Classifying Verbs in WordNet by Harnessing Semantic Resources

Svetlozara Leseva
Institute for Bulgarian Language – BAS
zarka@dcl.bas.bg

Ivelina Stoyanova
Institute for Bulgarian Language – BAS
iva@dcl.bas.bg

Maria Todorova
Institute for Bulgarian Language – BAS
maria@dcl.bas.bg

Abstract

This paper presents the principles and procedures involved in the construction of a classification of verbs using information from 3 semantic resources – WordNet, FrameNet and VerbNet. We adopt the FrameNet frames as the primary categories of the proposed classification and transfer them to WordNet synsets. The hierarchical relationships between the categories are projected both from the hypernymy relation in WordNet and from the hierarchy of some of the frame-to-frame relations in FrameNet. The semantic classes and their hierarchical organisation in WordNet are thus made explicit and allow for linguistic generalisations on the inheritance of semantic features and structures.

We then select the beginners of the separate hierarchies and assign classification categories recursively to their hyponyms using a battery of procedures based on generalisations over the semantic primes and the hierarchical structure of WordNet and FrameNet and correspondences between VerbNet superclasses and FrameNet frames. The so-obtained suggestions are ranked according to probability. As a result, 13,465 out of 14,206 verb synsets are accommodated in the classification hierarchy at least through a general category, which provides a point of departure towards further refinement of categories.

The resulting system of classification categories is initially derived from the WordNet hierarchy and is further validated against the hierarchy of frames within FrameNet. A set of procedures is established to address inconsistencies and heterogeneity of categories. The classification is subject to ongoing extensive manual verification, essential for ensuring the quality of the resource.

1. Introduction

This paper outlines the principles and procedures involved in the elaboration of a hierarchical classification of verbs through the mapping of three semantic resources – WordNet, FrameNet and VerbNet. The classification is induced from the relational hierarchies of WordNet and FrameNet and derives its categories from FrameNet frames and VerbNet superclasses which are assigned to WordNet synsets. More specifically, we adopt the FrameNet frames as the primary categories of the classification and transfer them to WordNet synsets through a set of procedures involving either (i) exact mapping or (ii) generalisations over the hierarchical structure of WordNet and FrameNet and correspondences between VerbNet superclasses and FrameNet frames. The hierarchical relationships between the categories are projected both from the hypernymy relation in WordNet and from the hierarchy of some of the frame-to-frame relations in FrameNet.

The classification exploits previously interconnected resources in a way that enables the study and use of structured representations of salient semantic and syntactic properties as realised in the hierarchical verb lexicon, the validation of semantic and syntactic generalisations derived from each of these

Keywords: WordNet, FrameNet, VerbNet, semantic frames, verb classification

resources against the data encoded in the other resources, the mutual enhancement and the expansion of coverage through generalisations over combinations of features of the different resources.

2. Linguistic prerequisites

The creation and mutual integration of complementary lexical semantic resources have presented a great interest in the research community. Notable efforts on mapping semantic resources include the work of Shi and Mihalcea on mapping WordNet, FrameNet and VerbNet (Shi and Mihalcea, 2005), Tonelli and Pighin's enrichment of FrameNet with WordNet mappings (Tonelli and Pighin, 2009), the system Semlink (Palmer, 2009) that unites these three resources with PropBank, and its follow-up Semlink+ that brings in mapping to Ontonotes (Palmer et al., 2014). While these efforts give rise to databases of integrated semantic knowledge, most of them deal with mapping of the units of the original resources to each other. Much less attention has been paid to exploring and exploiting the internal structure of these resources, especially with respect to how these structures relate and correspond to each other, how they can be mapped to each other, etc.

WordNet (Miller, 1995; Fellbaum, 1998b) is a large lexical database which represents conceptual and lexical knowledge in the form of a network whose nodes represent cognitive synonyms (synsets) interlinked through a number of conceptual-semantic and lexical relations. While the relations of hypernymy/hyponymy that provide the basic hierarchical organisation of verbs and nouns in WordNet are explicit, the membership of a synset to a hypernym tree only gives a very general idea about the semantic class to which this synset's members belong. Consider the tree dominated by the root change:1, alter:1, modify:3 (cause to change; make different; cause a transformation). We can only infer that the subordinate members of the tree denote some kind of change brought about by an entity and affecting another entity regardless of the depth of the hierarchy at which a particular synset is found.

On the other hand, more detailed classificatory features emerge from the mappings with VerbNet verb classes and superclasses and FrameNet frames. Both resources provide linguistic abstractions that either specify or translate into semantic classes of different granularity and different level of generalisation and in addition propose a certain hierarchical organisation.

