Two corpus based experiments with the Portuguese and English Wordnets

Alexandre Rademaker, Fabricio Chalub, and Cláudia Freitas

[1] IBM Research `alexrad@br.ibm.com`
[2] IBM Research `fchalub@br.ibm.com`
[3] PUC-Rio `claudiafreitas@puc-rio.br`

Abstract. This paper presents two experiments with real world applications of word sense disambiguation, wordnets and dependency parsing. The first is an effort towards a portuguese wordnet annotated corpus. We manually annotated 30 sentences using OpenWordNet-PT as a lexicon and then compared the results with an automatic annotation. In addition to the system's evaluation, the results provided valuable insights about how to deal with such an ambitious task. The second experiment deals with using Princeton Wordnet as part of an NLP pipeline for information extraction from technical texts in the mining domain and the issues found while integrating word sense disambiguation with a syntactic analysis of the sentences.

1   Introduction

In the current context of the computational processing of languages, in which systems are no longer prototypes, resources capable of handling meaning processing are in the spotlight. Such resources may take the form of semantically annotated corpora or computational lexicons, or lexical databases. For the English language, the Princeton WordNet (PWN) [8] is the canonical example of a general and robust lexical base, widely used by natural language processing (NLP) systems. For the Portuguese language, with respect to resources similar to PWN [19], we highlight OpenWordNet-PT [20] (OWN-PT)[4], aligned with Princeton's WordNet, and that has $47,700$ synsets ($33,604$ nouns, $6,805$ verbs, $6,233$ adjectives, and $1,058$ adverbs).

OWN-PT was chosen by Freeling Library [7] and Open Multilingual Wordnet [4]. While it does not yet have a corpus aligned to it, first steps to such a resource were reported in [10].

The purpose of this article is to discuss the contributions of an alignment with a wordnet as one of the stages of natural language understanding. We report two studies. The first, in Portuguese and based on OWN-PT, was carried out on journalistic texts. This experiment was first reported in [10], published in Portuguese, and we repeat here some of its results. Building on some of the

---

[4] Available for download at `http://github.com/own-en/openWordnet-PT/` and for online navigation at `http://wnpt.brlcloud.com/wn/`.

results of this first exercise, we developed the second study. This time, we used English as target language and Princeton Wordnet [8]. Instead of journalistic texts, we relied on texts from a specific domain, technical reports from Canadian mining companies.

In the second experiment, problems arising from the limitations of OWN-PT itself were eliminated and we tried to minimize the impact of polysemy, since the phenomenon is less expected in specific domains. If the first study is taken as an exploratory investigation of the difficulty to produce a corpus annotated with OWN-PT senses, the second one is a follow-up, an investigation of how the difficulties found in the first experiment can be mitigated with the use of the Princeton Wordnet (a more mature wordnet) and a domain specific corpus.

2 Wordnets evaluation

Many works in NLP uses a form of evaluation that measures both soundness and completeness, for which the existence of a golden data set is essential. However, for the evaluation of lexical databases, such measures are not easily applicable. In particular, what would mean the notion of completeness? The ratio of knowledge correctly acquired in relation to all the knowledge that must be acquired? The problem is how to define "all the knowledge that must be acquired", since the same set of facts can lead to different interpretations and, consequently, to different types of "knowledge".

Although there are attempts to evaluate wordnets or similar resources in Portuguese [19], such evaluations are always comparisons, and they do not tell us much about the intrinsic quality of each resource. In addition, we agree with [5] when they indicate that a possible evaluation criteria for ontologies is to map them to the data (a data-driven evaluation). Therefore, an alignment between existing synsets and a corpus is a good way to verify their completeness — even though we know that a corpus will always be a limited portion of the language.

