Creating Hierarchical Relations in a Multilingual Event-type Ontology

Zdeňka Urešová, Eva Fučíková, Jan Hajič
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University
{uresova,fucikova,hajic}@ufal.mff.cuni.cz

Abstract

This paper describes the work on hierarchization of the SynSemClass event-type ontology. The original resource has been extended by a hierarchical structure to model specialization and generalization relations between classes that are formally and technically unrelated in the original ontology. The goal is to enable one to use the ontology enriched by the hierarchical concepts for annotation of running texts in symbolic meaning representations, such as UMR or PDT. similar

The hierarchy is in principle built bottom-up, based on existing SSC classes (concepts). This approach differs from other approaches to semantic classes, such as in WordNet or VerbNet. Although the hierarchical relations are similar, the underlying nodes in the hierarchy are not.

In this paper, we describe the challenges related to the principles chosen: single-tree constraint and finding features for the definitions of specificity/generality. Also, a pilot inter-annotator experiment is described that shows the difficulty of the hierarchization task.

1 Introduction

The SynSemClass (SSC) multilingual[1] event-type ontology (Uresova et al., 2020; Urešová et al., 2023b) is a lexical-semantic resource that links similar resources, such as FrameNet (Baker et al., 1998; Fillmore et al., 2003), WordNet (Miller, 1995; Fellbaum, 1998), VerbNet (Schuler, 2006) and others, and unifies them under a single scheme.

Each entry in SynSemClass (Urešová et al., 2023b), a *class*, corresponds to one eventive *concept* (state or process). Every concept is specified in multiple ways so that the human reader can understand what the concept is. The following are the main features describing a class, e.g., *kill* (Fig. 1):

- the prototypical **name**, e.g., *kill* stands for the event type *killing*),
- a brief **class definition** (in all languages), which characterizes the common meaning of all synonymous **class members** contained in it, e.g., *A Cause deprives a Victim of life*,
- a fixed set (a **Roleset**) of defined "situational participants" (**"semantic roles"**), e.g., *Cause, Victim, etc.*,
- each class member is further **linked** to one or more existing syntactic or semantic lexical resources for each language (as referenced above, e.g., to WordNet entries),
- each class member is **exemplified** by instances of real texts (and their translations to English) extracted from translated or parallel corpora,[2] e.g., *This is not only because it kills the unborn.*

The organization of this paper is as follows: Sect. 2 explains why we have decided to build the hierarchy, and in Sect. 3 we mention other works on this topic. In Sect. 4, our approach to hierarchical scheme is presented, Sect. 5 describes some challenging issues (Sect. 5.1) and tools used (Sect. 5.2). Sect. 6 discusses the current state of the hierarchy with some statistics. We conclude and draw future plans in Sect. 7. Sect. 8 in the extra space lists the limitations of the current state of the hierarchy.

2 Motivation

Although SynSemClass is a resource that is meant to be used in document annotation (perhaps in addition to or on top of another meaning representation scheme, such as Uniform Meaning Representation (UMR) (Bonn et al., 2024)), such annotation would

[1] English, Czech, German and Spanish.

[2] Such as the Prague Czech-English Dependency Corpus (https://ufal.mff.cuni.cz/pcedt2.0/en/index.html) the Paracrawl corpus (http://paracrawl.eu), and the XSRL dataset (https://catalog.ldc.upenn.edu/LDC2021T09), among others.

Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025), pages 240–249
July 31, 2025

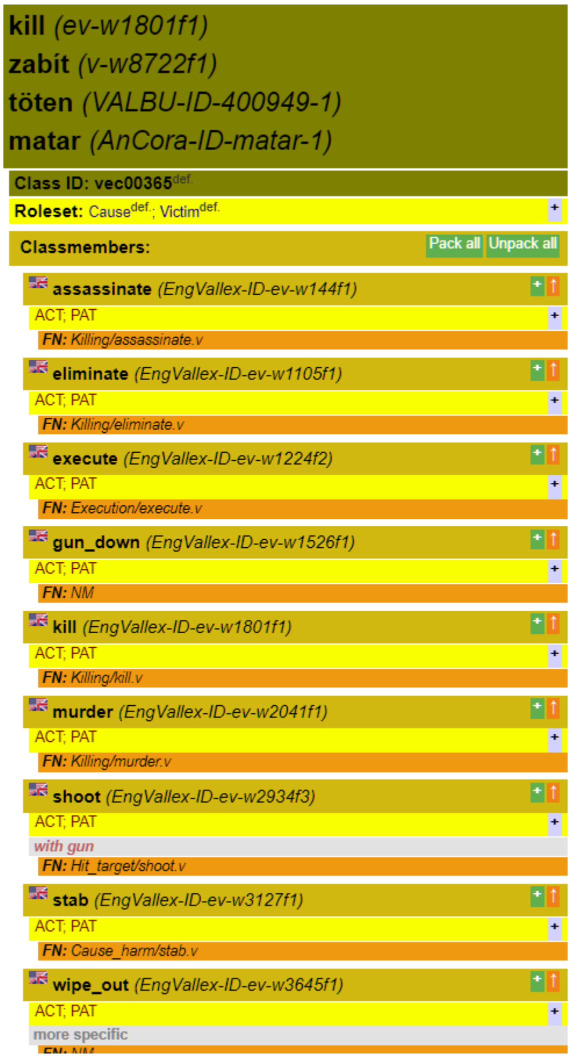

Figure 1: The abbreviated example of the SSC class *kill*.