FrameNet frames represent conceptual structures describing particular types of objects, situations, etc. along with their participants, or frame elements, FEs (Baker et al., 1998; Ruppenhofer et al., 2016). As such, they are abstract representations of lexical units that lexicalise these situations or objects. Though not exhaustively, FrameNet frames are related into a network through frame-to-frame relations part of which also provide a hierarchical organisation, the most prominent being Inheritance in which the child frame is a subtype of the parent frame, e.g. Change_of_temperature and Proliferating_in_number inherit from Change_position_on_a_scale. Other relations include Using, Perspective on, Subframe, Precedes, Inchoative_of, Causative_of, etc.

We adopt the FrameNet frames as the primary categories of the proposed classification which are to be further explored and enhanced. The hierarchical relationships between these categories are projected both from (i) the hypernymy (troponymy) relation in WordNet and (ii) the hierarchies formed by 2 frame-to-frame relations in FrameNet. As an illustration of (i), consider the hypernym–hyponym pair: change:1, alter:1, modify:3 > heat:1, heat up:2 where the verbs are mapped to the frames Cause_change and Cause_temperature_ change, respectively, which are adopted as classification categories. Given that, we posit a relation of hierarchy between the frame-derived classification categories: Cause_change > Cause_temperature_change. The FrameNet hierarchies are employed in augmenting the coverage of the mapping between frames and synsets as described in detail in Section 5.

The VerbNet (Kipper-Schuler, 2005) classes, which represent explicit natural groupings of verbs with shared semantic and syntactic properties, are structured in a shallow hierarchy of types (herein called superclasses), classes and subclasses (if any). Superclasses unite classes related to a particular type of eventualities, e.g. 'Verbs of putting', 'Verbs of removing', and provide semantically grounded linguistic generalisations. They are employed in addition to the semantic information derived from the FrameNet frames to support the mapping between FrameNet and WordNet, as well as to help resolving ambiguities and inconsistencies in the classification.

3. Related work and motivation of the proposal

While remaining a less explored area, FrameNet's frame structure and frame-to-frame relations has been employed in various domains, such as text understanding (Fillmore and Baker, 2001), semantic analysis (Burchardt et al., 2005), generation of lexical entailment rules (Coyne and Rambow, 2009; Aharon et al., 2010), paraphrase extraction (Hasegawa et al., 2011), construction of event ontologies (Palmer et al., 2014), role linking of implicit semantic arguments (Li et al., 2015).

Research into the enhancement of frame-to-frame relations has been proposed by a number of studies. Pennacchiotti and Wirth (2009) offer a definition of the notion of frame relatedness and different types of automatic measures to compute it. Ovchinnikova et al. (2010) describe a methodology for improving FrameNet's conceptual organisation through restructuring and axiomatisation of the frame relations. Frame relations have also been used in augmenting FrameNet's coverage with paraphrases (Rastogi and Van Durme, 2014).

Extension of frame relations has been another emerging area. Virk et al. (2016) propose a supervised model for enriching FrameNet's relational structure through predicting new frame-to-frame relations using structural features from the existing FrameNet network, information from the WordNet relations between synsets, and corpus-collected lexical associations. In devising the WordNet features, the authors employ similar logic to the one adopted in this paper by transferring the relational knowledge for pairs of related synsets to matching lexical units and frames in FrameNet. Botschen et al. (2017) present systems for predicting frame-to-frame relations based on text-based frame embeddings; the best-performing one uses the FrameNet hierarchy.

The research into verb classifications in terms of verbs' syntactic properties and behaviour (Levin, 1993; Pinker, 1989; Goldberg, 1994), among others, thematic structure (Chafe, 1970; Cook, 1979; Longacre, 1976; Foley and Van Valin, 1984; Van Valin Jr. and LaPolla, 1997), lexical conceptual structure (Gruber, 1965; Jackendoff, 1990), frame semantics (Fillmore, 1982) has culminated in resources such as VerbNet, FrameNet and WordNet and subsequent efforts at linking them in such a way as to maximise the merits of each resource (see Section 1).

Other researchers have brought together the ideas of a detailed description of verb semantic classes and a hierarchical classification of these classes. Hlaváčková et al. (2009) have proposed an outline of a verb ontology based on the data in the verb valency dictionary for Czech VerbaLex (Hlaváčková and Horák, 2005). A shallow hierarchy of Spanish verbs based on their semantic and syntactic properties has been described and implemented within the ADESSE project (García-Miguel and Albertuz, 2005).

The classification we propose aims at harnessing these already existing and interlinked resources while trying to bridge the divide between their use as a source of knowledge about the lexical items and natural semantic groupings and their combined potential to explore semantic and syntactic properties and generalisations, the distribution of these properties in the lexicon, the relations between these features, and so forth.

The classification combines information from the 3 outlined resources, each of which contains diverse lexical, semantic and syntactic information: detailed conceptual structures of lexical units with shared semantic and syntactic properties organised through frame-to-frame relations (FrameNet); more explicit verb semantics paired with a detailed representation of syntactic behaviour and linking between semantics and syntax (VerbNet); a hierarchical organisation of word senses within semantic domains connected in a lexicon-wide coverage network through a variety of conceptual and lexical relations (WordNet).