3 First experiment

The Freeling library [7] has a Word Sense Disambiguation (WSD) module, which performs an alignment between the words in the text and any semantic lexical database, it is an implementation of the graph-based method for WSD proposed in [2]. In order to verify the accuracy of the automatic disambiguation system, we created an experiment in which different annotators should select the appropriate synset (or synsets) for a word in context. Then, we compare the results obtained from the annotators and the Freeling WSD module.

We selected 30 phrases from the Brazilian portion of the Bosque corpus, the revised part of the Foresta sintá(c)tica [1]. The choice for the Brazilian variant aims to guarantee a good attribution of the senses, since the annotators were Brazilian. In addition, we consider only the nouns, and we select sentences with at least five nouns. The restriction on nouns is due to the well-known

verbal polysemy[5], which would make the task more difficult for the annotators. The total number of nouns evaluated was 226, with 204 different words. Each annotator received a form with the 30 sentences, and below each sentence we listed the target nouns, which in turn directed the annotators to the OWN-PT page with all the synsets in which the word parsed participated. The annotators should then select the appropriate synset, indicating in the form field the synset code. More than one synset could be chosen, as long as both fit the context equally, according to the annotators. The annotators were instructed to leave the field blank if they did not consider any suitable synset, regardless of the nature of the inadequacy.

It should be noted that the annotators did not receive any special training that would ensure familiarity with OWN-PT. Nine undergraduate students from linguistics (translation course) and one professional translator, considered "inexperienced" annotators, participated in the study. In addition, two linguists with annotation experience also participated, the "experienced" annotators.

3.1   Results and error analysis

Using the Kappa coefficient [6], which measures the degree of agreement among annotators, we performed two types of concordance assessment: only human concordance, and human concordance versus Freeling's disambiguation module.

In the inter-annotators agreement, considering only the inexperienced annotators and only one synset per annotator[6], the concordance index was 0.67. When, in the same group of annotators, we consider all the synsets chosen for the same word, the concordance index falls to 0.55. The low agreement rate is noteworthy, but it is equally astonishing that the agreement among only the experienced annotators was also 0.67.

Specifically for experienced annotators, when we compare Freeling's WSD module and annotator 1, the concordance is 0.45; the agreement between the WSD module and the annotator 2 is 0.52; and the agreement between both scorers and the WSD module is 0.56. Because agreement was low even among experienced scorers, the evaluation with the Freeling WSD module is poorly informative with respect to system quality. That is, if among humans it is difficult to agree on the appropriateness of synset, what performance should we expect from the system?

In about 20% of the cases it was pointed out the absence of an adequate synset. This absence, in turn, does not necessarily mean a gap in OWN-PT, since the alignment of words with synsets is preceded by the steps of tokenization and lemmatization. When some of these steps fail, the meaning assignment also fails.

---

[5] In PWN, for example, the average number of senses for verbs is 2.17 (with 36 verbs with over 20 senses), and for nouns is 1.22 (with only five nouns with over 20 senses).

[6] Throughout the evaluation, we noticed that there were more discerning annotators, who systematically chose to list all the synsets considered appropriate, as opposed to more economical annotators, who listed only the first appropriate synset they encountered. The option of evaluating one synset per annotator sought to avoid that the divergence in the quantity of the chosen synsets influenced the discordance.

The following are the situations in which this occurred:

1. Errors in the attribution of the part-of-speech class: six cases of mistaken nouns by adjectives or vice-versa;
2. Lemmatization errors regarding the 'number' feature: there are words with slightly different meanings when they are in the singular or plural: "recursos" (resources) can be the plural of "recurso" (resource) but, with the sense of goods and financial resources, it will always be used in the plural. The word "vésperas" (plural of eve) also has a less precise meaning than "véspera" (eve);
3. Tokenization errors and multiword units: it is difficult to find the appropriate synset when it contains a multiword unit, but tokenization is done on a word per word basis [21, 22], and this happened in about 20% of non-aligned words.

We know that some of these "failures" are not exactly errors, but rather non-consensual points in NLP and that are reflected in wordnets.