be very difficult to perform accurately and efficiently given the properties of the SynSemClass ontology as described in the previous paragraph (Urešová et al., 2023a).

The problem is the unrelatedness of the different classes in the ontology: in the hypothetical (but certainly not uncommon) case that the annotator sees an expression (verb, noun, and MWE) that is not found among the class members of any class (or is found, but it is used in a new or different sense clearly not corresponding to the concept of the class in which it is found), *all* the classes would have to be considered, one by one, to find a suitable one (or determine that it does not exist in the resource).[3] There are now 1500+ classes in SynSemClass - so this is unfeasible to do efficiently.

Therefore, we have determined that a hierarchy over the concepts (as represented by the classes) in SynSemClass is necessary. The existence of such a hierarchy, connecting all the classes by a generalization/specialization relation, would reduce the effort required to find the appropriate class in the hierarchy by going top-down and selecting an appropriate hierarchical node (and the class represented by (linked from) it) in just a few steps.

However, given the existence of hierarchies integrated in other resources, one might ask why to build a new one. We have had two main reasons: first, the underlying SynSemClass resource is richer than the aforementioned ones in terms of being multilingual (or "interlingual") from the start, build bottom up, interlinked to other resources, has explicit mappings to syntactic resources in the languages it refers to, and has exemplification based on real corpora. Second, when inspecting the links to resources with similar hierarchies (WordNet, FrameNet, VerbNet) included in SynSemClass, there was often a multiple number of possible generalizations.[4] While the differences might be due to a different view on the synset/class concept, it is clear that there is no simple way to get a common hierarchy.

That is why we are exploiting the gap and trying to fill it; the main novelty is the complexity of the linked resources in the combined resource, that is, the hierarchy plus the data in the underlining ontology. We believe that both the actual creation and the use for textual annotation in the future can benefit from this complex information, which can guide annotators' understanding of the concepts in the hierarchy. In addition, this approach combines the "bottom-up view ", built within the SynSemClass ontology itself, with the top-down view when starting with the top-level ontology, as most current approaches do.

We are aware of the fact that such a hierarchy cannot be fully built in a simple tree-shaped form. However, we do believe that the core of such hierarchical set of relations can, despite the fact that the individual languages might sometimes have incompatible tendencies in expressing hyperonymy and

[3] One can imagine a better way of pre-annotation, namely the use of current state-of-the-art technology, such as LLMs. However, even that assumes at least some data to be fully annotated manually, if only for the development and evaluation of such tool(s).

[4] When going from SynSemClass to WordNet to hyperonym synset in WordNet and back to SynSemClass, there have been over 3 suggested possible generalization classes on average. For example, for the SynSemClass *propose*, there are five different top-level aligned WordNet semantic classes (communication, social, possession and cognition), with 7 different synsets suggested as direct hyperonyms.

hyponymy. The fact that the underlying ontology stress concepts rather than lexical (syn)sets should help, since all the context (links to entries in the other resources, including WordNet), syntactic and semantic properties present at each entry, can be taken into account when considering the often conflicting grounds for determining the hierarchical structure.

At the same time, if this hierarchy exists, SynSemClass could also serve other purposes, such as enabling a comparison to other lexical resources and their hierarchies thanks to the rich linking scheme within SynSemClass, linguistic and cognitive research on generalization and specialization, or language acquisition.

3 Related Work

The work described here relates closely to other lexical resources that include information about hierarchical relations among concepts, for example, WordNet (Fellbaum, 1998) or FrameNet (Baker et al., 1998).

The Princeton WordNet (PWN) is a large lexical database of English that groups words into interrelated sets of cognitive synonyms (synsets) and that is organized as a network where the synsets' relations are encoded through a super-ordinate relation (hyponymy/hyperonymy). PWN represents a concept as lists of the word senses that can be used to express the concept. Verb synsets also add the relation of troponymy in such a way that the nodes at the bottom of the tree denote specifications of a more general event (Fellbaum, 2005; Miller and Fellbaum, 2007). The multilingual EuroWordNet (Pianta et al., 2002; Ellman, 2003) introduced some major design changes, among them new semantic and lexical relations that may be specific to individual languages [5] (Vossen, 1998; Vossen et al., 1998; Tufis et al., 2004). In addition, a framework for a 'Global Grid' was established that defines a universal core lexical inventory and establishes guidelines for its cross-linguistic encoding (Fellbaum and Vossen, 2007).

