Comparing Sense Categorization Between English PropBank and English WordNet

Özge Bakay
Boğaziçi University, Turkey
ozge.bakay@boun.edu.tr

Begüm Avar
Boğaziçi University, Turkey
begum.avar@boun.edu.tr

Olcay Taner Yıldız
Işık University, Turkey
olcaytaner@isikun.edu.tr

Abstract

Given the fact that verbs play a crucial role in language comprehension, this paper presents a study which compares the verb senses in English PropBank with the ones in English WordNet through manual tagging. After analyzing 1554 senses in 1453 distinct verbs, we have found out that while the majority of the senses in PropBank have their one-to-one correspondents in WordNet, a substantial amount of them are differentiated. Furthermore, by analysing the differences between our manually-tagged and an automatically-tagged resource, we claim that manual tagging can help provide better results in sense annotation.

1 Introduction

The main challenge in lexical semantics is generated from 'polysemy', which refers to the phenomenon of a single orthographic word having multiple, interrelated senses. Only with the help of the context one can pin down the particular sense in which a word is used. A further challenge in the field stems from multi-word expressions, i.e. groups of words having "a unitary meaning which does not correspond to the compositional meaning of their parts" (Saeed, 1997). Hence, there are two central concerns for semantic analysis, centered around compositionality: (i) At the lexical level, choosing the correct sense of a given word within a context, and (ii) at the sentence level, determining how a particular combination of words should be interpreted.

Having semantic analysis of annotated corpora along with the syntactic architecture enhances Natural Language Processing (NLP) applications such as information retrieval, machine translation, information extraction, and question answering. Using the added semantic layer, syntactic parser refinements can be achieved, which not only increases the efficiency but also improves application performance. PropBank (Kingsbury and Palmer, 2002; Kingsbury and Palmer, 2003; Palmer et al., 2005) is one of the studies on this concept, widely accepted by computational linguistics communities.

In this paper, we present a sense category evaluation between English PropBank and WordNet. In order to compare the sense categories, we first manually disambiguate English verbs in the input sentences with sense tags from English WordNet. Then, these annotations are compared with sense annotations in English PropBank.

This paper is organized as follows: Since we compare senses in English PropBank and WordNet, we provide information about these resources in Section 2 and touch upon the related work about combinations of these resources in Section 3. The details of our sense-annotated corpus and how it is constructed are given in Section 4. We give the comparison details and statistics in Section 5, and differences between automatic vs. manual tagging in Section 6. Lastly, we conclude in Section 7.

2 Resources

PropBank is a corpus where predicate-argument information is annotated and semantic roles or arguments each verb can take are posited (Babko-Malaya, 2005). PropBank uses conceptual labels for arguments from Arg0 to Arg5. Only Arg0 and Arg1 indicate the same roles across different verbs, standing for Agent/Causer and Patient/Theme, respectively. The rest of the argument roles can vary across different verbs. For instance, the roles of the predicate "attack" from PropBank are as follows: Arg0 is "attacker", Arg1 is "entity attacked", and Arg2 is "attribute".

WordNet is a graph data structure where the nodes are word senses with their associated word forms and edges are semantic relations between

the sense pairs. The first WordNet project was Princeton WordNet (PWN) which was initiated in 1995 by George Miller, (Miller, 1995). Over time, PWN evolved to become a comprehensive relational representation of the word senses of English (Fellbaum, 1998). WordNet includes relations between synsets such as hypernym, instance hypernym, hyponym, instance hyponym, meronym, holonym, antonym, entailment, etc.

3 Related Work

Among many previous studies similar to ours is the one by (Pazienza et al., 2006), which aims to extract frame pairs by combining the lexical database of WordNet with the syntactic and semantic information given in VerbNet and semantically-annotated corpus of verbs in PropBank. Having inferred 989 frame pairs with troponymy, entailment, causation, and antinomy; they conclude that NLP applications can benefit from such repositories by making use of automatic or semi-automatic techniques to map arguments across the frames.