Due to its features, which are essential for the structuring of the proposed classification, WordNet is used as the base resource, providing the classification's backbone. FrameNet frames are used as the classification categories for the grouping of WordNet verbs into semantic classes and their taxonomic organisation. The application of VerbNet superclasses lends a new dimension to the mapping among the resources and enables the direct exploitation of knowledge from the other two resources.

4. Mapping the resources

The efforts at linking lexical resources generally suffer from limited coverage and compatibility issues due to multiple release versions of the original resources. This reduces considerably their applicability and further development. A feature of the proposed classification is that it attempts at resolving this shortcoming by translating semantically salient groupings into classification categories, which, if not as exhaustive as appropriately assigned frames or verb classes, provide feasible semantic generalisations.

The mappings and mapping procedures implemented have been adopted from the works of Baker and Fellbaum (2009), Shi and Mihalcea (2005) and Laparra and Rigau (2010).

VerbNet 3.3[1] provides an integrated mapping between members of verb classes and WordNet literals in synsets with corresponding senses. Using this direct mapping we have assigned verb classes to 4,885 out of 14,206 verb synsets in WordNet 3.0. Verb classes are further combined into superclasses (cf. VerbNet Annotation Guidelines[2]) which add a more abstract semantically grounded level of description. The membership to a particular superclass is indicated by a common class number: thus, 'Verbs of ingesting', assigned the number 39 include the classes: eat-39.1, chew-39.2, gobble-39.3, devour-39.4, dine-39.5, gorge-39.6, feeding-39.7. We have mapped the superclass names proposed in the Guidelines to the respective number and assigned them to the corresponding synsets.

With respect to the linking with FrameNet, we have employed two existing mappings. The SemLink project[3] distributes a many-to-many mapping of VerbNet classes and FrameNet frames where a FrameNet lexical unit may be mapped to more than one VerbNet verbs, and more often, a VerbNet verb is mapped to more than one frame. We use this mapping to assign frames to synsets indirectly by using the FrameNet to VerbNet mapping first, and then the VerbNet to WordNet mapping. In order to increase the limited coverage (only 2,630 verb synsets are assigned frames from FrameNet), a direct FrameNet-to-WordNet mapping called WordFrameNet (Laparra and Rigau, 2010)[4] is also applied using a set of intermediate mappings between release versions. As a result the number of verb synsets with assigned frames has increased to 3,134. WordFrameNet also provides frame assignments to over 4,000 words of other parts of speech (adjectives and nouns).

5. Building the classification

We have devised a number of procedures for the augmentation of the mapping coverage, which in the scope of the classification translates as increasing the number of classification categories and extending the coverage of the individual classification categories.

These procedures are aimed at: (i) discovering existing but unmapped relations between synset members and FrameNet frames; (ii) transferring frames between synsets through relations of inheritance derived from WordNet and FrameNet; (iii) adopting additional classification categories from VerbNet.

5.1. Extending the coverage using FrameNet frames

Nouns and verbs in WordNet are classified into distinct semantic domains that roughly correspond to semantic primitives, such as person, artifact, act for nouns or change, motion, cognition for verbs. This information is made explicit through special synset labels, such as noun.person, noun.artifact, noun.act, verb.change, verb.motion, verb.cognition. There are 25 such categories for nouns and 15 for verbs (Miller, 1998). Unlike nouns which form several hierarchies, each originating from a different unique beginner (Miller, 1998), verbs pertaining to a given semantic domain do not belong to a single hierarchy, but often represent several independent trees (Fellbaum, 1998a). The picture is more complex in reality as WordNet 3.0 has 559 verbs with no hypernym, of which 225 have no hyponyms (single verbs) and only 254 have more than 1 direct or indirect hyponym (Richens, 2008).

A tree often does not consist entirely of synsets that have the same semantic primitive, but one primitive is clearly predominant. As a very preliminary approximation at a viable classification criterion,

[1] `http://verbs.colorado.edu/verbnet/index.html`
[2] `https://verbs.colorado.edu/verb-index/VerbNet\_Guidelines.pdf`
[3] `https://verbs.colorado.edu/semlink/`
[4] `http://adimen.si.ehu.es/web/WordFrameNet`

we have sorted the trees according to the predominant prime.

As the frames of the beginners are (the most) abstract nodes in a particular hierarchy and their properties are inherited by the subordinate nodes, it is important that we: (i) identify the beginners within each semantic domain; (ii) check if they are mapped to frames, correct wrong mappings and assign a frame where the automatic procedure has failed; (iii) attach those roots that do not qualify as beginners to a given hierarchy.