FrameNet, a resource containing information about lexical and predicate argument semantics, is based on the principles of frame semantics, where frames (conceptual structures that describe different types of entities, situations and events) are organized into a network where more abstract frames (*super-frames*) are connected to less abstract frames (*sub-frames*). These relations include, but are not limited to: *Inheritance* - the relationship between a parent frame and its child frame; *Using* (or weak-inheritance) - the relation between a frame that is related in some way to a super-frame; *Subframe* - a relation between a complex frame that denotes a sequence of states and transitions and the individual frames that separately denote each state; and *Perspective* - the relation between frames denoting different perspectives over a neutral frame and the neutral frame itself. In addition to the hierarchy of frames arranged according to the frame-to-frame relations, FrameNet works with the second hierarchy, i.e., hierarchy of semantic types, which indicates the basic types of fillers of frame elements, marks non-lexical types of frames, and records important semantic differences between lexical units belonging to the same frame (Materna, 2014 [cit. 2024-11-14]).

Various proposals have been put forward to align the information contained in both resources aiming at the development of an ontology of events. BabelNet (Navigli and Ponzetto, 2010) is a prime example. For Slavic languages specifically, (Leseva and Stoyanova, 2022) set the foundations for the development of an ontology of stative predicates in Bulgarian and Russian by elaborating on FrameNet hierarchical classification through its mapping with WordNet.

Another example of an ontology that integrates information from lexical resources (with upper-level ontologies such as DOLCE (Borgo et al., 2022)) is The Rich Event Ontology (Brown et al., 2017), which provides a structure of event concepts connected at various levels of specificity and establishes relations between events and between events and the key objects and participants involved.

There are other ontologies, but as far as we know, there is no multilingual synonyms ontology with a hierarchical scheme built bottom-up. i.e., as in SynSemClass, with so much empirical material available for determining the hierarchical relations with much higher certainty (than WordNet(s)' only lexically-based synsets). We also have to stress here that the multilingual wordnets are developed top-down working with a shared set of so-called Base Concepts and an equivalence relation for each synset to the closest concept from an Inter-Lingual-Index. The general approach of EuroWordNet is to build wordnets mainly from existing resources (Vossen et al., 1998; Vossen, 2002). Compatibil-

[5] Currently, WNs exist for some 40 languages, see http://www.globalwordnet.org.