In another study, Kwon and Hovy (2006) first assigned the frame for each verb sense in WordNet from FrameNet and then, aligned roles among FrameNet, WordNet, and LCS depending on their mappings. In total, 4240 senses are linked with FrameNet frames, 674 of which are also linked with LCS, 1250 with PropBank, and 1757 with both.

SemLink (Palmer, 2009) is another project which aims to combine different information provided by various lexical resources (VerbNet (Kipper-Schuler, 2005), FrameNet (Baker et al., 1998), PropBank (Palmer et al., 2005) and WordNet (Fellbaum, 1998)). With mappings among these resources, the project aims to develop an NLP resource with extended overall coverage.

Aiming for interoperability among the same resources used in SemLink, López de Lacalle et al. ((de Lacalle et al., 2014a; de Lacalle et al., 2014b; de Lacalle et al., 2016a; de Lacalle et al., 2016b)) focus on predicates and try to develop a common semantic infrastructure, which is called the Predicate Matrix (PM). They define a set of methods, such as advanced graph-based word sense disambiguation algorithms and various corpus alignment methods to automatically achieve this integration. While they base their work on the central motivation of SemLink, the authors criticize the limitations of the manual methods used in developing the SemLink project and argue that "building large and rich enough predicate models for broad-coverage semantic processing takes a great deal of expensive manual effort" (de Lacalle et al., 2016b).

López de Lacalle et al.'s (2016b) work is definitely a major progress for NLP studies centering around predicate structure. Their approach, however, does not seem to put enough emphasis on the cognitive nature of language. Undoubtedly, manual tagging requires significant human effort, hence is more costly, and manually tagged corpora may be more limited in terms of systematicity and coverage. However, whether any analysis of language can be fully automatized is still a very skeptical issue. According to the approach adopted in the present study, the use of human annotators is considered worthwhile for developing corpora by focusing on semantic information. There may well be 'human errors', but overall, we believe that manual tagging still gives us better results in qualitative terms, though maybe not in quantitative terms.

4 Sense Annotation

4.1 Annotation Tool

The annotators in the present study use a custom application (written in Java) for browsing sentences and annotating them with senses. The toolkit is publicly available[1]. The current implementation of the application is designed to import the text files that adhere to the Penn Treebank data format. Once a sentence has been imported into the semantic editor, the human annotator is presented with the visualized syntactic parse tree of that sentence. Annotators can click on the leaf nodes corresponding to words. When a word is selected, a drop-down list is displayed, in which all the available WordNet entries of the selected word's lemma are listed.

Moreover, sense options whose POS (parts of speech) do not agree with the given word's POS, are disabled to optimize the task/help the annotators. Upon the selection of the most appropriate sense, the drop-down list is hidden and the ID of the submitted synset is displayed under the word. Figure 1 shows a screenshot from the system interface, depicting the screen presented to the an-

[1] https://github.com/olcaytaner/DataCollector

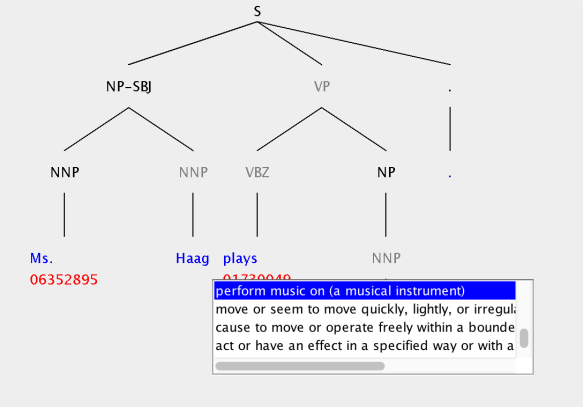

Figure 1: A screenshot from the system interface

notators when annotating the verb "plays" in the English sentence "Ms. Haag plays Elianti."

4.2 Sense Inventory

For the sense annotation, we use PWN version 3.1. Although PWN does not provide a web page for obtaining synsets and/or their relations, the data files are present. After retrieving the synset data files from the site, we constructed a WordNet XML file similar to the BalkaNet's (Stamou et al., 2002).