5.1.1. Identifying the beginners in each semantic domain

There is no explicit information which root verbs represent the right level of abstraction so as to qualify as beginners of verb hierarchies, but the best candidates are those that are homonymous or near-synonymous with the semantic primes assigned to them and/or have a multitude of hyponyms, e.g., change:2 (undergo a change...) and change:1, alter:1, modify:3 (cause to change) (semantic prime: verb.change). Beginners are or should be assigned a FrameNet frame that is high in the frame-to-frame hierarchy. The second-best candidates are other trees of considerable size (in terms of number of members) and level of abstraction (in terms of the FrameNet frames assigned). If not identified as beginners, they should be attached to the tree of a beginner, preferably one with the same dominant prime. Single verbs and verb roots with a couple of hyponyms tend to be lower in the conceptual hierarchy and should be attached to a larger tree.

5.1.2. Checking of the mapping and manual assignment of frames

We have checked manually the mapping of the root verbs to FrameNet frames, focusing on the beginners as errors tend to propagate down the tree. For instance, during the automatic procedure described in Section 4 change:2 was mapped to the Cause_change frame instead of Undergo_change and so, if the mapping was left uncorrected, all its hyponyms would be classified in a wrong hierarchy.

The procedure for validation and assignment of frames to verb roots includes the following steps: (i) check the correctness of the assigned frame(s) if any – at this point the frames are assigned from the existing mapping distributions (cf. Section 4); if no frame is assigned (ii) check the frames in which the literals of the synset in question are found as lexical units and select an appropriate one; if no (appropriate) match is found, (iii) try to assign a suitable frame to the given synset using other information.

In particular, step (i) has resulted in the manual check of the frames of 1,300 synsets which includes: (a) frames that in the course of the analysis of the data have been found to be erroneously assigned; these have been individually corrected (total of 139); (b) frames which have been consistently assigned instead of other semantically related frames, e.g. Attaching covers examples of the frames Attaching, Becoming_attached and Detaching; all the synsets to which such frames have been automatically assigned have been analysed and priority has been given to the most suitable frame (total of 373 examples); and (c) synsets which have been assigned more than one frame; in such cases, either one of them has been selected or corrected, or in case more than one frame is considered suitable, they have been ranked (total of 788 examples).

In step (ii) and step (iii) the process has been facilitated by exploring semantic and structural features, such as:

(a) the frames of the synset's hyponyms – this is a check if a hyponym's frame (especially one predominant among the hyponyms) describes appropriately the conceptual structure of the verb synset in question. An additional verification procedure keeps track of whether verbs are consistently assigned causative or inchoative frames depending on their membership to one or the other class. This is done by verifying that in a tree dominated by a causative root the causative counterparts of homophonous or similar pairs of frames are assigned; and respectively, that in an inchoative tree the inchoative counterparts are mapped. This procedure is aimed at removing the errors in the automatic mapping;

(b) the mapping of the VerbNet (super)class to FrameNet – check if a frame homonymous or similar to the (super)class of a given synset describes appropriately the conceptual structure of the verb;

(c) the gloss – check whether the generic term which is elaborated in the definition is or may be mapped to a frame and if this frame describes appropriately the conceptual structure of the verb;

(d) structural features – these are based on the relations of the root synset to other verbs or nouns. In the former case, we check if the FrameNet frame of (i) a semantically related verb synset, such as an antonym or a causative/inchoative counterpart, or (ii) a direct or an indirect hypernym of such a synset describes appropriately the verb in question. Consider, for instance, back:4 (cause to travel backward), which is an antonym of advance:5, bring forward:1 (cause to move forward). The latter's hypernym is move:2, displace:4 (cause to move or shift into a new position or place, both in a concrete and in an abstract sense) which is mapped to the frame Cause_motion. Cause_motion is thus suggested as a mapping of back:4, as well.

Another strategy employs the connectedness of nouns in WordNet (all the nouns except for the root (entity:1) have at least one hypernym) and the relations between eventive deverbal nouns and derivationally related verbs. Thus, given the Event relation between fall:17 (lose one's chastity) and fall:5 (a lapse into sin...), we establish that the latter's hypernym sin:2, sinning:1 and its Event derivative sin:1, transgress:3, trespass:4 are mapped to the frame Misdeed, which is also a suitable mapping for fall:17.

Finally, if no appropriate frame exists, we posit a new classification category (and a frame) provided that it is predictable from the FrameNet's frame structure. For instance, while Motion is linked to Cause_motion, Self_motion (e.g. jump:1, leap:1, bound:1, spring:1 – move forward by leaps and bounds) does not have a causative counterpart to which verbs such as jump:11, leap:4 (cause to jump or leap) can be mapped, so we formulate one.