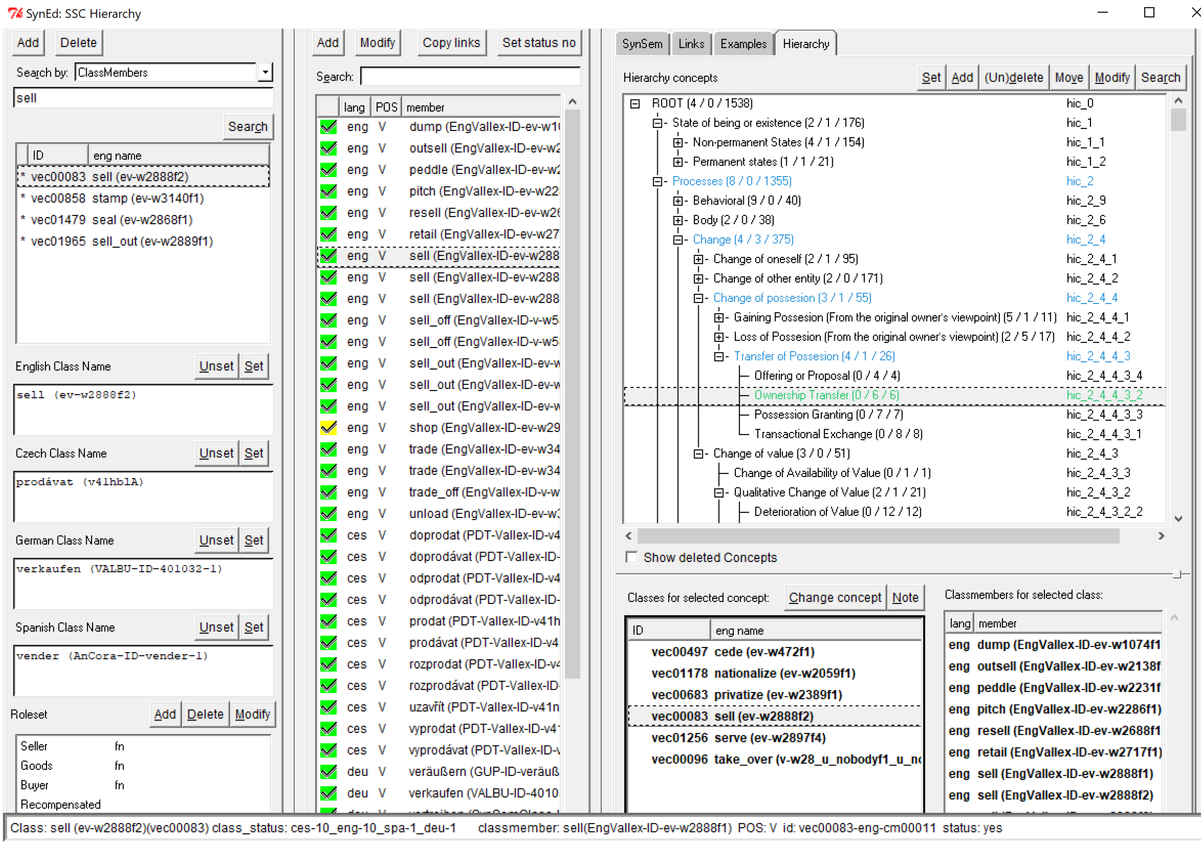

Figure 2: The hierarchical concept Ownership Transfer (abbreviated; shown in the editing tool)

ity between the EuroWordNet languages and the Inter-Lingual-Index with respect to lexical coverage and relations depends on which of the two basic methods for building the European wordnets was followed: either English synsets are translated into the target language and the relations are copied (Expand method), or synsets are created for the target language, interlinked with the PWN relations, and subsequently translated into English for mapping with ILI entries (Fellbaum and Vossen, 2007). For the discussion of near-synonymy, there are both theoretical lexicographic works such as (Lyons, 1968), and also the computationally-oriented view by (Edmonds and Hirst, 2002).

4 The Hierarchy

We have conceived the hierarchy as a single, rooted tree, in which ideally the SynSemClass classes are assigned 1:1 to its nodes and where the edges represent the *more general* or *more specialised* conceptual relation between the parent and the child nodes in the tree.

However, after testing a few examples, it was clear that this is not feasible to do directly, for the same reasons that the direct use of SynSemClass with its flat, set-like structure would be inefficient to use for annotation. Looking at any concept, the question that was not easy to answer was “which concept might be the next more general one among all the other SynSemClass concepts?” - without going through every other class. In Sect. 5 we explain how we proceeded, using some preprocessing to extract some candidates for these relations.

As a working solution, we have decided to scrap the 1:1 requirement of linking the hierarchy nodes to SynSemClass classes for now and temporarily allow both empty nodes in the hierarchy, as well as nodes with multiple SynSemClass classes assigned to them, to be split later. However, each SynSemClass class is (perhaps also temporarily) linked to *only one node* in the hierarchy to maintain at least some structure in it. We believe that this is not limiting at this time.

Having done so, we have to distinguish the original SynSemClass concepts as represented by the set of class members (verbs or nouns) in its flat structure (in this paper, we will call them **syc**s), and the nodes in the hierarchy tree (**hic**s, for hierarchical concepts).

Each **hic** (node in the hierarchy tree) is characterized by a series of features, or descriptors, as illustrated by the example in Fig. 2, for the **hic Ownership Transfer**:

- **definition**: *Refers to the complete shift of ownership or control from one party to another*,
- **mapping (linking) between the hic and syc(s)**: vec00497 (*cede*), vec01178 (*nationalize*), vec00683 (*privatize*), vec00083 (*sell* - highlited), vec01256 (*serve*), and vec00096 (*take_over*),
- **roleset(s)** coming from the **syc**(s) mapped: *Seller*, *Goods*, *Buyer* and *Recompensated*,[6]
- **class members from the classes mapped**, e.g., *dump, outsell, peddle, pitch, resell, retail, sell*,
- **example sentences** coming from **syc**(s) again (invisible on Fig. 2),
- its **parent** (more general concept) **hic** node: *Transfer of Possession*.

All of these parts constitute a complex description of **hic** (hierarchical concept). They serve (similarly to the SynSemClass class features and descriptors, as we see them) primarily for human understanding of the concepts.

We have created the base hierarchical structure (Sect. 5). To verify the approach fully, we have linked each class in the ontology (illustrated, e.g., in Fig. 1) to a node in the hierarchy.

The top level of the hierarchy is shown schematically in Fig. 3;[7] for the **hic** *Possession or Ownership*, we are showing the full expanded path (internal nodes in light blue) to this **hic** (which is a leaf in the hierarchy tree, shown in light green).