Another point is the need for a more systematic treatment of prefixes and other compounds with hyphen. In our exercise it was not possible to disambiguate "super-acordo" (super-agreement), which is absent from OWN-PT and it does not seem to us that it should be different. On the other hand, we would like "social-democrata" (social-democrat) to be in some synset. The existence of synsets related to US politics also poses challenges in annotating texts from another culture, and it may be necessary to create synsets relevant to the Portuguese-speaking countries. Finally, we do not know how to deal with stylistic effects, such as the use of the expression like "iron and fire", in the Example 1 which refer to the expression of "iron and fire", but also to the iron and fire of the grill.

(1) "Iti Fuji conquista clientela a ferro e fogo. O restaurante tem seu ponto forte no balcão de grelhados, que se sobrepõe aos prosaicos sushis e sashimis." (Iti Fuji wins over customers with iron and fire. The restaurant has the grill as its strong point, which upstages the prosaic sushis and sashimis.)

The possibility of assigning more than one synset to a word also contributed to the low agreement. Although we are aware of the perhaps excessive granularity of PWN, and of the well-known difficulty of clearly separating word senses [12], cases where more than one synset was suitable in the context of a sentence were far more common than we expected. Based on one of the experienced annotators, in at least 8% of annotated words more than one synset would be acceptable.

This last point was one of the main motivators for carrying out a second study. We know that polysemy/vagueness tends to be less frequent in terminology, and words typical of specific domains tend to be monosemic. When dealing with a more robust wordnet (PWN), and in a specific domain (mining), would the alignment be facilitated?

4 Second experiment

The second experiment reported here is part of a project of information extraction (references of entities and relation between those entities) from reports in the mining domain. Specialists in the domain have identified 16 PDF files to be used as seeds. These PDF files are scanned documents, and we used Apache Tika[7] for text extraction followed by manual fixing of the most obvious errors related to typos and formatting. The final corpus produced from these 16 PDF files contains $2,629$ sentences and $60,455$ tokens.

The corpus was processed by a NLP pipeline composed by tokenization, sentence segmentation, lemmatization, part-of-speech tagging, word sense disambiguation against the PWN and parsing. The pipeline produces a set of parsed trees enhanced with sense annotation using the PWN synset identifiers. The resulted parsed trees were used in two different stages: in the first stage, for queries for retrieving patterns of interest; in the second stage, the identified patterns obtained in the first stage were used to create rules for extracting facts to populate a knowledge base (e.g. mentions of entities or relations between entities). Once the rules and the NLP pipeline are refined, it can be reused for processing new documents to create a knowledge base. The final goal is to create a knowledge base to allow geologists to explore the data in an effective manner and obtain useful insights.

For example, the query `_ <amod|<compound L=deposit` (sentences that contain a token governed through a `amod` or `compound` dependency by a token with lemma 'deposit')[8] give us many candidates of references to styles of mineralization of rocks[9]: gold deposit, glacial sedimentary deposits, glaciofluvial deposits, mineral deposits, Zn-Cu deposit, stratiform gold deposits, etc.

However, the word 'deposition' may also be used as synonym of 'deposit' both included in the synset `13462191-n`. It would be desired to expand the query language to express a reference to a synset instead of only the lemmas or surface forms, i.e. `_ <amod|<compound S=13462191-n`. Another practical use of PWN synsets annotation is the canonical reference to chemical elements which are referred both by a name or a symbol: gold/Au `14638799-n`, zinc/Zn `14661977-n`, copper/Cu `14635722-n` etc. Of course, we can also explore the hyponym/hypernym relations to query for references to any chemical element `14622893-n`.

We used Freeling for the tokenization phase, POS tagging, lemmatization, sentence splitting, and word sense disambiguation. For the syntax analysis, since

---

[7] `https://tika.apache.org`

[8] The query language is loosely inspired by TGrep2 and TRegex, but is designed for querying general dependency graphs, see [13] and the online reference page `http://bionlp.utu.fi/searchexpressions-new.html`.

