Comparing Verb Synonym Resources for Portuguese

Jorge Teixeira, Luís Sarmento, and Eugénio Oliveira

Universidade do Porto - LIACC
Rua Dr. Roberto Frias, s/n - Porto, Portugal
`{jft,las,eco}@fe.up.pt`

Abstract. In this paper we compare verb synonym information contained in four public-available lexical-semantic resources for Portuguese: TeP, PAPEL, Wiktionary and OpenThesaurusPT. We quantify the extent to which verb synonymy information in four resources overlaps, and we quantify how much novelty each resource in comparison to the others. We demonstrate that the four resources vary *significantly* in respect to verb synonymy information. Also, we show that by merging the four resources we can obtain a more comprehensive verb thesaurus. Finally, we suggest that resource merging may actually be required in order to avoid *performance* and *evaluation bias* that arise from coverage problems when using only one of these resources.

Keywords: Language Resources, Synonymy, Evaluation.

1 Introduction

In the last few years there has been a significant increase in the number of lexical-semantic resources for Portuguese. This results, on one hand, from the products of several finished or on-going projects, such as TeP [1], WordNet.PT [2] and, more recently, MultiWordNet.PT (the Portuguese branch of the MultiWordNet project [3]) and PAPEL [4]. Essentially, these projects aim at producing *high-quality* WordNet-like resources for Portuguese, supported either by *human specialists* or by previously compiled *reference dictionaries*. On the other hand, community-edited resources, such as the Wikipedia, Wiktionary or OpenThesaurus, have become an extremely attractive option for obtaining *broad-coverage* lexical-semantic knowledge, at almost no cost.

As the number of choices of lexical-semantic resources increases, several questions immediately arise. The first is: *how do these resources relate to each other*? One typical claim is that manually edited resources - such as TeP or, indirectly, PAPEL - will have high-quality information at the cost of a lower coverage (or over-specialization), while community-edited resources tend to have much broader coverage (e.g. at the level of the semantic domains included) but provide no quality guarantees. The second question is: *how much benefit is there from merging the information of all resources*? Merging is an interesting option

not only because of *quality vs. coverage* issues but also because, quite para-
doxically, resources tend to *diverge* in many points. This is so because when
building lexical-semantic resources, developers need to follow a given linguistic
/ conceptual framework, which inevitable imprint a specific *semantic bias* [5] to
the corresponding resource. Since there are several possible linguistic / concep-
tual frameworks, the resulting resources will tend to be significantly different in
points where the frameworks do not agree.

In the current work, we compare the *verb synonymy* information offered by four
freely-available lexical-semantic resources for Portuguese: TeP, PAPEL,
Wiktionary and OpenThesaurusPT. These four resources result from four distinct
approaches to resource building and are thus expected to posses only partially
overlapping information. We will focus on comparing the information regarding
one specific semantic relation - synonym - for one word class only: verbs. We will
compute the degree to which the verb synonymy information in four resources
overlaps, and quantify how much novelty each resource has in relation to the oth-
ers. Finally, we will present statistics about the result of merging the verb synonym
information of the four resources into a more comprehensive verb thesaurus.

2   Related Work

Despite the significant increase in the number of available lexical-semantic re-
sources, including community-edited resources such as Wikipedia or Wiktionary,
in the last few years there has not been a corresponding effort by the research
community in *explicitly* comparing those resources. There are, however, a few
works in which a set of lexical-semantic resources is compared with the pur-
pose of (i) achieving cross-validation, (ii) measuring the benefit of merging their
information, or (iii) testing their usefulness in a specific application settings.

For example, Oliveira et al [6] evaluate PAPEL, a lexical resource built by
extracting relations from a large commercial Portuguese dictionary [4], and per-
form a comparative study against TeP thesaurus [1] by computing the overlap
between synsets of both the resources. Results obtained show that 50% of the
synonym pairs in PAPEL were found in TeP, while only 39% of the pairs in TeP
were found in PAPEL.