5 Building the Hierarchy

5.1 Issues of Full Hierarchization

The main identified problem is the very definition of the relation between **hic** s. At the beginning, we

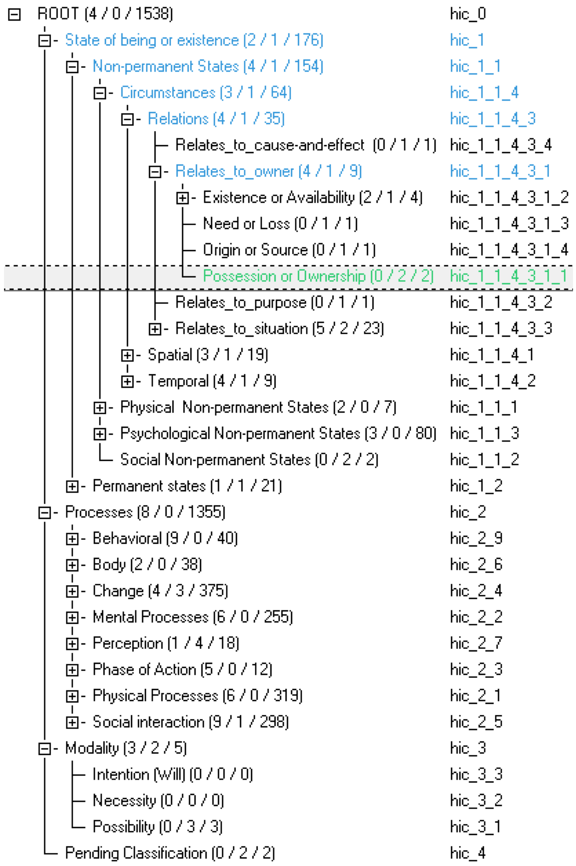

Figure 3: The tree w/path to Possession or Ownership

have intentionally abstained from using some predefined relation type(s), such as those from the Linguistic Linked Open Data (LLOD),[8] other Semantic Web ontologies, or even from the existing resources such as WordNet's hyponymy/hyperonymy (even though our idea was closest to this). Instead, we have been testing various node splits as we went along, refining the top-level hierarchy of essentially states vs. processes down the (sub)tree(s) being split from the root to the (current set of) leaves. We still see this relation as closest to "specialization" (of a higher-level concept in the hierarchy tree towards the lower-level one); the opposite direction would then be called "generalization."

Building such a hierarchical tree seems to be as difficult as categorization of things in the real world. The backbone of our scheme is the classification of real-world event types as states and processes. Since the resource used for our hierarchy, the SynSemClass ontology, represents the **syc**s concepts by a single class with a number of

[6] So far one **hic** may contain more rolesets, but ideally there should be only one, for the only class that should remain linked to (sec. 7).

[7] We are aware of the fact that *Modality* and *Phase of Action* (under *Processes*) are concepts that do not correspond to any **syc** "by definition" since SynSemClass does not cover non-content concepts. However, in our opinion, it is necessary to include them for full compositionality in the textual annotation, similarly to *abstract predicates* in UMR (Bonn et al., 2024).

[8] A sketch of possible conversion of SynSemClass into the relations and schemas available in LLOD is provided in (Uresova et al., 2020), but no hierarchical relations are included in that schema(s).

possible realizations (class members, i.e., words) with a unified roleset containing the situational participants (semantic roles), we found it convenient to use this feature as a starting point to build the initial classification.

Some **syc**s seemed to be classified and grouped under one **hic** quite easily due to the same set of roles. For example, under the **hic** *Communication* initially included all the **syc**s with *Speaker, Audience_Addressee, Information*. However, sorting then all the classes that fell within *Communication* was no longer easy. The appropriate criteria for further splitting and sorting have to be found. Questions arose not only regarding which meaning is more general and more specific but also regarding the subtle semantic distinctions that could be used to categorize (split) the given **hic** in a more subtle way, such as in *Transfer message*, *Discussion*, *Request*, *Communicated relation*, and *Mode of Speaking*.

Analyzing the relationships between individual **syc**s was difficult mainly because it posed a challenge:

- to specify what the (more general) parent **hic** is, especially when no suitable **syc** for the parent node has been found,
- to determine which sorting criteria are the most relevant,
- to determine which feature (criterion) of the concept is preferred when splitting a **hic** with a number of **syc**s assigned to it,
- to specify how to distinguish the specialized semantic relations within one **hic** due to the different views on the distinctive criteria of meaning,
- to be consistent in applying the criteria.

Because some **hic**s overlap in certain features, distinguishing and classifying their meanings is particularly complex. For example, some might argue that **hic** *Change* and **hic** *Transformative* are much alike; however, we believe that this splitting has its merits, and they thus belong to different second-level concepts.

For example, verbs of motion might be divided into different sets of **hic**s according to the criteria used. One might prefer to use the criterion of *way of the movement* and distinguish the concepts of going vs. the concept of driving, but it is also possible to prefer the criterion of *speed* and classify the concept of running vs. the concept of crawling, or the criterion of *who does the motion*: *Self-Motion* (movement driven by the entity itself) vs. *Transport* (movement driven by external factors). In all cases, eventually we will be able to arrive at a full tree and employ all the criteria mentioned above, but the trees will differ substantially. The general criterion of explicability, simplicity, and linguistic adequacy should then be applied to determine the order of application of the criteria (i.e., at which level, which criterion shall be used).