5.1.3. Attaching non-beginner roots into the hierarchy (hypernym assignment)

It is most likely to find both the beginners and suitable positions for attachment of smaller trees into a particular hierarchy among synsets that pertain to the same semantic domain (semantic prime) as the one we explore. Hypernym assignment involves techniques similar to the ones presented in Section 5.1.2. with certain differences. Thus we look at:

(a) the frame of the synset – check if a verb homonymous or similar to the frame assigned to a given synset or other verbs evoking this frame represent a suitable hypernym;

(b) the mapping of the synset to a VerbNet (super)class – check if a verb homonymous or similar to the assigned (super)class is a suitable hypernym;

(c) the gloss – check whether the generic term elaborated in the definition is a suitable hypernym;

(d) structural features – these are more or less the same as in (d) above, but in this case we check if a semantically related verb (a hypernym of an antonym, a derivative of a derivative's hypernym, etc.) is a suitable hypernym; these features also involve analysing a tree structure more globally.

Consider the root verb look:1 (perceive with attention; direct one's gaze towards): it is assigned the frame Perception_active and its definition contains as a generic term the verb 'perceive' so an obvious choice of hypernym is perceive:1, comprehend:2 (to become aware of through the senses). The tree structure of the latter synset confirms this choice as among its hyponyms one finds both active perception verbs such as listen:1 and passive perception verbs such as see:1 and hear:1.

As a result of the procedures in 5.1, incorrect frames assigned to the original set of roots have been replaced by appropriate ones, frames have been assigned to those roots for which the automatic mapping has failed and several new frames have been defined. Most of the original beginner roots have been assigned a FrameNet frame. In addition, those of them that could be successfully mapped to a larger tree have been assigned a suitable hypernym.

5.2. Procedures for automatic classification

We have developed 4 procedures for frame identification (STEPS 1–4) aimed at expanding the frame-to-synset coverage for the purposes of the classification, which we then apply recursively down the WordNet trees. We have also implemented a procedure for ranking candidate frames in order to facilitate their manual verification and selection.

5.2.1. Employing hierarchical relations in WordNet (hypernymy–hyponymy)

FrameNet frames may correspond to different hierarchical levels in the hypernym hierarchy in WordNet which results in the fact that lexical units evoking the same frame may be in a hypernymy relation to each other. For instance, amble:1, totter:2, wade:1 are found in hyponyms of walk:1, but all of them, including walk:1, evoke the frame Self_motion. This asymmetry in the grouping of lexical units and synsets has been used in mapping procedures, cf. Tonelli and Pighin (2009).

We therefore assume that, by default, a hyponym synset inherits the frame(s) of the hypernym synset. So as a first approximation, given a hypernym–hyponym pair, we consider the set of all the frames assigned to the hypernym and transfer them to the hyponym (STEP 1).

We also employ the hierarchical relations in FrameNet (frame-to-frame inheritance); more particularly, we make use of 2 hierarchical frame-to-frame relations which to a different degree correspond to the structure of WordNet. These are: Is_Inherited_by – the child frame is a subtype of the parent frame (e.g. Motion Is_Inherited_by Fluidic_motion); Is_Used_by – the child frame presupposes the parent frame as background (e.g. Theft Is_Used_by Robbery).

These two relations may be construed as potentially corresponding to the hypernymy relation: for instance, the lexical unit 'influence' evokes the frame Objective_influence which is inherited by the frame Manipulate_into_doing, and the latter is evoked by the lexical unit 'manipulate'. In WordNet the synset influence:1, act upon:1, work:15 (have and exert influence or effect) is the hypernym of manipulate:1, pull strings:1, pull wires:1 (influence or control shrewdly or deviously). Similarly, the lexical unit 'arrive' and 'reach' evoke the frame Arriving which is in the relation Is_Used_by with the frame Having_or_lacking_access, and the latter is evoked by the lexical unit 'access', among others. In the corresponding synset pair reach:1, make:22, attain:4, hit:4, arrive at:1, gain:4 (reach a destination, either real or abstract) is the hypernym of access:2, get at:1 (reach or gain access to).

Based on the above considerations, given a synset S, we select as a mapping of S any frame F if S contains a literal L coinciding with a lexical unit LU in FrameNet, such that LU evokes the frame F1, where F1 is assigned to the hypernym of S and F1 is related to F through one of the following set of frame-to-frame relations R{Is_Inherited_by, Is_Used_by} (STEP 2).

5.2.2. Combining hierarchical relations in WordNet and FrameNet

We also apply the reverse operation up the hypernym-hyponym tree so as to identify frames that generalise over the frames assigned to a synset's hyponyms. Thus, for each synset S with no frame assigned, we add to the mapping any frame F such that there is a synset S1 which is a direct hyponym of S and S1 is assigned the frame F1 where F1 is a direct descendant of F (STEP 3).

5.2.3. Employing VerbNet superclasses to identify possible frame candidates

We use the VerbNet to FrameNet mapping to identify indirectly frames to be assigned as classification categories to synsets based on the correspondences between the VerbNet superclasses and FrameNet frames. For each synset S with an assigned VerbNet class and thus, superclass C, but no frame assigned, we add to the mapping any frame F such that there is a synset S1 which has been assigned both superclass C and frame F (STEP 4).