Navarro et al [7] evaluate how two lexical-semantic resources, WordNet and
Dicosyn (a compilation of synonym relations extracted from seven dictionaries)
can be used to augment synonymy information in Wiktionary. They compute
two overlap measures: *recall* - the fraction of the common entries between the
resources, and *precision* - the fraction of common synonyms for the common
entries. The comparison between the English Wiktionary and WordNet yields
recall values from 0.45 (for verbs) to 0.12 (for nouns), while precision ranges from
0.23 (for nous) to 0.08 (for verbs). As for the comparison between the French
Wiktionary and Dicosyn, the recall values ranges from 0.35 for nouns to 0.32 for
verbs, while precision values varies from 0.08 (for nouns) to 0.04 (for verbs).

In [8] the authors compare the performance of several word selection proce-
dures that use WordNet-like resources for computing two semantic relatedness

measures - *path length* and *concept vector*. Four different resources were used to ground the computation of such measures: Wiktionary, Wikipedia, GermaNet and WordNet. Results achieved using Wiktionary were found to be *comparable*, and sometimes *better*, than those achieved when using expert-made resources (i.e. GermaNet and WordNet).

Muller et al [9] compare the effectiveness of using Wiktionary, Wikipedia and WordNet in reducing the vocabulary gap between query terms and documents in a Information Retrieval setting. They showed that all the three resources contain sufficient information for achieving such goal by helping to expand queries with related words (obtained by computing a word relatedness metric presented in [8]). However, results differed depending on the semantic relation between the words in query and best words for expansion. For example, while WordNet was especially effective when hypernyms/hyponyms were the optimal expansion keywords, Wiktionary was the most effective in providing synonyms and co-hyponyms, and Wikipedia was the best in providing words related by non-classical relations (e.g. functional).

3   Resources Being Compared

In this section we describe the four resources being compared: Wiktionary, PAPEL, TeP and OpenThesaurusPT. Among other information, all the four resources store synonym information about verbs (which are the focus of our current study), and are freely available for download. There are other lexical-semantic resources for Portuguese, such as WordNet.PT and MultiWordNet.pt, which provide Web interfaces available for users to query[1]. However, as far as we know, none of these two resource is available on the Web *for download*.

3.1   Wiktionary

Wiktionary is a comunited-edited resource that stores a wide variety of information about words, such as grammar, pronunciation, etymology, definitions, a list of synonyms, a list of related terms, a list of derived terms, and also translations to other languages. Wiktionary project started in 2002, and it currently has more than 6.9 million entries for 172 languages. The Portuguese version of Wiktionary was started in 2004 and contains approximately 92,000 entries. The structure of a Wiktionary page for verbs is generally divided in two sections from which synonymy information can be obtained. The first is the *synonyms list* that explicitly presents synonyms, but is only available for a relatively small number of words. The second section is the *word definitions* that may be found in a larger number of pages, and provide extensive information regarding synonyms. The version of the Wiktionary for Portuguese we used[2] contains 92,144 pages,

---

[1] WordNet.PT: http://cvc.instituto-camoes.pt/wordnet/
MultiWordNet.pt: http://mwnpt.di.fc.ul.pt/
[2] http://download.wikimedia.org/ptwiktionary/20090809/ptwiktionary-20090809-pages-articles.xml.bz2

each containing a word/expression. We found 4,818 pages addressing verbs in Portuguese, of which 1,656 have synonym information.

3.2   PAPEL

PAPEL (*Palavras Associadas Porto Editora Linguateca*) [4] is a freely available[3] lexical database for Portuguese, automatically constructed by extracting relations from a large commercial Portuguese dictionary. The process consisted in parsing 237,246 dictionary definitions (using a chart parser), and using a set of manually-developed lexical-syntactical patterns to mine 10 semantic relations: (i) *hypernym*, (ii) *cause_of*, (iii) *part_of*, (iv)*means_to_end*, (v) *place_of*, and its corresponding inverse relations. PAPEL contains information about 125,871 words: 50,201 are nouns, 17,932 verbs, 14,025 adjectives and 43,713 adverbs (however, this value seems to be too large and is probably a spelling mistake in [4]). It contains 80,429 pairs of *synonym* relations, of which we will focus on the ones that connect pairs of verbs.

3.3   TeP

TeP (*Thesaurus Electrônico para o Português do Brasil*) [1] is a freely-available[4] electronic thesaurus for the Brazilian Portuguese, strongly inspired by WordNet. TeP development was based on four dictionaries used as reference corpus and relies on three main concepts: (i) *synsets*, i.e. sets of synonyms, (ii) the *lexical matrix*, which defines a biunivocal correspondence between synsets and *senses*, (iii) and the *index*, which establishes the antonymy relation by using indexes. TeP contains information about approximately 44,000 words: 17.000 nouns, 15,000 adjectives, 11,000 verbs and 1,000 adverbs. There are 19,885 synsets, 4,145 of which related to verbs. As far as we know, there is only one study [10] (unpublished at the time of this writting) that compares *TeP* thesaurus with *WordNet.pt* and *PAPEL*.

