TOWARDS DETAILED SEMANTIC CLASSES OF NOUNS[1]

Svetla Koeva

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

svetla@dcl.bas.bg

Abstract. The primary objective of the study is to present an approach for effectively classifying noun classes, facilitating the specification of syntagmatic combinations between verbs and nouns. It offers a concise overview of the semantic classifications of nouns in WordNet, Corpus Pattern Analysis, and FrameNet from the perspective of verb-noun compatibility. Mapping the appropriate WordNet synsets with the semantic types of Corpus Pattern Analysis, FrameNet and with the selective preferences of VerbAtlas allows the creation of an Extended Ontology of Noun Semantic Classes. The contribution of this development lies in detailing the steps for selecting an appropriate set of semantic classes whose members are suitable for the syntactic realisation of certain frame elements.

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

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), and frame elements in FrameNet are often specified according to the semantic types of nouns (Ruppenhofer et al. 2016: 86 – 87).

The aim of this study is to define an approach to the appropriate identification of noun classes that allows the specification of syntagmatic combinations between verbs and nouns, or, in other words, of noun fillers for FrameNet frame elements. The paper (as part of the brother's investigation) shows why the available classifications are not at all suitable to be used ad hoc to illustrate syntagmatic combinations. Based on the existing language resources, we offer a complex approach that combines appropriate (ontological) representations of noun classes and extends them with some other classifications when necessary. As a result we aim at the developing of an Extended ontology of semantic classes of nouns.

The paper is organised as follows: The next three sections provide a concise overview of the semantic classifications of nouns in WordNet, Corpus Pattern Analysis, and FrameNet from the perspective of verb-noun compatibility. Section five provides a brief discussion of additional resources determining groupings of nouns that can co-occur with specific verbs. Section six outlines our approach to identifying noun classes with appropriate candidates for particular verb-noun compatibilities.

2. WordNet Semantic Classes (Primitives)

WordNet, a lexical database of the English language, uses a series of semantic primitives (semantic classes) for nouns and verbs to categorise concepts and place them in a hierarchical structure (Miller 1990/1993: 16; Fellbaum 1990/1993: 41). These semantic primitives serve as basic building blocks that capture the essential semantic distinctions and help to categorise words on the basis of their meanings.

Nouns are categorised into twenty-five semantic classes: {**act**, action, activity,} {**animal**; fauna}, {**artifact**}, {**attribute;** property}, {**body;** corpus}, {**cognition;** knowledge}, {**communication**}, {**event;**

[1] This paper 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).

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, communication, 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).

The nouns are divided according to sets of semantic primes, selecting a (relatively small) number of fundamental concepts and treating each one as the distinct starting point of a separate hierarchy. These various hierarchies align with relatively distinct semantic domains, each possessing its own vocabulary (Miller 1990/1993: 16). Noun hypernymy subtrees could be connected in a hierarchical structure with the most abstract root {entity}. There are more than 550 verb subtrees in WordNet, and here one and the same semantic class can appear in many subtrees; again, the verb subtrees could be artificially united in an abstract root {eventuality}.

Noun synonymy sets (synsets) such as {communication; communicating}, {language; linguistic communication}, {visual communication}, {email; e-mail; electronic mail}, {body language}, etc., have the semantic class noun.communication. Verb synsets such as {communicate; pass on; pass; pass along; put across}, {articulate; enunciate; vocalise}, {sign; signal; signalise}, {broadcast; air; transmit; beam; send}, etc., are specified with the semantic class verb.communication. It is intuitively clear that nouns and verbs referring to verbal and non-verbal communication form different semantic groups; the same can be said for communication in written form, communication by signalling, communication by broadcasting, etc. Also, combinations such as *he/she passes on conversation*; *he/she shares or relays conversations to others*; *he/she communicates by email* are grammatically correct and acceptable, while combinations such as **he/she signals language*; **he/she articulates email* (although grammatically correct) are unusual and unacceptable. It can be concluded that there is still much scope for further specification and specialisation of WordNet semantic classes to capture semantic compatibility between verbs and nouns.

It has been noted that the synsets within the WordNet noun hierarchy may not universally correspond to their associated semantic classes. Additionally, these classes may not consistently fulfil specific slots within verb argument structures (Hanks, Pustejovsky 2005: 66).