4.3 Extracting Candidate Sense List

For extracting senses, we only ask for the available senses of the English word in PWN. Complexities arise for verbs marked for third person (-s), gerund (-ing), past participle (-ed); and for adjectives in comparative (-er) or superlative (-est) forms. For those cases, we strip down the affixes and search for the root form in PWN. For irregular forms (such as irregular verbs) we use the exception list of PWN to get the root forms.

For collocated verbs, we just search for 2 or 3 word collocations in PWN with respect to the adjacent words of the current word. For instance, consider the sentence "They get up early". While showing the sense list of "get", we do not only show the sense list of "get" in isolation, but also add the senses of "get up" to that list.

4.4 The Comparison Process

For the comparison of sense categorization of verbs between English PropBank and English WordNet, a list of sentences (7576 sentences with 1554 senses of 1453 distinct verbs), all of which had been annotated by human annotators, was extracted. Instead of single-word annotations, we preferred to have the annotations of all the words appearing in those sentences. Two human annotators who are both graduate students in language-related departments were then provided with the list which displayed all the verbs alphabetically with all of the sentences they were used in. Seeing all the exemplary sentences of the verbs together is believed to have helped the annotators analyze the meaning differentiations within the verbs more closely (See Table 1). Before moving onto the comparison between the two sense categorizations, annotators also checked the accuracy of the annotated meanings of the verbs. This second step is considered to have strengthened the accuracy of sense annotation as well as the comparison of PropBank and WordNet. During the comparison stage, human annotators analyzed the two datasets to find out how similarly or differently the senses in English PropBank were reflected in English WordNet.

5 Comparison Details

In the comparison of the senses in English PropBank and English WordNet, what has been found out is that whereas most of the senses in WordNet seem to match the ones in PropBank, some of them do not. We came across these mismatches for three reasons; because of (i) the senses going under differentiation, (ii) the senses getting combined in English WordNet or (iii) the overlaps of senses in a given verb in that one sense of a verb in WordNet corresponds to more than one single sense of the same verb in PropBank.

In Sections 5.1 and 5.2, we will review one-to-one and one-to-many sense matchings between English PropBank and English WordNet, whereas in Sections 5.3 and 5.4 we will review many-to-one and many-to-many sense matchings between English PropBank and English WordNet.

5.1 One-to-one Sense Matches between English PropBank and English WordNet

The majority of the senses in English PropBank (1184 senses of 1118 different verbs) is found to match the ones in English WordNet. In other words, the sense categorizations in English PropBank seem to be retained in English WordNet once the senses are replaced with their WordNet equivalents. For example, as it is shown in Case 1, the sense of "abate" in PropBank, "to decrease, become less strong", is observed to be equal to "be-

Table 1: Example sense categorizations for English PropBank verbs

Case	Verb	PropBank Sense	WordNet Sense	Example
1	abate_01	to decrease, become less strong	become less in amount or intensity	The dollar posted gains in quiet trading as concerns about equities abated.
2	strengthen_01	(cause to) become stronger	gain strength	As Wall Street strengthened, the London trading room went wild.
	strengthen_01	(cause to) become stronger	make strong or stronger	In 1986, Congress strengthened the penalty by making it a felony.
3	absorb_01	suck up	assimilate or take in	Most dealers can not continue to absorb this supply.
	absorb_01	suck up	take up, as of debts or payments	Deal stocks led the market down as they absorbed the heaviest losses.
4	buy_01	purchase	accept as true	U.S. officials, however, said they are n't buying the Israeli argument.
	buy_01	purchase	obtain by purchase	Everybody was out buying Monets.
	buy_01	purchase	buy what had previously been sold, lost, or given away	So far, the company had bought back 1.6 million shares.
5	celebrate_01	honor, show respect to	have a celebration	The ads celebrate the achievements of some of Lake View 's residents.
	celebrate_02	have a party, occasion to mark an event	have a celebration	They don't even give a nod to human sensibilities by celebrating Halloween.
6	build_01	construct	make by combining materials and parts	A Taiwanese steelmaker recently announced plans to build a Nucor-like plant.
	build_01	construct	develop and grow	You built your career on prejudice and hate.
	build_02	grow	bolster or strengthen	Seagram says the promotion is designed to build brand loyalty rather than promote heavy drinking.
	build_02	grow	develop and grow	The great silver clouds on the horizon build themselves on the pale water.

come less in amount or intensity" in WordNet (See Table 1).