Another example is whether an additional role in the Roleset can be used as criterion for a split into more specialised **hic**s (such as in the case of a general class "change" (roleset: (thing, person) `Changing`) vs. the more specialised class "overcome" (roleset: `Protagonist`, `Hindrance`)), or the opposite, when a role from the Roleset becomes "built-in" into the more specialised class (such as in the case of the general class describing transport with roles `Transporter`, `Transported`, `Area_1`, `Area_2`, with a more specialised sub-**hic** *Setup Placement* (with class "plant" with its roleset `Transporter`, `Transported`, `Place`), which removes `Area_1` given that it is irrelevant to plant something. Another example of specialization is positivity vs. negativity: Loss vs. Gain, Improvement vs. Deterioration; granularity of cause (concepts of Contamination or Pollution vs. Water- and Liquid-induced damage), and several others.

These splitting criteria might differ between higher-level concepts. For example, while the difference in actor-caused (or actor-less) movement can prevail for the concepts of motion, for mental concepts, the "manner" criterion might prevail.

5.2 Tools Used

We have used an open source editor that was used in version 5.0 of SynSemClass[9] by adapting it - adding one additional tab to its editing canvas which shows the hierarchy as created so far and allows for assigning a **syc** to any **hic** in the hierarchy. It also allows for editing the **hic** tree by moving nodes around, adding new ones, and deleting them; definitions can also be added to its nodes.

To aid in creating the **hic** nodes of the hierarchy tree, we have also created a preprocessing tool that suggests **syc**s (i.e., the original SynSemClass

[9] `https://github.com/fucikova/SynSemClass_multi/tree/main/Editor`

classes) that appear to be semantically close enough to form either a subtree in the hierarchy, or the cluster could be used when considering a new general concept unifying them. The tool uses the sharing of semantic roles assigned to the classes and other hints to propose the clustering. Its results are collected in a table to aid in the effort to form the **hic** tree as a side resource.

6 Current State

All the classes (**syc**s) from SynSemClass have been assigned to the tree nodes of the conceptual hierarchy tree nodes (**hic**s). There are 1538 **syc**s in the version of SynSemClass that we have been working with. The current hierarchy has 663 nodes; this means that there are around 2.5 classes (**syc**s) per node in the hierarchy. This is still far from the goal of having (close to) 1:1 correspondence between **hic**s and **syc**s, but a larger number of nodes than many existing hierarchies currently have. In this section, we present some quantitative indicators.

6.1 Statistics and Description of the Hierarchy

The top level of the hierarchy (just under its root) has three branches[10] (Fig. 3):

1. States of Being or Existence: 139 nodes in total; they describe "static" concepts (existence, position, qualities, possession, mental states, etc.), linked from 176 **syc**s in total.

2. Processes: 518 nodes in total, describing processes (as opposed to states, as in the previous branch). There are 1355 **syc**s linked to these **hic**s, clearly indicating that there are still many split candidates in this branch, however typically with only 2-3 classes in them;

3. Modals: 4 nodes in total, describing modalities that are to be used as full concepts in textual annotation; given the SynSemClass principles, there are no classes that can be assigned to such "modality" concepts, except for five (e.g., *have a choice* in the "possibility" sense). This set of **hic**s will in fact need more work, since the **syc**s required to be linked to might not fit the philosophy of concepts in SynSemClass (which excludes auxiliaries, modals, copulas, etc.). Nevertheless, we believe that we need to have independent concepts for modals, phase-denoting and some "light" verbs, given the meaning they convey, which is then combined with the "content" eventives when annotating running texts.

[10] Pending Classification is meant for undecided classes yet, so this branch is an artificial node only.

A total of 35 conceptual nodes in the hierarchy tree have no class assigned to them yet, but they were introduced to keep the hierarchy tree fully connected (and might be populated later).

6.2 Structure of the Hierarchy Files

The current version of SynSemClass is 5.5;[11] For complete reproducibility, we also include the version used for the work that led to this paper.[12] After unpacking, there is

- File `hierarchy-tabular.xlsx`: tabular form of the hierarchy tree, one **hic** per row, sorted by the ID (column C). The hierarchy node name is in column A. In column B, the following statistics on **hic** are posted: number of sub-**hic**s, number of classes in **hic** and number of all classes in **hic** s within the subtree rooted in the current one.

- The XML files that represent both the SynSemClass version used and the proper hierarchy (`synsemclass_hierarchy.xml`).

7 Conclusions and Future Work

We have created a novel hierarchy of eventive concepts linked to an existing event-type ontology, SynSemClass. Each its class is linked to one node in the hierarchy. The hierarchy is a fully connected rooted tree, currently containing 663 **hic**s, with about 2.5 SynSemClass classes linked to each **hic**.