To address the key syntactic properties of verbs, WordNet presents, for each verb synset, one or several **sentence frames**. These frames delineate the subcategorisation properties of the verbs within the synset by outlining the types of sentences in which they can appear (Fellbaum 1990/1993: 55). For example, the sentence frame **Somebody ------s something to somebody** includes, in addition to the number and type of syntactically obligatory constituents, minimal information about the selective restrictions on combinability: whether a given element in the frame can be realised as a noun that is human or not, and minimal syntactic information: whether the element is realised as a noun, prepositional phrase, or clause.

3. Semantic Types in Corpus Pattern Analysis

Corpus Pattern Analysis (CPA) is a technique for mapping meaning to words in a text. It 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). The patterns consist of ordered sentence roles that are usually filled with nouns that share some aspects of their meaning, which is described as a semantic type (Cinkova, Hanks 2010: 4; Hanks 2012: 66). In other words, semantic types (e.g., [Human], [Animal], [Part], etc.) generalise properties expressed by frequently occurring words in certain pattern positions (Hanks 2012: 57 – 59).

The semantic types are organised in a shallow ontology (up to 10 sublevels for some types) based on the analysis of corpus data, and they can be extended with additional semantic types as they appear in new verb patterns. Some patterns involve very broad preferences, such as the semantic type [Anything], while others differentiate preferences for a limited number of words grouped into semantic types. Semantic types serve to express the semantic preferences that determine the range of nouns and noun phrases that typically occur in a given sentence position. Compared to WordNet semantic classes, semantic types are corpora-driven generalisations about the grouping of nouns into classes based on their collocations with particular verbs. CPA semantic types are about 10 times more than WordNet semantic classes. Examples of semantic types included in the CPA ontology and related to the communication domain are: [Information Source], [Medium], [Broadcast], [Document], [Agreement], [Language], [Radio Programme], and so on. The verb *tell*, for example, is linked to 21 verb patterns, for some of which a pattern position is given as [Information Source]:

1. Human 1 | Institution 1 | Information_Source **tells** Human 2 | Institution 2 QUOTE THAT WH+
[[Human 1 | Institution 1 | Information_Source]] **informs** [[Human 2 | Institution 2]] {that [CLAUSE] | WH- [CLAUSE] | [QUOTE]}

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 *news* fits the following pattern for the verb *tell*:

2. Human 1 **tells** news
[[Human 1]] **informs** [[Human 2]] about {news}

One of the most important features of semantic types is that they are corpus-based, i.e., they are created from corpora using real-world examples. Although semantic types represent cognitive concepts that play an important role in the use of words, they remain abstract notions as they are not associated with collections of nouns that are members of the respective class. It is therefore not specified which particular nouns can be interpreted, for example, as [Information Source].

4. FrameNet Semantic Types

FrameNet is based on the theory of frame semantics (Fillmore 1982; Fillmore, Baker 2010). Semantic frames are defined as a *schematic representation of speakers knowledge of the situations or states of affair that underlie the meanings of lexical items* (Fillmore 2007: 130). Frame elements are components of a semantic frame and stand for the participants, props, phases, and parts of situations (Fillmore 2007: 130). Some frame elements are associated with a semantic type, indicating the basic categorisation of their fillers (Ruppenhofer et al. 2016: 86).

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 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 can be directly translated into WordNet synset nodes and ontologies. They are arranged in a hierarchy, with the inheritance relation (or is-a relation) between a parent node and a child node (Ruppenhofer et al. 2016: 86). At the top-level, a distinction is made between five types: [Attribute], [Physical entity], [Event], [Group], and [Relations]. The remaining forty ontological semantic types fall under one of these five top-level types, e.g., [Shape] is-a [Attribute]; [Human] is-a [Sentient] which is-a [Animate being] which is-a [Living thing] which is-a [Physical object] which is-a [Physical entity]; [Activity] is-a [Event] which is-a [State of Affairs]; [Organisation] is-a [Group]; [Path] is-a [Locative relation], and so on.

Not all frame elements are given a semantic type, or the semantic types are too general, and in some cases, they do not show the actual constraints on lexical combinations (Koeva 2020: 11). For example, the following frame elements of the semantic frame **Telling** are provided with semantic types: **Speaker** with the semantic type [Sentient]; **Addressee** with the semantic type [Sentient]; **Message** with the semantic type [Message]; **Manner** with the semantic type [Manner]; **Time** with the semantic type [Time], while the frame element **Topic** remains unspecified.

