et al. 2022: Borgo, S., R. Ferrario, A. Gangemi, N. Guarino, C. Masolo, D. Porello, E. M. Sanfilippo, L. Vieu. DOLCE: A Descriptive Ontology for Linguistic and Cognitive Engineering. – *Applied Ontology,* 17 (1), pp. 45 – 69.

Cinkova, Hanks 2010: Cinkova, S., P. Hanks. Validation of Corpus Pattern Analysis. – In: *Assigning Pattern Numbers to Random Verb Samples,* pp. 1 – 16. <https://ufal.mff.cuni.cz/project/spr/data/publications/annotation_manual.pdf> [20.01.2024]

Di Fabio et al. 2019: Di Fabio, A., S. Conia, R. Navigli. VerbAtlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. v 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).* Hong Kong: Association for Computational Linguistics, pp. 627 – 637. <https://aclanthology.org/D19-1058/> [20.01.2024]

EAGLES 1999: *Preliminary Recommendations on Lexical Semantic Encoding. Final report*. <https://www.ilc.cnr.it/EAGLES96/EAGLESLE.PDF> [20.01.2024]

Fellbaum 1990/1993: Fellbaum, C. English Verbs as a Semantic Net. – *International Journal of Lexicography*, Volume 3, Issue 4, Winter 1990, pp. 278 – 301; reprinted in 1993, pp. 40 – 51. <https://wordnetcode.princeton.edu/5papers.pdf> [20.01.2024]

Fillmore 1976: Fillmore, C. J. Frame semantics and the nature of language. – *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech,* 280(1), pp. 20 – 32.

Fillmore 1982: Fillmore, C. J. Frame semantics. – In: *Linguistics in the morning calm*. Seoul: Hanshin Publishing, pp. 111 – 137.

Fillmore 2007: Fillmore, C. J. Valency issues in FrameNet. – In: Herbst, T., K. Gцtz-Votteler (Eds.). *Valency: Theoretical, descriptive and cognitive issues*. Berlin: Mouton de Gruyter, pp. 129 – 160.

Fillmore, Baker 2010: Fillmore, C. J., C. Baker. A frames approach to semantic analysis. – In: Heine, B., H. Narrog (Eds.). *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press, pp. 313 – 340.

Fillmore et al. 2003: Fillmore, C. J., C. R. Johnson, M. R. L. Petruck. Background to FrameNet. – *International Journal of Lexicography,* 16(3), pp. 235 – 250.

Gangemi et al. 2003: Gangemi, A., R. Navigli, P. Velardi. The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. – In: Meersman, R., Z. Tari, D. C. Schmidt (Eds.). *On The Move to Meaningful Internet Systems 2003*: *CoopIS, DOA*, *and ODBASE*. *OTM 2003. Lecture Notes in Computer Science, vol 2888*. Berlin, Heidelberg: Springer, pp. 820 – 838.

Hanks 2004: Hanks, P. Corpus pattern analysis. – In: *Proceedings of the 11th EURAL- EX International Congress (EURALEX 2004), Lorient, France, 6–10 July 2004*. Lorient: Universitй Bretagne Sud, pp. 87 – 97.

Hanks 2012: Hanks, P. How people use words to make meanings: Semantic types meet valencies. – In: Boulton, A., J. Thomas (Eds.). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, pp. 54 – 70. <https://www.ceeol.com/search/chapter-detail?id=837374> [29.02.2024]

Hanks, Pustejovsky 2005: Hanks, P., J. Pustejovsky. A Pattern Dictionary for Natural Language Processing. – *Revue franзaise de linguistique appliquйe*, 2 (Vol. X), pp 63 – 82. <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2005-2-page-63.htm> [20.01.2024]

Koeva 2010: Koeva, S. Lexicon and grammar in Bulgarian FrameNet. – In: *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*. Valletta: European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2010/pdf/705_Paper.pdf> [20.01.2024]

Koeva 2012: Koeva, S. Bulgarian Sense-annotated Corpus – between the tradition and novelty. – In: *Études Cognitives / Studia Kognitywne,* 12, pp. 181 – 198.