[9] In geology, mineralization is the deposition of economically important metals in the formation of rocks. Mineralization may also refer to the product resulting from the process of mineralization. For example, mineralization may introduce metals (such as iron) into a rock. That rock may then be referred to as possessing iron mineralization.

we wanted to use the universal dependencies [18], we opted for the SyntaxNet[10] parser for the ease of use and for the fact that we could customize its tokenization step. We used SyntaxNet with the pre-trained model with UD 1.3 release English corpus[11]. Figures 1a and 1b present examples of sentence annotated with UD 1.3 relations and POS tags.
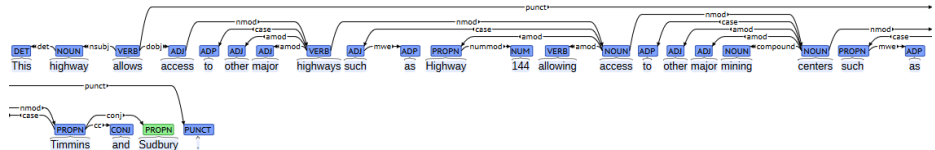
(a) The dependency tree of the sentence "This highway allows access to other major highways such as Highway 144 allowing access to other major mining centers such as Timmins and Sudbury."
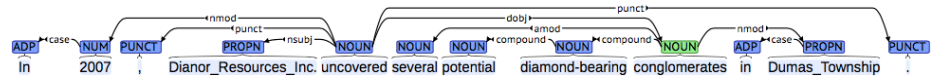
(b) The dependency tree of the sentence "In 2007, Dianor Resources Inc. uncovered several potential diamond-bearing conglomerates in $Dumas_T$ ownship."

Fig. 1: Example of parse trees from the mining domain in UD 1.3

Combining both systems posed a challenge: in order to use SyntaxNet we need to provide it a sequence of tokens that is compatible with the way the model that we are using was trained. That means that several modules had to be disabled in Freeling, including ones that affect directly the word sense disambiguation, such as the multiword recognition and numbers detection. For example, in the sentence of Figure 1a, it is very likely that 'such as' would be tokenized by Freeling as a single token `such_as`, which differ in how UD would annotate this expression (using the `mwe` relation). On the other hand, we chose to keep the Freeling's Named Entity Recognition (NER) module active, and this module identifies names such as 'Dianor Resources Inc.' and produce a single token (see Figure 1b).

4.1 Results and error analysis

Since both SyntaxNet and Freeling produce POS tags for each token, one of the first obvious idea was to measure the agreement between the systems. The disagreements happens mostly between: (1) tokens that Freeling tag as NOUN but SyntaxNet tag as ADJ, ADP, ADV or AUX; and (2) Freeling tagged as VERB

---

tokens that SyntaxNet tagged as ADJ or NOUN. Many of these errors are expected since both POS tagging models were trained with corpora from a different domain, Freeling was trained with the Penn TreeBank [16] and SyntaxNet was trainned with the UD 1.3 release (English corpus [23]). The errors in POS tag naturally propagate to errors in the WSD. The word 'till' is one such case, in 75% of its occurrences, it was tagged by Freeling as preposition, leading to a lack of sense in PWN. Nevertheless, almost all occurrences in the corpus should be tagged as noun with the sense `15074772-n` ('till' has only three noun senses in PWN). The word 'fault' is another such case, with 84 occurrences, 62% were tagged as `14464203-n` and only 2% received the expected sense `09278537-n` (a crack in the earth's crust resulting from the displacement of one side with respect to the other), this word has 8 senses in PWN but only three were used in our corpus. Table 1 presents the words with more than 100 occurrences in the corpus and the respective used senses. High frequency words tends to be more polysemic, but Table 1 shows that, at least for the words with more than 100 occurrences in this corpus, in most cases, one sense predominate over the others. In Table 1, we also noted that the most frequent senses chosen by the Freeling's WSD module is the right one that we expected to be chosen in all occurrences, which is directly related with our expectation discussed in the Section 1. As we discussed in the end of Section 3.1, a clear distinction between word senses can be difficult to sustain – for example, the two senses for the word 'rock' are both acceptable.