As can be seen, semantic types provide a high level of abstraction and, in most cases, can hardly be used to predict divergent context realisations.

5. Other Representations of Noun Fillers

Each of the presented resources (WordNet, PDEV and FrameNet) is built on the basis of its own methodology. On the other hand, none of them is intended to specify the allowed and possible verb-noun combinations realised in context, although in PDEV and FrameNet the semantic types of noun fillers are assigned.

Another manually crafted resource, VerbAtlas, provides extensive coverage of English verbs; prototypical argument structures for each cluster of synsets that define a semantically coherent frame; a small set of explicit semantic roles; selectional preferences for the frames' arguments; and a link to WordNet and BabelNet (Di Fabio et al. 2019: 627).

VerbAtlas has a smaller collection of semantic roles than FrameNet frame elements, with 27 roles compared to FrameNet's practically unlimited number of frame elements (Di Fabio et al. 2019: 628). The selectional preferences in VerbAtlas have been manually labelled from a set of 122 "macro-concepts" defined by WordNet synsets, whose hyponyms are expected to be likely candidates for the corresponding argument slot (Di Fabio et al. 2019: 630), a strategy similar to that previous one, which, instead, was algorithm-based (Agirre, Martinez 2001).

The comparison in the interpretation of the verb *inform* in FrameNet and VerbAtlas shows that the semantic frame in FrameNet **Telling** includes verbs such as *advise*.v, *apprise*.v, *assure*.v, *brief*.v, *confide*.v, *inform*.v, *let know*.v, *notify*.v, *tell*.v, while the range of verbs with a similar meaning in VerbAtlas frame **Inform** is higher: *acquaint; advise, apprise, notify; apprise, apprize, instruct; awaken; circularise; communicate, pass, pass along; convey, communicate; evidence, tell; fill in; follow, keep abreast, keep up; give; give, impart, leave; hear; inform; input; keep track; misadvise, misguide; nark; popularise, popularize; read, scan; report; sell out; send a message; tell, assure; tip, tip off; update, refresh; wake; wise up*, and the synonymy is explicit. In VerbAtlas, the semantic roles include **Agent**, **Recipient**, **Object**, and **Instrument**, as opposed to the five core frame elements in FrameNet. A comparison between selectional preferences and semantic types reveals that in VerbAtlas, the **Agent** and **Recipient** are specified as [individual] (human being) and [social group] respectively, while in FrameNet, the corresponding frame elements **Speaker** and **Addressee** are classified as [Sentient]. Similarly, VerbAtlas specifies **Object** and **Instrument** as [entity], while FrameNet assigns the semantic type [Message] to the frame element **Message**. Relations between frame elements are delineated in FrameNet: the pairs {**Message**, **Topic**} and {**Medium**, **Speaker**} constitute core sets, allowing either frame element or both to appear within a sentence. In both representations, defining sets of verb-noun combinations poses challenges, whether employing automatic or corpus observational methods, due to the abstract level of both semantic types and selectional preferences and inherent difficulties in reconciling figurative but acceptable usage.

In the context of Brazil FrameNet, each core frame element undergoes analysis based on the aspect of the scene it represents, leading to the assignment of one or more frames to the frame element (Torrent et al. 2022: 4 – 5). Only frames that signify events, states, attributes, and relations qualify for frame element-to-frame relations. The information available from the core frame element definition or semantic type is used to determine the type of concept it refers to (e.g., *people*, *place*, *event*) and to identify the top-level frame representing it. In this way, the FrameNet is enriched with additional semantic information by linking the conceptual structures that built it. The approach appears akin to the introduction of morphosemantic relations between verb and noun synsets in WordNet; however, this extension is not applicable as nouns have not yet been introduced in the Bulgarian FrameNet.

Some conclusions can be drawn as follows: The definition of noun fillers for frame elements uses ontological representations of abstract entities, some of which come from WordNet or can be linked to WordNet. However, due to the relatively high level of abstraction (at most 250 semantic types are used to classify nouns), accurate prediction of noun fillers for frame elements (semantic types for verb patterns' slots; selectional preferences for semantic roles) is fraught with challenges. First, generalised concepts tend to encompass a wide range of words, some of which may not be semantically compatible with a particular lexical unit. Second, the WordNet noun hierarchy is constructed for a different purpose, which may lead to hypernym-hyponym trees containing synsets belonging to different semantic classes or to different ontological classes, i.e. concrete and abstract nouns.