5.2.4. Ranking of candidate frames

After the implementation of the above procedures, we apply ranking to all new candidate frames in order to identify the most likely frames and to facilitate their manual verification and selection. The scoring system reflects the source of the candidate: score ranges indicate the procedure via which the frame has been assigned, including information whether the assignment has been made through the application of more than one procedure, which increases the probability of a candidate being a valid suggestion as a classification category for that synset. The actual ranking then combines these into a single score and orders the candidates in decreasing order.

The ranking is based on the following factors:

(i) Priority is given to direct over indirect inheritance of frames from hypernym to hyponyms;

(ii) The score of frames assigned by more than one procedure is a sum of the individual scores given to the frame individually via each procedure;

(iii) Frame score is adjusted with respect to the collective frequency of the literals belonging to the respective synset found in the candidate frame's lexical units list.

Example 1. presents the result of the assignment and ranking of frames to the synset spy:3, sight:1.

Example 1. eng-30-02163746-v spy:3, sight:1 'catch sight of; to perceive with the eyes'
WordNet semantic prime: verb.perception
VerbNet class: sight-30.2 (Verbs of Perception)
Ranking of possible FrameNet frames in descending order: Perception (104); Becoming_aware (100); Awareness (100); Perception_experience (100); Categorization (100); Sensation (20); Perception_active (20); Intentionally_act (1);

The similar scores of Perception, Becoming_aware, Awareness and Perception_experience reflect the similarity among these conceptual structures in the FrameNet hierarchy: Perception is inherited by Becoming_aware and Perception_experience. The values of these particular frames correspond very neatly to the WordNet structure, as well: perceive:1, comprehend:2, the hypernym of spy:3, sight:1, is mapped to Perception_experience and most of the hyponyms of spy:3, sight:1 are mapped to Becoming_aware. These two frames are therefore very likely candidates.

In addition, while Awareness and Becoming_aware are not linked through a relation, it is obvious that such a relation needs to be specified as Becoming_aware (referring to a Cogniser's adding a Phenomenon to their model of the world) is the inchoative counterpart of the stative Awareness (referring to a Cogniser's having a piece of Content in their model of the world)[5], so this merits further analysis.

6. Classification of verbs in WordNet

6.1. Resulting classification of verbs in WordNet

Using the above procedures, we obtain a detailed classification of verbs in WordNet based on a number of distinctive features: (a) verb semantic primes; (b) FrameNet frame(s); (c) WordNet hierarchical relations; and (d) VerbNet classes and superclasses. The classification categories (FrameNet frames) are assigned within each general semantic category of verbs (as defined by their semantic prime) and form a hierarchy determined by the WordNet hypernymy/hyponymy hierarchy (where the respective synsets have been assigned FrameNet frames in the existing mappings between the resources, cf. Section 4) in conjunction with the hierarchy of FrameNet frames based on the Is_Inherited_by / Inherits_from and Is_Used_by / Uses relations and the VerbNet (super)class assignments to fill the gaps in the cases where frames have not been assigned to synsets in the existing mappings. These procedures align well, as the inheritance between frames roughly follows the route from more general to more specific conceptual structures, which corresponds to the branching of the WordNet hierarchy from more general to more specific senses.

Initially, there are 3,134 classified synsets (i.e synsets assigned a frame as a classification category through the mapping procedure described in Section 4) out of a total of 14,206 verb synsets in WordNet. Using the algorithm in Section 5, we have been able to assign at least one classification category derived from the inventory of FrameNet frames to 10,331 synsets. At STEP 1, by transferring frames from the hypernym to the hyponym, 10,217 candidate frames have been assigned. Another 1,192 candidate frames have been assigned at STEP 2, and further 1,311 frames have been mapped at STEP 3. More than half of the synsets (5,606) are assigned more than one candidate frame, about a third (3,012) – more than two candidate frames, and about 10% (981) – more than five candidate frames.

The candidate frames for each synset obtained through the different procedures are ranked according to their scores in descending order. A total of 1,210 candidate frames have been assigned using more than one STEP rendering them more probable and bringing them forward in the ranking. The results are then subjected to manual validation so as to ensure high quality of the assignments, including through removing errors stemming from the automatic FrameNet to WordNet mapping. The classification is available at: `http://dcl.bas.bg/en/verb-classification/`.

6.2. Towards the validation and streamlining of the classification

Proper automatic validation of the method is not possible as there is no gold standard against which to evaluate the classification. We propose partial validation of the resulting structure against the relational

[5]cf. `https://framenet2.icsi.berkeley.edu/`

structure of FrameNet. The comparison is focused on the consistency between the hierarchy of categories (frames) automatically built from the relational structure of WordNet and the hierarchy on which FrameNet itself is constructed. Consistency is measured in terms of identical or near-identical paths in the two hierarchies according to a particular relation: the hypernymy/hyponymy relation in WordNet and the Inheritance and Uses relations (and their combinations) in FrameNet. Each category is evaluated in relation to its direct parent in the hierarchy and is labeled either: (a) OK – the direction of inheritance of the frames coincides with the direction of the hypernym-to-hyponym relation in WordNet (Example 2a); (b) UPPER – the direction of inheritance of the frames is reversed (Example 2b); or (c) DIFFERENT – the frame appears elsewhere in the FrameNet hierarchy (Example 2c).