3.4   OpenThesaurusPT

OpenThesaurusPT is the official thesaurus used for OpenOffice and is freely available online[5]. The OpenThesaurusPT project [11] started as an effort to migrate the German OpenThesaurus project to the Portuguese. The OpenThesaurus project is based on Wordnet, and also uses synsets to hold information about synonyms and antonyms. It also stores some information about superordinate/subordinate (i.e. hypernymy/hyponymy) relations between synonym sets. The current available version of OpenThesaurusPT dates from 2006-08-17, and contains 4,010 synsets for 12,941 words, 2,815 of them verbs.

---

[3] http://www.linguateca.pt/PAPEL/PAPELv1.0.zip
[4] http://www.nilc.icmc.usp.br/tep2/download.htm
[5] http://openthesaurus.caixamagica.pt/index.php

4   Experimental Set-Up

4.1   Pre-processing and Normalization

The resources being compared provide information about synonymy in different formats: TeP and OpenThesaurusPT provide *synsets*, PAPEL provides a list with *pairs* of synonyms and Wiktionary provides information about the synonyms of a word embedded in the corresponding page, formatted according to the standard Wiki formatting style. For Wiktionary, we developed our own parsing procedures, highly optimized for extracting synonyms from lists and from available definitions. However, because Wiktionary is a community-edited resource, several pages have formatting inconsistencies which lead to extraction problems. We thus, manually evaluated a sample of 20% of the synonym data extracted. We concluded that our extraction procedures achieve over 98% of *precision* and over 90% *recall*, which is compatible with our evaluation purposes.

Additionally, each resource deals with the symmetry of synonymy in different ways. For example, in Wiktionary if we find from the page of verb v_A that v_B is one of its synonyms, there is no guarantee that we find in the page of v_B information indicating that v_A is one of its synonyms. We have thus inverted all synonym relations extracted from Wiktionary to obtain explicit representations for *both* directions of the relation. In TeP and OpenThesaurusPT, information is given in synsets, so we established bidirectional synonymy links between all possible pair in the set. In PAPEL, there is one explicit synonymy link for each direction of the synonym relation so no further processing was needed. Finally, for each of the four resources, generically denoted by $\mathcal{R}$, the (symmetric) synonymy pairs for each verb were aggregated in order to create mappings between verbs $v^{\mathcal{R}}(i)$ the corresponding list of synonyms, $S^{\mathcal{R}}(i)$:

$$v^{\mathcal{R}}(i) \longrightarrow S^{\mathcal{R}}(i) = [s_1^{\mathcal{R}}, s_2^{\mathcal{R}}...s_j^{\mathcal{R}}] \tag{1}$$

We will denote, $v^{\mathcal{R}}(1)$, $v^{\mathcal{R}}(2)$, ... $v^{\mathcal{R}}(n)$, the set of all verbs for which synonym mappings were obtained as $\mathcal{V}^{\mathcal{R}}$. Thus, for Wiktionary, PAPEL, OpenThesaurus PT and TeP we will have, $\mathcal{V}^{wk}$, $\mathcal{V}^{pp}$, $\mathcal{V}^{ot}$ and $\mathcal{V}^{tp}$, respectively. It is important to clarify that information about the different senses of verbs was available only in TeP and in the OpenThesaurusPT thesaurus (i.e. the synsets provided address different senses). However, while producing the synonym mappings we ignored information about multiple senses. Thus, for all verb considered in this work, the set of synonyms includes synonyms for all senses found, so as to compare all resources in equal conditions. Table 1 presents some statistics about the number of verbs, $|\mathcal{V}^{\mathcal{R}}|$, and the average number of synonyms gathered for each verb, $\overline{|S^{\mathcal{R}}(i)|}$, for the four resources under study. Also, we present an indicative measure of amount of synonymy information found in each resource, by computing the number of synonymy links it stores i.e. $|\mathcal{V}^{\mathcal{R}}| \times \overline{|S^{\mathcal{R}}(i)|}$.