6. Towards an Extended Ontology of Semantic Classes of Nouns

The accepted approach involves selecting the highest-ranking synset, or a combination of synsets, from the (Bulgarian) WordNet that encompasses all suitable noun synsets for the frame element fillers (Koeva 2020: 17; Koeva 2021: 184 – 185). Such an approach facilitates the development of a robust training dataset for automatically labelling nouns and their semantic classes as frame element instances, enabling valence pattern annotation and the assignment of noun fillers to frame elements. In contrast, the manual annotation process in FrameNet, which entails identifying valence patterns and potentially extracting sets of noun fillers from annotated examples, is more labor-intensive. Additionally, it lacks automatic classification of noun classes due to the limited size of sense annotated corpora (with disambigated senses), such as the Bulgarian sense annotated corpus (Koeva 2012).

As noted, WordNet categorises nouns into broad semantic classes, which may not adequately reflect the semantic preferences of diverse range of verbs. Additionally, multiple hypernymy in WordNet arises from the consolidation of diverse taxonomic relations into a single hypernymy. To tackle these challenges, we propose: a) linking WordNet synsets to more detailed ontological representations of noun semantic classes to improve 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 Mapping ontological representations to WordNet noun hierarchy

Expanding the 25 semantic classes of WordNet involves linking WordNet synsets to various hierarchies of semantic categories. Such sub-categorisation was achieved through a manual mapping process, aligning CPA semantic types with the corresponding WordNet synsets (Koeva et al. 2018). As a result, the WordNet semantic classes are categorised into 253 CPA semantic types, wherein hyponym noun synsets inherit both the semantic class and the semantic type of their superordinates.

For example, the synset {beverage; drink} is marked both as noun.food and [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 noun.food and [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. The taxonomic organisation facilitates inheritance between semantic classes and types along the hierarchy, ensuring a more precise delineation of verb-noun compatibility.

Other ontological representations that are additionally mapped to WordNet synsets are those of the VerbAtlas selective preferences and the FrameNet semantic types.

The comparison between the semantic types in the CPA and the selective preferences in the VerbAtlas reveals extensive overlap, with only a few unique selective preferences found in the VerbAtlas but not in the CPA. For instance, the selective preference [Liquid] exists in both the CPA and the VerbAtlas and is already mapped to the WordNet synset {liquid} with the semantic class noun.substance. Only seven selective preferences are not present in the CPA, and they were additionally added in the Extended ontology of semantic classes of nouns. For example, the selective preference [Shot] is mapped to the synset {stroke, shot}. The mapping process was facilitated by VerbAtlas's explicit referencing to BabelNet synsets and, as a result, to WordNet synsets.

The hierarchy of semantic types in FrameNet (Ruppenhofer et al. 2016: 86) is also aligned with CPA semantic types and VerbAtlas selective preferences. Given the abstract nature of semantic types in FrameNet, only four additional types are included in the Extended ontology of semantic classes of nouns. For example, the semantic type [line] is linked to the synset {line} with the semantic class noun.shape. In the mapping process, any terminological differences used to denote the same concepts were standardised.

The initial formation of the Extended ontology of semantic classes of nouns involves mapping the relevant ontologies to the hierarchy of WordNet nouns, which enables the selection of nodes indicating the appropriate noun fillers for particular frame elements. As a result of this mapping, the original 25 semantic classes in WordNet were previously enriched with additional 253 semantic types from the CPA and now with additional 7 selective preferences from VerbAtlas and 4 semantic types from FrameNet. The alignment of the semantic types in FrameNet, the semantic types in CPA and the selective preferences in VerbAtlas to the same WordNet synset serves as an indicator of an appropriate level of abstraction for selecting suitable noun fillers when annotating FrameNet frame elements. The Extended ontology of semantic classes of nouns can be enhanced with new, more specific classes if the number of annotated verbs and frame elements increases and requires further characterisation.

6.2. 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.