Freeling's POS tagger is highly dependent of its dictionary, since for every token, first Freeling assign for each token all possible pairs of lemma/POS and only after this step the system choose the best choice in the context (based on a trained statistical model). If a word is not in the dictionary, Freeling tries to guess the possible POS tag and lemma for the token. Table 2 shows the most frequent words not found in the Freeling's dictionary. For instance, 'volcanic rock' are often shortened to 'volcanics' in scientific contexts. In all occurrences of 'volcanics' the POS tag was guessed correctly but the lemma was kept 'volcanics', not found in the PWN, but 'volcanic rock' is in the `14933314-n`. This case shows that PWN is missing the word 'volcanics' in this same synset.

A serious challenge for NLP understanding is how to combine word sense disambiguation and parsing. The universal dependency model, chosen as our morphossintactic framework, emphasizes single words as the logical unit of analysis. Multiword expressions are to be related to other words via specific dependencies, such as *name*, *compound* or *mwe* (in UD 1.3). The motivation behind this decision is the goal to deal with discontinuous expressions that would not be possible to be joined in a single token, the main limitation of the Freeling's multiword module.

How to specify a sense that is comprised multiple words, such as `14933314-n` ('volcanic rock') that in the parse trees are represented as two tokens connected by the `amod` relation, `volcanic <amod rock`?

Another case is the phrasal verb particles, annotated with the dependency relation `compound:prt`. This relation holds between the verb and its particle

lemma	freq	used senses	pwn
gold	392	14638799-n (1.00)	7
sample	372	05821775-n (0.77) 01195299-v (0.21) 06026635-n (0.02) ? (0.00)	4
claim	347	06729864-n (0.97) 01018352-v (0.02) 02275365-v (0.01)	11
property	312	13244109-n (0.78) 08513718-n (0.17) 04916342-n (0.04) 05849040-n (0.01)	5
have	304	00065639-v (0.87) 00056930-v (0.04) 01156834-v (0.03) 02630189-v (0.02) 02204692-v (0.01) 02203362-v (0.01) 02108026-v (0.01) 01733477-v (0.01) 02236124-v (0.00) 00770437-v (0.00) 02210119-v (0.00)	20
zone	298	08688247-n (0.94) 08509442-n (0.06) 00332835-v (0.00)	6
meter	277	03753077-n (0.93) 13659162-n (0.05) 01273974-v (0.01) 07094093-n (0.01) 04991738-n (0.00)	6
area	270	08497294-n (0.73) 14514039-n (0.23) 05128519-n (0.03) 02735688-n (0.01) 05221895-n (0.01)	6
rock	240	14696793-n (0.78) 09416076-n (0.22)	9
drill	226	00606093-v (0.60) 00894552-n (0.34) 01443021-v (0.04) 00604094-v (0.02)	9
within	215	? (1.00)	1
survey	197	00881649-n (0.60) 00644503-n (0.34) 02150328-v (0.04) 00646542-v (0.01) 00646738-v (0.01) 06469694-n (0.01) 02198602-v (0.01)	9
along	157	? (0.94) 00069164-r (0.06)	5
north	154	01600333-a (0.83) 13831176-n (0.08) 00244043-r (0.06) ? (0.02)	9
hole	148	03526198-n (0.97) 03526805-n (0.02) 01282023-v (0.01)	10
value	144	05856388-n (0.85) 00681429-v (0.12) 13413493-n (0.03)	11
over	133	? (0.98) 00226550-r (0.02)	7
work	132	04599396-n (0.42) 00575741-n (0.24) 05755883-n (0.20) 04602044-n (0.09) 01525666-v (0.05) 00584367-n (0.01)	34
occur	131	02612368-v (0.53) 00339934-v (0.47) 00723349-v (0.01)	3
shear	128	04186709-n (0.62) 01560369-v (0.34) 01320816-v (0.04)	6
anomaly	124	09606527-n (1.00)	3
trend	121	02033295-v (0.57) 06197664-n (0.40) 05750163-n (0.02) 08681222-n (0.02)	5
locate	119	02694933-v (1.00)	4
south	117	01602330-a (0.89) 00243938-r (0.06) 13833375-n (0.03) ? (0.03)	7
east	113	00823350-a (0.62) 00323786-r (0.37) ? (0.01)	7
may	110	? (1.00)	2
line	109	04402057-n (0.44) 00582388-n (0.19) 02934168-n (0.09) 08377085-n (0.06) 08593924-n (0.05) 08593262-n (0.04) 06799897-n (0.03) 06260121-n (0.03) 05748786-n (0.03) 03671272-n (0.02) 03670849-n (0.02) 06626286-n (0.02)	36
part	109	08630039-n (0.61) 05929008-n (0.26) 05867413-n (0.08) 00787465-n (0.06)	18
program	107	06568978-n (0.87) 05898568-n (0.06) 06619428-n (0.05) 05899087-n (0.03)	10
till	106	? (0.75) 13357891-n (0.25) 15074772-n (0.01)	4
out	104	00233413-r (0.91) 00233295-r (0.04) ? (0.03) 02061487-a (0.02)	17
two	103	? (0.99) 13743269-n (0.01)	3
unit	103	00003553-n (0.82) 13583724-n (0.18)	6
report	102	06681551-n (0.47) 00965035-v (0.37) 07218470-n (0.11) 06409752-n (0.03) 00967098-v (0.01) 00965687-v (0.01)	13
carry	100	01449974-v (0.88) 02079933-v (0.08) 00235110-v (0.02) 01061017-v (0.01) 00738951-v (0.01)	41