Table 1. Verb Synonyms in each resource after normalization

Resource	$|\mathcal{V}^{\mathcal{R}}|$	$\overline{|S^{\mathcal{R}}(i)|}$	$|\mathcal{V}^{\mathcal{R}}| \times \overline{|S^{\mathcal{R}}(i)|}$
OpenThesaurusPT	2,815	1.67	4,701
PAPEL	4,142	5.38	22,284
Wiktionary	1,964	2.34	4,596
TeP	10,827	3.28	35,512

4.2 Comparison Measures

We are mainly interested in comparing two parameters over all resources. The first is the *overlap* between them, i.e. the amount of information that resources have in common. This implies measuring two more specific parameters: (i) the number of common verb entries extracted, and (ii) for each common verb entry, the number of common synonyms found. The second parameter is the relative *novelty* between resources, i.e. the amount of information that one resource has while the others do not. Again this can be seen both at the level of verb entries or, for each verb entry in common, at the level of synonyms.

Let us assume that $\mathcal{A}$ and $\mathcal{B}$ are two generic verb synonymy mappings to be compared. Let us define *Entry Overlap*, $\mathcal{O}_{ent}$, as the number of common verb entries between resources $\mathcal{A}$ and $\mathcal{B}$:

$$\mathcal{O}_{ent}^{\mathcal{A}:\mathcal{B}} = |\mathcal{V}^{\mathcal{A}} \cap \mathcal{V}^{\mathcal{B}}| = |\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}| \tag{2}$$

We can also define *Entry Novelty from $\mathcal{A}$ to $\mathcal{B}$* to as the number of entries in $\mathcal{A}$ for which there is *no* corresponding entry in $\mathcal{B}$:

$$\mathcal{N}_{ent}^{\mathcal{A}\rightarrow\mathcal{B}} = |\mathcal{V}^{\mathcal{A}} - \mathcal{V}^{\mathcal{A}\cap\mathcal{B}}| \tag{3}$$

Let us now consider only the set of verbs that have synonyms in both $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}$. For all verbs $v(i)$ in $\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}$ we can measure the *synonym overlap* $O_{syn}^{\mathcal{A}:\mathcal{B}}(i)$ between the corresponding synonyms sets, i.e. the number of common synonyms between the set of synonyms found in $\mathcal{A}$, $S^{\mathcal{A}}(i)$, and in the set found in $\mathcal{B}$, $S^{\mathcal{B}}(i)$:

$$O_{syn}^{\mathcal{A}:\mathcal{B}}(i) = |S^{\mathcal{A}}(i) \cap S^{\mathcal{B}}(i)| \tag{4}$$

A resource-wide synonym overlap measure, $\mathcal{O}_{syn}^{\mathcal{A}:\mathcal{B}}$ can be computed by averaging the values of overlap for all the verbs at stake, $\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}$:

$$\mathcal{O}_{syn}^{\mathcal{A}:\mathcal{B}} = \frac{\sum_i^{|\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}|} O_{syn}^{\mathcal{A}:\mathcal{B}}(i)}{|\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}|} \tag{5}$$

For any $v(i) \in \mathcal{V}^{\mathcal{A}\cap\mathcal{B}}$ we can also define *synonym novelty from $\mathcal{A}$ to $\mathcal{B}$*, $N_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}(i)$, as the number synonyms found in $\mathcal{A}$, $S^{\mathcal{A}}(i)$, that are not found in $\mathcal{B}$, $S^{\mathcal{B}}(i)$:

$$N_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}(i) = |S^{\mathcal{A}}(i) - S^{\mathcal{A}}(i) \cap S^{\mathcal{B}}(i)| \tag{6}$$

Likewise, the resource-wide *average synonym novelty from $\mathcal{A}$ to $\mathcal{B}$, $\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}$*, can be computed by averaging the values of $N_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}(i)$ for all verbs in $\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}$:

$$\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}} = \frac{\sum_{i}^{|\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}|} N_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}(i)}{|\mathcal{V}^{\mathcal{A}\cap\mathcal{B}}|} \tag{7}$$

5   Results and Analysis

Table 2 presents the values of *entry overlap*, $\mathcal{O}_{ent}^{\mathcal{A}:\mathcal{B}}$, between all resources. Wiktionary has the lowest entry overlap among all the evaluated resource, eventhough its overlap with TeP is the highest. TeP, which is the largest resource being evaluate, is, as expected, the one with the highest entry overlap.