Based on the premise that a synset should only link to a single hypernym, additional semantic relations (**origin**, **form**, and **function**) are defined in the context of multiple hypernyms (Koeva, Hristov 2023). This enables the redefinition of some of the hypernymy relations and the resolution of the occurrence of multiple hypernyms. By distinguishing different semantic relations, we can use only the **is-a** inheritance relation to divide the semantic classes of noun synsets into more specific groups and predict the compatibility of verbs and nouns.

6.3. Further identification of noun fillers based on corpus analysis

For the set of verbs evoking a particular semantic frame, or their instances in the Bulgarian FrameNet, called conceptual frames (Koeva 2020: 17; Koeva 2021: 184 – 185), suitable examples are sought in the Bulgarian National Corpus or in other sources that illustrate the realisation of the frame elements at the syntactic level, and thanks to an annotation showing the combination of the frame elements and their syntactic realisation for a particular lexical unit, the so-called valency patterns of the lexical units are constructed.

However, 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 activity is performed by annotators who select WordNet nodes that may be representative for the set of potential fillers and check whether the concept is already included in the Extended ontology of semantic classes of nouns; whether the subordinate noun synsets are suitable for pairing with the concrete verb as the syntactic realisation of its frame element; and decide whether to use, for the definition of noun fillers, a single concept from the Ontology, a set of concepts, or define a new concept and align it with the corresponding WordNet synset.

To illustrate the proposed approach, let us consider the relevant noun fillers for the frame elements of the **Statement** frame evoked by the verbs *обяснявам* and *обясня* ‘explain’, defined as ‘представям нещо в разбираема и достъпна форма’ (WordNet: make plain and comprehensible). The core frame elements are defined in FrameNet as follows: ‘The **Speaker** is the sentient entity that produces the Message (whether spoken or written)’; ‘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 **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***; ‘The **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*.

The noun fillers for the frame element **Speaker** are either nouns classified with the semantic class noun.person in WordNet or non-sentient nouns whose meaning can express unions of people such as *party, ministry, organisation, company*, etc., denoting organisations responsible for specific functions, policies, or services. In this context, such nouns embody abstract concepts of administrative authority, policy formulation, regulatory oversight, etc., referring not to physical, tangible entities but to the collective functions and responsibilities associated with human activities. The concepts in the Ontology are represented a) by the WordNet synset eng-30-02472293-n {human, human being} and are aligned with the semantic type [Human] in CPA and FrameNet (although the frame element is specified as [Sentient]) and the selective preference [human] in VerbAtlas; b) by the WordNet synset eng-30-08008335-n {organisation} and it is aligned with the semantic type [Institution] in CPA and [Organisation] in FrameNet and the selective preference [social group] in VerbAtlas. The same selection applies to the frame element **Addressee**.

The most obvious observation about the frame element **Message** fillers is that they should be nouns classified as noun.communication or noun.cognition in WordNet. However, these nouns differ in how they express communication and cognition. Therefore, it is important to develop a technique to eliminate the nouns that cannot be colocated with the verb *explain* as direct objects. The synset {communication} with the definition 'something that is communicated by or to, or between people or groups' is the highest in the hierarchy of nouns marked with the semantic class noun.communication. However, this meaning is too abstract to appear as a filler for the frame element **Message**. Although the hyponyms of this synset are appropriate in some cases, some inappropriate nouns appear in the respective subtrees: {receipt} – 'an acknowledgment (usually tangible) that payment has been made'; {mail} – 'the bags of letters and packages that are transported by the postal service'; and {publication} – 'the communication of something to the public; making information generally known', among others. Some of the non-combinable nouns within the subtree are marked with the semantic class noun.act. Other non-combinable nouns have the semantic class noun.communication, but are concrete. In this case, a working strategy is to restrict the set to nouns of semantic class noun.communication and add another level of classification: abstract and concrete nouns.

The synset at the top of the hierarchy, with the semantic class noun.cognition, is {cognition; knowledge}, defined as 'the psychological result of perception, learning, and reasoning'. However, not all of its hyponyms are compatible with the frame element **Message**. To address this, two concepts are selected: eng-30-05816287-n {information} and eng-30-05833840-n {idea; thought}, both also defined in the VerbAtlas.

For the frame element **Topic**, the concept represented by the synset eng-30-00002137-n {abstraction; abstract entity} 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 synset or combination of synsets in the WordNet noun hierarchy that semantically dominate appropriate nouns.