Ideally, the two structures coincide (with possible omissions or additions of nodes in either of them) which means that for a pair of categories in the classification where F1 (evoked by synset S1) is a parent of F2 (evoked by synset S2) and S1 is a hypernym of S2, F1 is a direct or an indirect parent of F2 in FrameNet (F1 Is_Inherited_by / Is_Used_by F2).

Example 2.
(a) OK
eng-30-01835496-v travel:3; go:16; (verb.motion, Motion) is hypernym of
eng-30-01886488-v slither:1; slide:3 (verb.motion, Self_motion)
and the frame Motion Is_Inherited_by the frame Self_motion
(b) UPPER
eng-30-01463963-v arrange:4; set up:5 (verb.contact, Arranging) is hypernym of
eng-30-01543000-v drape:2 (verb.contact, Placing)
and the frame Placing Is_Inherited_by the frame Arranging
(c) DIFFERENT
eng-30-00983824-v utter:4; emit:2 (verb.communication, Communication_noise) is hypernym of
eng-30-01197208-v smack:4 (verb.consumption, Body_movement)
and the frame Communication_noise is not related to the frame Body_movement neither directly nor indirectly via the Is_Inherited_by / Inherits_from or Is_Used_by / Uses relation.

Deviations from the ideal as presented in (b) and (c) may be indicative of: (i) different logic underlying the hierarchical organisation of the conceptual structures and the synsets; (ii) inconsistencies or errors in the frame assignment or in the WordNet structure, e.g. mingling transitive/intransitive synsets in a single subtree or inconsistencies in the semantic prime and/or the WordNet hierarchy (as in (c)).

Further, in order to deal with the inconsistencies, an intermediate level of abstraction may be introduced in the hierarchy based on the (super)classes in VerbNet. This information can be used to combine heterogeneous examples in more general categories and to further ensure the validity of the classification.

7. Future development

An immediate task to be pursued is to identify inconsistencies in the data through exploring the inheritance among classification categories. Future work will be focused on elaborating techniques for increasing the depth of the classification hierarchy through maximising the weight of the similarity between the WordNet hypernym-hyponym structure and the FrameNet frame structure, through correlating VerbNet (super)classes with frames in such a way as to reflect the degree of semantic generality/specificity, etc.

Another venue of research is extending FrameNet's hierarchy with new frame relations and relation instances identified in the proposed WordNet classification, as well as mutually enhancing the two hierarchies on the basis of linguistic information retrievable from them.

Acknowledgements

This study has been undertaken within the project *Towards a Semantic Network Enriched with a Variety of Semantic Relations* funded by the National Scientific Fund of the Republic of Bulgaria under the Fundamental Scientific Research Programme (Grant Agreement No. 10/3/2016). We would like to thank three anonymous reviewers for their valuable comments.

References

Aharon, R. B., Szpektor, I., and Dagan, I. (2010). Generating Entailment Rules from FrameNet. In *Proceedings of the ACL 2010 Conference Short Papers, Uppsala, Sweden, 11-16 July 2010*, pages 241–246. Association for Computational Linguistics, Stroudsburg, PA, USA.

Baker, C. F. and Fellbaum, C. (2009). WordNet and FrameNet as Complementary Resources for Annotation. In *Proceedings of the Third Linguistic Annotation Workshop (ACL-IJCNLP '09), Association for Computational Linguistics, Stroudsburg, PA, USA*, pages 125–129.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *COLING-ACL '98: Proceedings of the Conference. Montreal, Canada*, pages 86–90.

Botschen, T., Mousselly-Sergieh, H., and Gurevych, I. (2017). Prediction of Frame-to-Frame Relations in the FrameNet Hierarchy with Frame Embeddings. In *Proceedings of the 2nd Workshop on Representation Learning for NLP, Vancouver, Canada, August 3, 2017*, pages 146–156. Association for Computational Linguistics, Stroudsburg, PA, USA.

Burchardt, A., Frank, A., and Pinkal, M. (2005). Building Text Meaning Representations from Contextually Related Frames – A Case Study. In *Proceedings of the Sixth International Workshop on Computational Semantics*.

Chafe, W. L. (1970). *Meaning and the Structure of Language* . University Press, Chicago.

Cook, W. A. (1979). Case Grammar: Development of the Matrix Model (1979-1978) . Georgetown University Press.

Coyne, R. and Rambow, O. (2009). Lexpar: A Freely Available English Paraphrase Lexicon Automatically Extracted from FrameNet. In *Proceedings of the Third IEEE International Conference on Semantic Computing*.