Table 2. Entry Overlap $\mathcal{O}_{ent}^{\mathcal{A}:\mathcal{B}}$ between resources

| $\mathcal{A}:\mathcal{B}$	PAPEL	Wiktionary	TeP
OpenThesaurusPT	2,119	1,407	2,664
PAPEL	-	1,816	3,647
Wiktionary	-	-	1,883

Entry Novelty results are presented in Table 3 and show that Wiktionary is the resource with less novel information, while PAPEL and TeP are the ones the highest number of novel entries.

Table 3. Entry Novelty, $\mathcal{N}_{ent}^{\mathcal{A}\rightarrow\mathcal{B}}$, between the resources

$\mathcal{A}\setminus\mathcal{B}$	OpenTh.PT	PAPEL	Wiktionary	TeP
OpenTh.PT	-	696	1,408	151
PAPEL	2,023	-	2,326	495
Wiktionary	557	148	-	81
TeP	8,163	7,180	8,944	-

Figure 1 presents perspective that combines both the relative sizes of the four resources (already presented in Table 1) and the entry overlap between all combinations of resources (including the aggregation of all resources). As it can be seen, the entry overlap between *all* resources is relatively low: there are only 1339 common entries between all resources in an universe of 11,450 verbs. It is also interesting to confirm that the two largest resources exhibit a relatively large number of exclusive entries, while the two smallest (and community-edited resources) have much fewer novel entries (i.e. Wiktionary with only 29 novel verbs and OpenThesaurusPT with 98 novel verbs).

Tables 4 and 5 present statistics about the average synonym overlap, $\mathcal{O}_{syn}^{\mathcal{A}:\mathcal{B}}$ and the average synonym novelty $\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}$. There are two interesting facts that can be observed in these tables. The first is that the synonym overlap between
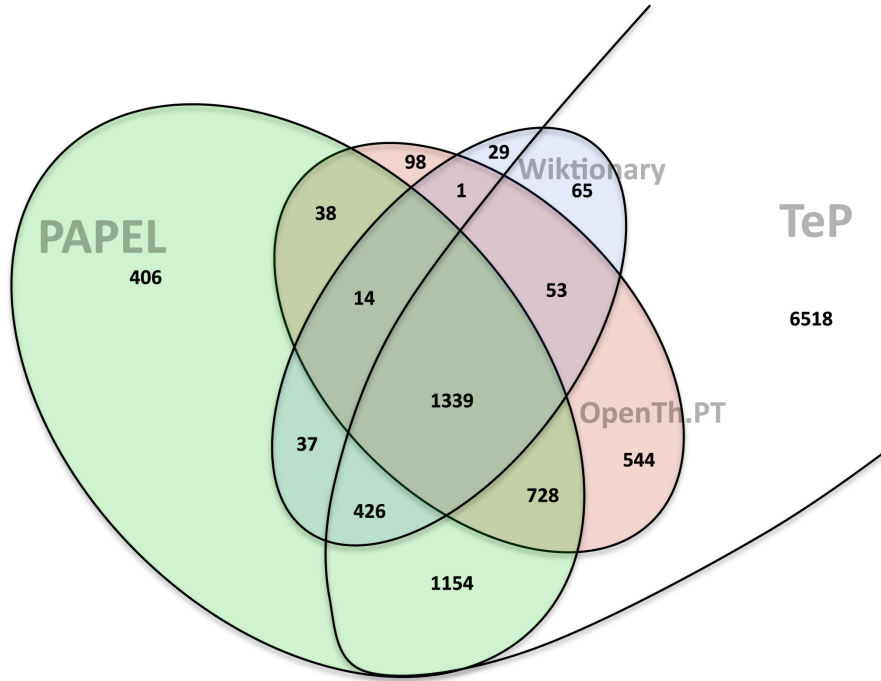
Fig. 1. Perspective of all the resources and their overlap

resources is *very low*. In most cases, the average synonym overlap is less than 1 meaning that there are many verbs for which the resources do not share any synonym. There is a relatively higher synonym overlap between PAPEL on one side and OpenThesaurusPT and Wikitionary on the other, but not between these two. This suggests that PAPEL aggregates part of the synonymy information that can be either on OpenThesaurusPT or on Wikitionary. Notably TeP has a very low synonymy overall with all other resources despite the fact that it is the resource with larger number of verb entries.