Introducing additional elementary semantic types and classifying nouns based on these types so that 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. Additionally, the mapping with the VerbAtlas selectional preferences and FrameNet semantic types highlights the significance of high-rated semantic classes. The alignment of these resources underscores the relevance of certain semantic classes. Furthermore, it illustrates how the ontology of description-relevant semantic classes remains open to augmentation.

7. Conclusions

The development of modern technologies and generative large language models makes it possible to predict the next word in relation to the previous context with great accuracy (especially for English, but increasingly also for other languages). This development raises the question of the need for a classification of classes of nouns in terms of their combinability with verbs. According to the author of this study, the necessity is justified by the following factors:

The presented approach to classification is primarily useful for the in-depth analysis and study of a particular language (as a mother tongue or a foreign language), as it illustrates not only the compatibility of words but also their ontological and independent class grouping based on shared semantic properties.

The classification can also help fine-tune large language models so that they correctly predict the compatibility of verbs and nouns not only in common contexts but also in rarer and more unusual ones.

The goal is to identify classification models that enable effective description of noun classes, achievable either by experts, automatically, or by a combination of both methods.

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]

Cinkova, Hanks 2010: Cinkova, S., P. Hanks. *Validation of Corpus Pattern Analysis – Assigning Pattern Numbers to Random Verb Samples*. <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. – 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, China: Association for Computational Linguistics, pp. 627 – 637. <https://aclanthology.org/D19-1058/> [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 1982: Fillmore, Ch. Frame semantics. – In: *Linguistics in the morning calm*. Seoul: Hanshin Publishing, pp. 111 – 137.

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

Fillmore, Baker 2010: Fillmore, Ch., 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.

Hanks 2004: Hanks, P. Corpus pattern analysis. – In: *Proceedings of the 11th EURALEX International Congress* (EURALEX 2004). Lorient: Université 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 française de linguistique appliquée*, 2 (Vol. X), pp. 63 – 82. <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2005-2-page-63.htm> [20.01.2024]

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

Koeva 2019: Коева, Св. Комплементите в български. – В: Мичева, В., Д. Благоева, С. Колковска, Т. Александрова, Хр. Дейкова (отг. ред.). *Доклади от Международната годишна конференция на Института за български език „Проф. Любомир Андрейчин“, София, 14 – 15 май 2019 година*. София: Издателство на БАН „Проф. Марин Дринов“, с. 57 – 68. (Koeva, Sv. Komplementite v balgarski. – In: Micheva, V., D. Blagoeva, S. Kolkovska, T. Aleksandrova, Hr. Deykova (Eds.). *Dokladi ot Mezhdunarodnata godishna konferentsiya na Instituta za balgarski ezik „Prof. Lyubomir Andreychin*“*, Sofia, 14 – 15 may 2019 godina*. Sofia: Izdatelstvo na BAN „Prof. Marin Drinov“, pp. 57 – 68.)

Koeva 2020: Koeva, Sv. Semantic Relations and Conceptual Frames. – In: Koeva, Sv. (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, Sv. 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, Hristov 2023: Koeva, Sv., 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. 2018: Koeva, Sv., Ts. 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.

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]
Ruppenhofer et al. 2016: Ruppenhofer, J., M. Ellsworth, M. Petruck, Ch. Johnson, C. Baker, J. Scheffczyk. *FrameNet II: extended theory and practice*. Berkeley, California: International Computer Science Institute.
Torrent et al. 2022: Torrent, T., E. Matos, Fr. Belcavello, M. Viridiano, M. Gamonal, Al. da Costa, M. 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]

КЪМ ДЕТАЙЛНИ СЕМАНТИЧНИ 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 представи подход за ефективна класификация на съществителните имена, улесняваща спецификацията на синтагматичните комбинации между глаголи и съществителни. Предлага се кратък преглед на семантичните класификации на съществителните в Уърднет, Корпусния анализ на изреченски модели и Фреймнет от гледна точка на съчетаемостта между глаголи и съществителни. Съотнасянето на Уърднет със семантичните типове на Корпусния анализ на изреченски модели, с Фреймнет и със селективните предпочитания на Върбатлас позволява създаването на Разширена онтология на семантичните класове на съществителните. Приносът на разработката се състои в детайлизиране на стъпките за подбор на набор от семантични класове, чиито членове са подходящи за реализация на определени фреймови елементи.

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://doi.org/10.7546/ConfIBL2024.36