Fellbaum, C. (1998a). A Semantic Network of English Verbs . In Fellbaum, C., Ed., *WordNet: An Electronic Lexical Database*, pages 69–104. MIT Press, Cambridge, MA.

Fellbaum, C., Ed. (1998b). *WordNet: an Electronic Lexical Database*. MIT Press, Cambridge, MA.

Fillmore, C. J. and Baker, C. F. (2001). Frame Semantics for Text Understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*.

Fillmore, C. J. (1982). Frame Semantics . In *Linguistics in the Morning Calm*. The Linguistic Society of Korea, Soeul: Hanshin.

Foley, W. and Van Valin, R. (1984). *Functional syntax and Universal grammar*. CUP.

García-Miguel, J. M. and Albertuz, F. J. (2005). Verbs, Semantic Classes and Semantic Roles in the ADESSE project . In Erk, K., Melinger, A., and Schulte im Walde, S., Eds., *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes, Saarbrücken, 28 February – 1 March 2005*.

Goldberg, A. (1994). *Constructions: A Construction Grammar Approach to Argument Structure* . University of Chicago Press.

Gruber, J. (1965). *Studies in Lexical Relations, Doctoral dissertation, MIT, Cambridge, MA*. Also published in J. Gruber (1976) Lexical Structures in Syntax and Semantics, North-Holland, Amsterdam, 1–210.

Hasegawa, Y., Lee-Goldman, R., Kong, A., and Akita, K. (2011). FrameNet as a Resource for Paraphrase Research. *Constructions and Frames*, 3:104–127.

Hlaváčková, D. and Horák, A. (2005). VerbaLex – New Comprehensive Lexicon of Verb Valencies for Czech. In *Proceedings of the Computer Treatment of Slavic and East European Languages 2005, Bratislava, Slovakia*, pages 107–115.

Hlaváčková, D., Khokhlova, M., and Pala, K. (2009). Semantic Classes of Czech Verbs. In *Recent Advances in Intelligent Information Systems*, pages 207–217.

Jackendoff, R. S. (1990). *Semantic Structures*. Cambridge, Mass., The MIT Press.

Kipper-Schuler, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon. PhD Thesis*. Computer and Information Science Dept., University of Pennsylvania. Philadelphia, PA.

Laparra, E. and Rigau, G. (2010). eXtended WordFrameNet. In *Proceedings of LREC 2010*, pages 1214–1219.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago and London: The University of Chicago Press.

Li, R., Wu, J., Wang, Z., and Chai, Q. (2015). Implicit Role Linking on Chinese Discourse: Exploiting Explicit Roles and Frame-to-frame Relations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, July, 2015*, pages 1263–1271. Association for Computational Linguistics.

Longacre, R. E. (1976). *An Anatomy of Speech Notions* . Peter de Ridder Press.

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

Miller, G. A. (1998). Nouns in Wordnet . In Fellbaum, C., Ed., *WordNet: An Electronic Lexical Database*, pages 21–46. MIT Press, Cambridge, MA.

Ovchinnikova, E., Vieu, L., Oltramari, R., Borgo, S., and Alex, T. (2010). Data-driven and Ontological Analysis of FrameNet for Natural Language Reasoning. In *Proceedings of LREC 2010*.

Palmer, M., Bonial, C., and McCarthy, D. (2014). SemLink+: FrameNet, VerbNet and Event Ontologies. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929–2014), Baltimore, Maryland USA, June 27, 2014*, pages 13–17. Association for Computational Linguistics.

Palmer, M. (2009). Semlink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*. 9–15.

Pennacchiotti, M. and Wirth, M. (2009). Measuring Frame Relatedness . In Lascarides, A., Gardent, C., and Nivre, J., Eds., *Proceedings of EACL 2009*, pages 657–665. Association for Computer Linguistics.

Pinker, S. (1989). *Learnability and Cognition: The acquisition of argument structure*. MIT Press.

Rastogi, P. and Van Durme, B. (2014). Augmenting FrameNet Via PPDB . In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation, Baltimore, Maryland, USA*, pages 1–5. Association for Computational Linguistics.

Richens, T. (2008). Anomalies in the WordNet Verb Hierarchy. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, August 2008*, pages 729–736.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F., and Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California.

Shi, L. and Mihalcea, R. (2005). Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing. In Gelbukh, A., Ed., *Computational Linguistics and Intelligent Text Processing. CICLing 2005. Lecture Notes in Computer Science*, volume 3406. Springe, Berlin, Heidelbergr.

Tonelli, S. and Pighin, D. (2009). New Features for Framenet – Wordnet Mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL'09), Boulder, USA*.

Van Valin Jr., R. D. and LaPolla, R. J. (1997). *Syntax: Structure, meaning and function* . Cambridge University Press.

Virk, S. M., Muller, P., and Conrath, J. (2016). A Supervised Approach for Enriching the Relational Structure of Frame Semantics in FrameNet . In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan*, pages 3542–3552.