5.2 Sense Differentiations in English WordNet

A significant difference between the two sense categorizations is that in 352 senses of 329 different verbs, the senses given in English PropBank branch up to 12 distinct, and hence more specific, senses in English WordNet. Those differentiations may be meaning- or syntax-related. Regarding the syntax-related differentiation as indicated in Case 2, for example, in English PropBank, for the verb "strengthen", there is one sense for both transitive and intransitive forms, "(cause to) become stronger". However, in English WordNet, this sense is differentiated into two; "gain strength" for the intransitive and "make strong or stronger" for the causative, i.e. transitive form (See Table 1).

As an example for meaning-related differentiations, when we look at the verb "absorb" in Case 3, we see that whereas the only sense provided for it by Propbank is "suck up", two different senses for that same sense are given in WordNet; (i) assimilate or take in and (ii) take up, as of debts or payments. Although the former can be considered as the equivalent of "suck up", the latter indicates a different sense, which seems to be miss-

Table 2: Verbs with the highest number of senses

Verbs	Senses Annotated	Senses in WordNet
take_01	12	42
give_01, have_03, see_01	11	44, 19, 24
break_01, move_01	9	59, 16
come_01, know_01, turn_01	8	21, 11, 26
do_02, draw_02, find_01, lead_01, look_01, place_01	7	13, 36, 16, 14, 10, 16

ing in PropBank (See Table 1). It is also important to note that the number of the added senses in WordNet may vary depending on the verb. Table 2 shows the list of the 15 verbs with the highest number of senses. For instance, while PropBank lists only one sense for the verb "take", which is "take, acquire, come to have, choose, bring with you from somewhere, internalize, ingest", WordNet lists 42 different ones and in the current dataset, 12 of them were assigned.

Apart from the higher number of senses provided by English WordNet, another factor playing a role in presenting new senses is collocations. For example, as Case 4 shows, for the verb "buy", three senses were assigned in total (See Table 1). While two of them are among the senses given for "buy" in English WordNet, the third one is the sense given for the collocation of "buy back". Thus, in addition to the senses of the verbs as individual forms, the senses of collocations including the verbs were also annotated. In total, for 18 senses of 15 verbs, senses of their collocations were also assigned.

5.3 Sense Combinations in English WordNet

Another difference between the two categorizations is that some senses in English PropBank are combined into a single sense in English WordNet. For instance, in Case 5, the two senses used in PropBank for "celebrate" are combined into one (See Table 1). In total, 6 verbs senses of which are combined in WordNet are "celebrate, cite, clear, explode, scuttle, and prepare".

5.4 Overlapping Categorizations

In some of the verbs, there is no one-to-one sense match between the two categorizations. In other words, a single sense in WordNet is annotated for at least two different senses of the same verb in PropBank. These overlapping categorizations are different from the sense combinations explained in 5.3 since in these verbs, a particular sense given in WordNet may replace more than one sense of a verb in PropBank in some of the sentences, whereas in the rest of the sentences, different senses, other than the overlapping one, are still annotated. For example, for the verb "build" given in Case 6, the sense of "develop and grow" was annotated for two different senses of "build" in English PropBank: "construct" and "grow" (See Table 1). However, in both of these sense categories of "build", other senses were also annotated: "make by combining materials and parts" in "build_01" and "bolster and strengthen" in "build_02". So, instead of combining these two senses, the sense of "develop and grow" seems to occur across different senses of the same verb. The number of verbs with sense overlaps are 29.