Table 1: Most frequent words and their senses (numbers is parentheses are relative frequency)

freq	lemma	POS
162	mineralization	NOUN
51	volcanics	NOUN
45	ppb	NOUN
37	sediments	NOUN
35	till	NOUN
30	showings	NOUN
27	metavolcanic	NOUN
25	metavolcanics	NOUN
21	ppm	NOUN
20	north-south	NOUN
20	metasedimentary	NOUN
20	metagabbro	NOUN
20	greenstone	NOUN
18	mineralized	VERB
18	metasediments	NOUN
17	veining	VERB
15	three	NOUN
14	pillowed	VERB
14	per	NOUN
11	mafic	NOUN

Table 2: Most frequent unknown words and their POS tags

but in many cases, only the verb was tagged with a sense and we end-up losing semantics. One example is the expression 'carry out' with two senses in PWN versus the verb 'carry'. In our corpus we found 100 occurrences of 'carry' being two of them part of the expression 'carry out', in 88% of them, the word 'carry' was tagged with `01449974-v` (move while supporting) and 8% of them it was tagged as `02079933-v` (transmit or serve as the medium for transmission), senses very similar[12]. When we search in PWN for senses of all pairs of tokens connected by `compound:prt` in our corpus, we could find: 'carry out' (two senses), 'put down' (eight senses), 'drop off' (five senses), 'open up' (seven senses), 'make up' (nine senses), 'follow up' (two senses), 'pick up' (16 senses).

Unfortunately, the parser did not produce a consistent annotation of the expressions already presented in PWN. If we try the inverse of the previous analysis, that is, if we search how the MWE found in PWN were annotated in our corpus, we find many different dependency relations being used: 'carry out' (`compound:prt` and `advmod`), 'drill hole' (`compound`), 'as well' (`mwe`), 'base metal' (`compound`), 'up to' (`mwe`), 'at least' (`case`), 'east side' (`amod`), 'be due' (`cop`), 'representative sample' (`amod`).