Koeva 2020: Koeva, S. Semantic Relations and Conceptual Frames. – In: Koeva, S. (Ed.). *Towards a Semantic Network Enriched with a Variety of Semantic Relations*. Sofia: Institute for Bulgarian Language Prof. Lyubomir Andreychin, pp. 7 – 20.

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

Koeva 2022: Koeva, S. The system of diatheses in Bulgarian and stative predicates. – In: Koeva, S., E. Ivanova, Y. Tisheva, A. Zimmerling (Eds.). *Ontology of Stative Situations – Linguistic Modeling. A Contrastive Bulgarian-Russian Study*. Sofia: Prof. Marin Drinov Publishing House of the Bulgarian Academy of Sciences, pp. 117 – 160.

Koeva, Hristov 2023: Koeva, S., D. Hristov. Resolving Multiple Hyperonymy. – In: *Proceedings of the 12th Global WordNet Conference*. University of the Basque Country, Donostia – San Sebastian, Basque Country: Global WordNet Association, pp. 343 – 351.

Koeva et al. 2003: Koeva, S., E. Doychev, G. Cholakov. 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*, pp. 41 – 47.

Koeva et al. 2008: Коева, С., Р. Влахова, Р. Декова, П. Несторова, А. Атанасов. Представяне на езиковата информация в Българския ФреймНет – лингвистична мотивировка. – В: *Българският ФреймНет. Семантико-синтактичен речник на българския език*.

Коева, С. (съст.). София: Институт за български език „Проф. Любомир Андрейчин“, с. 58 – 78. (Koeva, S., R. Vlahova, R. Dekova, P. Nestorova, A. Atanasov. Predstavyane na ezikovata informatsia v Balgarskiya FreymNet – lingvistichna motivirovka. – In: *Balgarskiyat FreymNet. Semantiko-sintaktichen rechnik na balgarskiya ezik*. Koeva, S. (sast.). Sofia: Institut za balgarski ezik „Prof. Lyubomir Andreychin“, pp. 58 – 78.)

Koeva et al. 2018a: Koeva, S., T. Dimitrova, V. Stefanova, D. Hristov. Mapping WordNet concepts with CPA ontology. – In: *Proceedings of the 9th Global WordNet Conference (GWC'2018).* Singapore: Global WordNet Association, pp. 70 – 77.

Koeva et al. 2018b: Коева, Св., В. Стефанова, Д. Христов. Семантични релации в рамките на многократната хиперонимия в УърдНет. – *Чуждоезиково обучение*, 45, 4, 2018, с. 343 – 365. (Koeva, S., V. Stefanova, D. Hristov. Semantichni relacii v ramkite na mnogoktratnata hiperonimiya. – *Chuzhdoezikovo obuchenie*, 45, 4, 2018, pp. 343 – 365.)

Laparra et al. 2012: Laparra, E., G. Rigau, P. Vossen. Mapping WordNet to the Kyoto ontology. – In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association (ELRA), pp. 2584 – 2589.

Loenneker-Rodman, Baker 2009: Lönneker-Rodman, B., C. F. Baker. The FrameNet model and its applications. – *Natural Language Engineering,* 15(3), pp. 415 – 453.

Miller 1986: Miller, G. Dictionaries in the mind. – *Language and Cognitive Processes*, 1, pp. 171 – 185.