6 Automatic vs. Manual Tagging

Although automatic corpus alignment methods are preferred over manual tagging because of their systematicity and lower cost in many lexical resource integration studies, we argue that the manual tagging method is still highly needed. In an attempt to investigate the effectiveness of automatic and manual taggings, we compared our manually-tagged lexical resource integration with the automatically-tagged PM created by López de Lacalle et al.'s ((de Lacalle et al., 2014a; de Lacalle et al., 2014b; de Lacalle et al., 2016a; de Lacalle et al., 2016b)) based on PropBank senses. As a result of this comparison, we found that while the matchings of 418 WordNet senses in 413 verbs in PropBank and WordNet are the same in our and their integrations (See Case 1 in Table 3), the higher number of items with differences (1721 senses in 1387 verbs) in their matchings are worth attention. To mention some of those differences, for 307 PropBank senses in 281 verbs, our work and the PM do not match in any of the assigned senses. Also, for 199 PropBank senses in 178 verbs, there are both matching and mismatching senses.

Table 3: Example sense categorizations for English PropBank verbs by manual and automatic annotation

Case	Verb	Propbank Sense	WordNet Sense	
			Manual	Automatic
1	abdicate_01	to relinquish (power or responsibility)	give up, such as power, as of monarchs and emperors, or duties and obligations	give up, such as power, as of monarchs and emperors, or duties and obligations
	accelerate_01	make be faster, the act of speeding up	move faster	move faster
	accelerate_01	make be faster, the act of speeding up	cause to move faster	cause to move faster
2	zap_01	destroy	-	strike at with firepower or bombs
	zigzag_01	(cause to) move in zigzag fashion	-	travel along a zigzag path
3	emote_01	express emotion	give expression or emotion to, in a stage or movie role	-
	encrypt_01	encode, scramble digital information	convert ordinary language into code	-
4	accept_01	take willingly	consider or hold as true	consider or hold as true
	accept_01	take willingly	give an affirmative reply to	give an affirmative reply to
	accept_01	take willingly	receive willingly something given or offered	receive willingly something given or offered
	accept_01	take willingly	-	tolerate or accommodate oneself
	accept_01	take willingly	-	react favorably to
5	answer_01	give an answer, reply	react verbally	react verbally
	answer_01	give an answer, reply	give the correct answer or solution to	-
6	appeal_01	legal transaction	take a court case to a higher court for review	be attractive to
	appeal_03	be attractive	be attractive to	be attractive to

First of all, as we cover only a small part of PropBank data in our incomplete and still ongoing study, we lack 593 senses of 323 verbs included in the PM. In other words, those senses (such as "zap_01" or "zigzag_01" as shown in Case 2 in Table 3) are not included in our comparison at all. However, 8 verbs that are annotated in our limited integration, namely "emote, encrypt, franchise, indemnify, jell, motorize, outsell and squeegee" in Case 3 in Table 3, do not seem to be automatically annotated in the PM, which could be taken as the first evidence to suggest that automatic tagging may not be sufficient.

Secondly, when we look at the number of the matches assigned for each sense, we observe that the PM has matches that do not currently exist in our integration. For example, in Case 4 given in Table 3, while our integration provides only three senses for the item "accept_01" ((i) consider or hold as true, (ii) give an affirmative reply to and (iii) receive willingly something given or offered), the PM has two additional ones ((iv) tolerate or accommodate oneself and (v) react favorably to), adding up to five in total. The number of items with additional sense annotations is 497 in 477 verbs. The reason we suggest for those missing senses in our integration is that our work captures a portion of the whole Penn Treebank and as larger portions get annotated, those senses will be added to our integration, as well. On the other hand, in

addition to finding missing senses in our integration, we also came across senses that are included in our corpus but not in their PM. In total, 125 senses of 123 verbs in our integration do not exist in theirs. For instance, for the item "answer_01" in Case 5 in Table 3, we annotated two senses, which are (i) react verbally and (ii) give the correct answer or solution to. In the PM, only the first sense is annotated. So, we take the lack of those unassigned senses in the PM as our second evidence.