Also with regards the way PWN deals with multiword expressions, we do find a number of inconsistencies when attempting to verify the completeness of a particular domain. For example, we find synsets such as `14996395-n` ('porphyritic rock'), but not one for 'aplitic rock'. There are adjectives for 'porphyritic' and 'aplitic', which suggests that we could opt out of having a noun for 'porphyritic rock' and use a more compositional model, combining adjectives and nouns to form new types of rocks (sort of a special case of semantic decomposable MWEs

---

[12] The synset `02079933-v` in PWN seems to contain an error in the verb frames associated to it, the gloss suggest that 'Something —s something' is missing.

from [22]). While this is in line with the universal dependency model, it carries the disadvantage losing some semantic information, as there is no hierarchy in PWN for adjectives. For example, while we know that `14697485-n` ('arenaceous rock') is a hyponym of `14698000-n` ('sedimentary rock'), there isn't a connection between `00142040-a` ('arenaceous') and `02952109-a` ('sedimentary'), nor should there be, since those adjectives can be applied to other nouns, not necessarily only types of rocks. On the other hand, it is also well known that the compositional model does not work for certain types of MWEs, like 'round' and 'round robin'.

5   Conclusion

One important conclusion of this article can be taken from the discussion about Table 1, where we noted that the most frequent senses chosen by the Freeling's WSD module were the right ones, supporting our expectation from the introduction. It is widely recognized that the fine grained senses of PWN makes word sense disambiguation harder. We also have shown that some senses can be equally acceptable in some contexts. These observations suggest that methods taking domain information into consideration (word domain disambiguation), such as the ones explored in [15, 14, 11] as well as the use of the 'Domain of synset' semantic relations from PWN or the 'subject area tests' from the English Slot Grammar are all valuable alternatives for the direction sense tagging of words with PWN synsets.

Bringing together word sense disambiguation is a delicate balance that needs to consider the POS tagging and dependency training corpus, WSD algorithm and tokenization. We have shown that it is possible to combine multiple NLP pipelines to achieve this goal, at the expense of losing valuable information, such as MWE senses. But how to tag word senses together with a dependency model that favors single words as the basic lexical unit? One possible idea is that sense should be assigned to the head of the MWE and none of the other words that belong to that MWE, provided that the dependencies are correctly annotated. This difficult is an unfortunate side effect of having each NLP step being independent of one another, with separate training models and such.

We believe that a more integrated approach of parsing, WSD and morphological analysis seems to be worth to investigate, such as the one taken by grammar based formalism like constraint grammar [3], English Slot Grammar [17] and the HPSG based English Resource Grammar [9].

We also aim to investigate possible approaches to guess the general area of a unknown word. For example, the word 'mafic' does not exist in PWN, but in our corpora it is connected to a number of words in the geology domain, such as 'volcanics', 'rock', 'gabbro', 'breccia', which all have senses.

References

1. Afonso, S., Bick, E., Haber, R., Santos, D.: Floresta sintá(c)tica: um treebank para o português. In: Gonçalves, A., Correia, C.N. (eds.) Actas do XVII Encontro

Nacional da Associação Portuguesa de Linguística (APL 2001). pp. 533–545. APL, Lisboa, Portugal (2-4 de Outubro de 2001 2002)

2. Agirre, E., Soroa, A.: Personalizing pagerank for word sense disambiguation. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. pp. 33–41. Association for Computational Linguistics (2009)

3. Bick, E., Didriksen, T.: Cg-3—beyond classical constraint grammar. In: Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania. pp. 31–39. Linköping University Electronic Press (2015)

4. Bond, F., Foster, R.: Linking and extending an open multilingual wordnet. In: Proceedings of the 51st annual meeting of the Association for Computational Linguistics (ACL). vol. 1, p. 1352–1362 (2013)