Miller 1990/1993: Miller, G. Nouns in WordNet: A Lexical Inheritance System. – *International Journal of Lexicography*, 3(4), 1990, pp. 245 – 264; reprinted in 1993, pp. 10 – 25. <https://wordnetcode.princeton.edu/5papers.pdf> [20.01.2024]

Miller et al. 1990/1993: Miller, G., R. Beckwith, C. Fellbaum, D. Gross, K. Miller. Introduction to WordNet: an On-line Lexical Database. – *International Journal of Lexicography*, 3(4), 1990, pp. 233 – 244; reprinted in 1993, pp. 1 – 9. <https://wordnetcode.princeton.edu/5papers.pdf> [20.01.2024]

Navigli et al. 2021: Navigli, R., M. Bevilacqua, S. Conia, D. Montagnini, F. Cecconi. Ten Years of BabelNet: A Survey. – In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence Survey Track*. pp. 4559 – 4567. <https://www.ijcai.org/proceedings/2021/0620.pdf> [20.01.2024]

Niles, Pease 2003: Niles, I., A. Pease. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. – In: *Proceedings of the 2003 International Conference on Information and Knowledge Engineering*. Las Vegas, Nevada. <https://www.adampease.org/Niles-IKE.pdf> [20.01.2024]

Pustejovsky et al. 2004: Pustejovsky, J., P. Hanks, A. Rumshisky. Automated Induction of Sense in Context. – In: *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. Geneva: COLING, pp. 924 – 930. <http://acl.ldc.upenn.edu/C/C04/C04-1133.pdf> [20.01.2024].

Ruppenhofer et al. 2016: Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. Johnson, C. F. Baker, J. Scheffczyk. *FrameNet II: extended theory and practice*. Berkeley, California: International Computer Science Institute.

Sikos, Pado 2018: Sikos, J., S. Pady. Framenet's using relation as a source of concept-based paraphrases. – *Constructions and Frames,* 10(1), pp. 38 – 60.

Torrent et al. 2022: Torrent, T. T., E. E. da Sylva Matos, F. Belcavello, M. A. Viridiano, M. Gamonal, A. D. da Costa, M. C. Marim. Representing Context in FrameNet: A Multidimensional, Multimodal Approach. – *Frontiers in Psychology,* Vol. 13. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.838441/full> [20.01.2024]

РАЗШИРЕНА ОНТОЛОГИЯ НА СЕМАНТИЧНИТЕ КЛАСОВЕ НА СЪЩЕСТВИТЕЛНИТЕ ИМЕНА

Светла Коева

Институт за български език „Проф. Любомир Андрейчин“, Българска академия на науките

svetla@dcl.bas.bg

Резюме. Основната цел на изследването е да се предложи ефективен подход за характеризиране на фреймовите елементи от концептуалните фреймове на Българския фреймнет посредством класове съществителни имена, с цел да се предложат изчерпателни и семантично валидни комбинации между глаголните лексикални единици, предизвикващи представата за концептуален фрейм, и подходящи съществителни. Студията предлага кратък преглед на семантичните класификации на съществителните в Уърднет, Корпусния анализ на изреченските модели, Фреймнет и Върбатлас, като се подчертава тяхното значение за определяне на съчетаемостта между глаголи и съществителни. Представя се структурата на Българския фреймнет (Булфрейм), който съдържа валидните за български компоненти, заимствани от Фреймнет, но и голямо количество лексикална, морфологична, синтактична и семантична информация, специфична за български език. Една от отличителните характеристики на Българския фреймнет е спецификацията на класове съществителни имена, които показват множеството от подходящи съществителни за лексикална реализация на фреймовите елементи. Чрез съотнасяне на синонимни множества от Уърднет със семантичните типове на Корпусния анализ на изреченските модели и Фреймнет, както и със селективните предпочитания на Върбатлас, се изгражда основната структура на Разширената онтология на семантичните класове на съществителните. Понятията в Онтологията са свързани (не изключително) със синонимните множества от Уърднет и следователно, с множества от съществителни, подходящи за свързване с глаголните лексикални единици, които предизвикват представата за съответните концептуални фреймове. Приносът на разработката се състои в детайлизиране на стъпките за подбор на семантичните класове, изграждащи Разширената онтология на семантичните класове на съществителните.

Ключови думи: *семантичен клас, семантичен тип, селективни предпочитания, семантичен фрейм, фреймов елемент*

Svetla Koeva
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., bl. 17
Sofia 1113, Bulgaria

https://orcid.org/0000-0001-5947-8736