Thirdly, when we analyze the senses within the same verbs in the PM, we see that while some of them were annotated by taking into account their differences, some of them were merged, which resulted in the loss of some meanings. As an example, the same sense is assigned to "appeal_01" and "appeal_03" in the PM while different senses are annotated in our work as shown in Case 6 in Table 3. Due to this merger, PM fails to capture the sense that is needed, for example, for the sentence "Minpeco attorneys said they would appeal the decision to a federal district court.". Although not all the verbs with multiple senses were subject to that kind of wrong merging, we still consider this as an issue that needs to be resolved and take it as our third evidence to show the importance of manual tagging. Given that those errors cannot be noticed in automatic tagging, we suggest that manual tagging still has a crucial role in detecting those systematic errors resulting from automatic tagging.

7 Conclusion

In this paper, we reported our comparison results of English verb sense annotations in PropBank with senses in English WordNet for the sentences from Penn Treebank. In opposition to the idea that automatic tagging is good enough to eliminate the necessity for manual tagging, based on our comparison of our work with the PM, we contend that manual tagging is still needed to have qualitatively-better results and that it would be quite useful to apply it, at least in combination with automatic tagging.

Another issue that makes our work promising is its extendibility to a larger Turkish dataset. Related to that, Ak et al. (Ak et al., 2018) have recently constructed a Turkish Proposition Bank using translated sentences of English PropBank. So far, 9560 translated sentences are annotated with semantic roles and framesets are created for 1914 verb senses. In spite of its limited size, their study constitutes a base for Turkish Proposition bank. Therefore, we hope that our English Propbank and English Wordnet parallelization study can be used to extend many larger datasets in other languages, starting with the Turkish Proposition bank.

References

K. Ak, C. Toprak, V. Esgel, and O. T. Yildiz. 2018. Construction of Turkish proposition bank. *Turkish Journal of Electrical Engineering & Computer Sciences*.

O. Babko-Malaya, 2005. *Guidelines for Propbank Framers*.

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

M. Lopez de Lacalle, E. Laparra, and G. Rigau. 2014a. First steps towards a predicate matrix. In *Proceedings of the 7th Global WordNet Conference*, Tartu, Estonia, Jan 25-29.

M. Lopez de Lacalle, E. Laparra, and G. Rigau. 2014b. Predicate matrix: extending semlink through wordnet mappings. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*.

M. Lopez de Lacalle, E. Laparra, I. Aldabe, and G. Rigau. 2016a. A multilingual predicate matrix. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, Paris, France, may.

M. Lopez de Lacalle, E. Laparra, I. Aldabe, and G. Rigau. 2016b. Predicate matrix: automatically extending the semantic interoperability between predicate resources. *Language Resources and Evaluation*, 50(2):263–289, Jun.

C. Fellbaum. 1998. ed. wordnet: an electronic lexical database. *MIT Press, Cambridge MA*, 1:998.

P. Kingsbury and M. Palmer. 2002. From treebank to propbank. In *LREC*. European Language Resources Association.

P. Kingsbury and M. Palmer. 2003. Propbank: The next level of treebank. In *Proceedings of Treebanks and Lexical Theories*, Växjö, Sweden.

K. Kipper-Schuler. 2005. *VerbNet: a broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

M. Palmer, D. Gildea, and P. Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106, March.

M. Palmer. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*, Pisa, Italy.

M. T. Pazienza, M. Pennacchiotti, and F. M. Zanzotto. 2006. Mixing wordnet, verbnet and propbank for studying verb relations. 01.

J. I. Saeed. 1997. *Semantics*. Blackwell.

S. Stamou, K. Oflazer, K. Pala, D. Christodoulakis, D. Cristea, D. Tufis, S. Koeva, S. Totkov, D. Dutoit, and M. Grigoriadou. 2002. Balkanet: A multilingual semantic network for balkan languages. In *Proceedings of the First International WordNet Conference*, pages 21–25, Mysore, India.