We have identified problems that arise while building such a hierarchy: defining each concept clearly, finding criteria for splitting nodes into its child nodes when multiple possibilities exist, and finding a set of SynSemClass classes representing each concept (node in the hierarchy) efficiently.

Perhaps not surprisingly, the existing resources do not consistently define its entries, as demonstrated by the multiplicity and fuzziness of relation mappings between these resources (using SynSemClass links). The hierarchies in these resources also differ substantially (FrameNet's vs. WordNet's hyponymy/hyperonymy relation vs. the shallow VerbNet hierarchy).

[11] http://hdl.handle.net/11234/1-5915

[12] https://github.com/ufal/SynSemClassHierarchy/tree/main/Lexicons-LAW-XIX-2025

No. of judgments	Both agree	1 annotator only (avg.)	IA agreement
50	20	26.5	28
100%	40%	53%	56%

Table 1: Gold data and inter-annotator agreement for assigning a class to the hierarchy tree

All of this poses a challenge for the refinement of the hierarchy over SynSemClass as we have developed it so far, in several respects:

- the hierarchy nodes which map to multiple SynSemClass classes must be split, after suitable criteria are identified for where to do the split, especially for nodes with a large number of classes;[13]
- the child nodes of **hic**s with no **syc** currently mapped to must be investigated in detail, to find out if there is a mistake in the composition of the **syc**(s), and if a split of the **syc** could be done to create such a (more general) concept that would be suitable to link from the currently empty **hic**s (which entails modifying SynSemClass);
- test the hierarchy in "real life", i.e., to use it for annotation of text in such a setup that will make clear in which way, and what proportion of running real text can be done with SynSemClass alone and what need the hierarchy;
- consider adding semantic features (such as animateness, abstractness) to the nodes of the hierarchy, or even to the SynSemClass entries themselves, to represent distinctions which did not make it into the hierarchy itself as a criteria for specialization.

We have performed a pilot annotation comparison (annotator agreement experiments) for the (re)assignment of 50 classes to the current hierarchy tree (Table 1). Two annotators independently assigned classes to the hierarchy, and the result was compared to the gold annotation and also between them.

The numbers indicate low accuracy against the data when annotators also agree, and only slightly above 50 percent accuracy for each of the two independently, and between themselves. This is to be expected since it is a very hard task, both mentally and from the statistical point of view (the random uniform baseline is 1/663). But it is an approximation of the text annotation task, since the SynSemClass classes (**syc**s) to be assigned to the hierarchy nodes (**hic**s) correspond, by and large, to the verb senses that text annotators will have to determine during such annotation, which will also serve as the relevant test and evaluation experiment.

The current full version of the hierarchy is published in a new version of SynSemClass (v5.5).[11] Nevertheless, there is still work to do, such as split some of the leaves of the hierarchy tree, populate some nodes with new links to the SynSemClass classes, and refine the concepts definitions.

8 Limitations

As is usual with any introspective approach in semantics in general and ontology work in particular, albeit supported by multiple lexical and corpus resources, the major limitation is our ability to understand the distinctions in the concepts we try to hierarchize and distinguish.

It might be the case that the fully connected tree constraint that we have chosen at the start is eventually untenable.[14] However, unless we specify the full hierarchy, no conclusions can be drawn.

Another limitation is that SynSemClass coverage needs to be improved (Fučíková et al., 2024).[15] In addition, the work on some abstract concepts, like modalities and concepts represented often by phase-denoting and some light verbs (i.e., concepts that take other eventives as arguments), has not been finished. Some SynSemClass classes would need to be rearranged to populate some internal **hic**s.

Finally, we acknowledge that this is work in progress and that additional work on splitting the remaining concepts in the hierarchy that are linked to more than one SynSemClass entry is needed. However, having the 663 current **hic**s assigned and structured in the hierarchy was, as we believe, the hardest part, both on the top levels and providing enough problems to solve at the more detailed levels down the hierarchy. The rest should go much more smoothly, despite the criteria selection problem discussed in Sect. 5.1.

[13] Ongoing work in progress: 217 additional hierarchy nodes are under evaluation and verification, and will appear in the final version.

[14] There are both cognitive and technical arguments in the literature; even WordNet does not follow this restriction, at least technically.

[15] It has not been used for annotation yet, except for small experiments (Urešová et al., 2019).