5. Brewster, C., Alani, H., Dasmahapatra, A.: Data driven ontology evaluation. In: In Int. Conf. on Language Resources and Evaluation (2004)

6. Carletta, J.: Assessing agreement on classification tasks: the kappa statistic. Computational linguistics 22(2), 249–254 (1996)

7. Carreras, X., Chao, I., Padró, L., Padró, M.: Freeling: An open-source suite of language analyzers. In: Proceedings of the 4th LREC (2004)

8. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database (Language, Speech, and Communication). The MIT Press (1998)

9. Flickinger, D.: Accuracy vs. robustness in grammar engineering. Language from a cognitive perspective: Grammar, usage, and processing pp. 31–50 (2011)

10. Freitas, C., Real, L., Rademaker, A.: Anotação de corpus com a openwordnet-pt: um exercício de desambiguação. In: Freitas, C., Rademaker, A. (eds.) Proceedings of the 10th Brazilian Symposium in Information and Human Language Technology. pp. 51–55. Natal, Brazil (Nov 2015), `http://www.aclweb.org/anthology/W15-5607`

11. Gella, S., Strapparava, C., Nastase, V.: Mapping wordnet domains, wordnet topics and wikipedia categories to generate multilingual domain specific resources. In: LREC. pp. 1117–1121 (2014)

12. Kilgarriff, A.: I don't believe in word senses pp. 1–33 (Dec 2014)

13. Luotolahti, J., Kanerva, J., Pyysalo, S., Ginter, F.: Sets: Scalable and efficient tree search in dependency graphs. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. pp. 51–55. Association for Computational Linguistics (2015), `https://aclweb.org/anthology/N/N15/N15-3011.pdf`

14. Magnini, B., Cavaglia, G.: Integrating subject field codes into wordnet. In: LREC. pp. 1413–1418 (2000)

15. Magnini, B., Strapparava, C., Pezzulo, G., Gliozzo, A.: Using domain information for word sense disambiguation. In: The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems. pp. 111–114. Association for Computational Linguistics (2001)

16. Marcus, M., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K., Schasberger, B.: The penn treebank: Annotating predicate argument structure. In: Proceedings of the Workshop on Human Language Technology. pp. 114–119. HLT '94, Association for Computational Linguistics, Stroudsburg, PA, USA (1994), `http://dx.doi.org/10.3115/1075812.1075835`

17. McCord, M.C.: The slot grammar lexical formalism. Tech. Rep. RC23977 (W0607-020), IBM Research (Jul 2006)

18. Nivre, J., de Marneffe, M.C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., Zeman, D.: Universal dependencies v1: A multilingual treebank collection. In: Chair), N.C.C., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Paris, France (may 2016)

19. Oliveira, H.G., de Paiva, V., Freitas, C., Rademaker, A., Real, L., Simões, A.: As Wordnets do Português, vol. 7, pp. 397–424. OSLa, Oslo, Noruega (Março 2015), `https://www.journals.uio.no/index.php/osla/issue/view/100/showToc`

20. de Paiva, V., Rademaker, A., de Melo, G.: OpenWordNet-PT: An Open Brazilian WordNet for Reasoning. In: Proceedings of 24th International Conference on Computational Linguistics. COLING (Demo Paper) (2012)

21. Ramisch, C.: A generic framework for multiword expressions treatment: from acquisition to applications. In: Proceedings of ACL 2012 Student Research Workshop. pp. 61–66. Association for Computational Linguistics (2012)

22. Sag, I.A., Baldwin, T., Bond, F., Copestake, A., Flickinger, D.: Multiword Expressions: a pain in the neck for NLP. In: Conference on Intelligent Text Processing and Computational Linguistics. pp. 1–15. Springer Berlin, Heidelberg (2002)

23. Silveira, N., Dozat, T., de Marneffe, M.C., Bowman, S., Connor, M., Bauer, J., Manning, C.D.: A gold standard dependency corpus for English. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014) (2014)