Acknowledgements

The work described here has been started while the authors have been supported by the Grant Agency of the Czech Republic under the EXPRO program as project “LUSyD” (grant No. GX20-16819X), and finished under the MŠMT international cooperation project "UMR" (grant No. LUAUS23283) and the LINDAT/CLARIAH-CZ Research Infrastructure (projects LM2018101 and LM2023062, supported by the Ministry of Education, Youth and Sports of the Czech Republic). The authors would like to thank all the annotators of the SynSemClass ontology and especially the ones who helped with the pilot hierarchy inter-annotator agreement experiment, Anna Staňková and Tomáš Razím. We would also like to thank the reviewers for their insightful comments.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Julia Bonn, Matthew J. Buchholz, Jayeol Chun, Andrew Cowell, William Croft, Lukas Denk, Sijia Ge, Jan Hajič, Kenneth Lai, James H. Martin, Skatje Myers, Alexis Palmer, Martha Palmer, Claire Benet Post, James Pustejovsky, Kristine Stenzel, Haibo Sun, Zdeňka Urešová, Rosa Vallejos, and 4 others. 2024. Building a broad infrastructure for uniform meaning representations. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2537–2547, Torino, Italia. ELRA and ICCL.

Stefano Borgo, Roberta Ferrario, Aldo Gangemi, Nicola Guarino, Claudio Masolo, Daniele Porello, Emilio M. Sanfilippo, and Laure Vieu. 2022. DOLCE: A descriptive ontology for linguistic and cognitive engineering1. *Applied Ontology*, 17(1):45–69.

Susan Brown, Claire Bonial, Leo Obrst, and Martha Palmer. 2017. The Rich Event Ontology. In *Proceedings of the Events and Stories in the News Workshop*, pages 87–97, Vancouver, Canada. Association for Computational Linguistics.

Philip Edmonds and Graeme Hirst. 2002. Near-synonymy and lexical choice. *Computational Linguistics*, 28(2):105–144.

Jeremy Ellman. 2003. Eurowordnet: A multilingual database with lexical semantic networks: Edited by Piek Vossen. Kluwer Academic Publishers. 1998. isbn 0792352955, 179 pages. *Natural Language Engineering*, 9:427 – 430.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA. 423 pp.

Christiane Fellbaum. 2005. Wordnet and wordnets. In Keith Brown, editor, *Encyclopedia of Language and Linguistics*, pages 2–665. Elsevier.

Christiane Fellbaum and Piek Vossen. 2007. Connecting the universal to the specific: Towards the global grid. In *Intercultural Collaboration*, pages 1–16, Berlin, Heidelberg. Springer Berlin Heidelberg.

Charles J. Fillmore, Ch. R. Johnson, and M. R. L.Petruck. 2003. Background to FrameNet: FrameNet and Frame Semantics. *International Journal of Lexicography*, 16(3):235–250.

Eva Fučíková, Cristina Fernández-Alcaina, Jan Hajič, and Zdeňka Urešová. 2024. Textual coverage of eventive entries in lexical semantic resources. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15835–15841, Torino, Italy. European Language Resources Association.

Svetlozara Leseva and Ivelina Stoyanova. 2022. *Stative verbs: Conceptual structure, hierarchy, systematic relations*, pages 68–114. Prof. Marin Drinov Publishing House of BAS.

J. Lyons. 1968. *Introduction to Theoretical Linguistics*. Cambridge University Press.

Jiří Materna. 2014 [cit. 2024-11-14]. *Probabilistic Semantic Frames [online]*. Doctoral theses, dissertations, Masaryk University, Faculty of Informatics, Brno. SUPERVISOR : Karel Pala.

George Miller and Christiane Fellbaum. 2007. Wordnet then and now. *Language Resources and Evaluation*, 41:209–214.

George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM*, 38(11):39–41.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. Multiwordnet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*.

Karin Kipper Schuler. 2006. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Dan Tufis, Dan Cristea, and Sofia Stamou. 2004. BalkaNet: Aims, methods, results and perspectives – A general overview. *Romanian Journal of Information Science and Technology Special Issue*, 7:9–43.

Zdeňka Urešová, Cristina Fernández-Alcaina, Eva Fučíková, and Jan Hajič. 2023a. SynSemClass Czech and English Annotation Guidelines. Technical Report 73, UFAL MFF UK.

Zdeňka Urešová, Eva Fučíková, Cristina Fernández Alcaina, and Jan Hajič. 2023b. Synsemclass 5.0. available from the lindat/clariah-cz digital repository. http://hdl.handle.net/11234/1-5230.

Zdeňka Urešová, Eva Fučíková, Eva Hajičová, and Jan Hajič. 2019. Parallel Dependency Treebank Annotated with Interlinked Verbal Synonym Classes and Roles. In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*, pages 38–50, Paris, France. Université Paris Sorbonne Nouvelle, Association for Computational Linguistics.

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

Piek Vossen, editor. 1998. *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*. Kluwer.

Piek Vossen. 2002. Wordnet, EuroWordNet and Global WordNet. *Revue Française de Linguistique Appliquée*, 7.

Piek Vossen, Laura Bloksma, and Horacio Rodriquez. 1998. The EuroWordNet Base Concepts and Top Ontology. Workingpaper, Vrije Universiteit.