The second interesting fact observed is that most values of average synonymy novelty, $\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}$, are relatively high. For example, for shared verbs PAPEL stores 6.97 synonyms not found in the OpenThesaurusPT, 6.54 not found in Wikitionary and 5.00 not found in TeP. Notably, the inverse is also true: for the shared verbs TeP has 3.90 synonyms not found in PAPEL, and OpenThesaurusPT has 1.05 synonyns not found in PAPEL. Of all resources, OpenThesaurusPT seems to be

Table 4. Synonym Overlap $\mathcal{O}_{syn}^{\mathcal{A}:\mathcal{B}}$ between resources

$\mathcal{A}:\mathcal{B}$	PAPEL	Wikitionary	TeP
OpenThesaurusPT	0.68	0.42	0.44
PAPEL	-	1.62	0.87
Wikitionary	-	-	0.50

Table 5. Synonymy Novelty $\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}$ between the resources

$\mathcal{A}\setminus\mathcal{B}$	OpenTh.PT	PAPEL	Wiktionary	TeP
OpenTh.PT	-	1.05	1.43	1.22
PAPEL	6.97	-	6.54	5.00
Wiktionary	2.12	0.80	-	1.88
TeP	4.92	3.90	5.80	-

the one with less average synonymy novelty in relation to the others but, still, the corresponding $\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}$ values are always higher than one.

The low values of $\mathcal{O}_{syn}^{\mathcal{A}:\mathcal{B}}$ combined with the relatively high values of $\mathcal{N}_{syn}^{\mathcal{A}\rightarrow\mathcal{B}}$ suggest that the synonym information differs *considerably* among the four resources. Additionally, since entry novelty among resources is also quite high (check Figure 1) one can say that the four resources are *significantly different*. By merging the information of the four resources we found that it us possible to obtain synonym mappings for 11,450 verb with 5.49 synonyms in average, which results in 62,902 synonymy links. When comparing these values with the corresponding values for each of the four resources (see Table 1)one can see that there is a very significant increase in all the parameters, especially in the number of synonym links. This confirms that merging existing resources is, in fact, a good option for reducing the coverage and recall gaps of each individual resource.

6   Conclusion and Future Work

We compared the overlap and novelty of verb synonymy information contained in four freely-available resources. Results show that there are *significant differences* among *all* the resources both at the level of *verb coverage* and at the level of *synonym links* between shared verbs. Globally, as far as verb synonymy is concerned, TeP and PAPEL are clearly much more comprehensive resources than Wiktionary or OpenThesaurusPT. However, the verb synonymy information contained in these two community-edited resources is non-negligible. In the particular case of Wiktionary, we believe that it is a very interesting resource for developers to follow since it is constantly growing (contrary to OpenThesaurusPT which has not been updated recently) and has the potential to efficiently cover more the frequently used lexicon.

Overall, these results should draw our attention to the fact that applications using any of these resources may obtain rather different performances depending on the actual resource chosen. Likewise, when using any of these resources as *gold-standard* data to evaluate automatic procedures, evaluation figures may vary significantly depending on the resource used as reference. Additionally, since each resource covers different parts of the lexicon, it is unsafe to compare automatic procedures using only one of these resources. It is possible that both procedures being evaluated produce different yet *equally valid results*, although only one set of results intersects with the specific gold-standard chosen. By merging information all the four resources we reduce the chances of these problematic events.

In future work, we wish to extend the comparative work performed on synonymy verbs both to other parts-of-speech, such as nouns and adjectives, and to other semantic relations such as antonymy, hypernymy/hyponymy and meronymy. However, not all resources available (e.g. the OpenThesaurusPT thesaurus) provide information about all these semantic relations. In the case of Wiktionary, new extraction procedures have to be developed to mine information regarding additional semantic relations.

Acknowledgments

References

1. da Silva, B.D.: Construcão de um thesaurus electrónico para o português do brasil. Processamento Computacional do Português Escrito e Falado (PROPOR) 4, 1–10 (2000)
2. Marrafa, P., Amaro, A., Chaves, P., Lourosa, S., Martins, C., Mendes, S.: Wordnet.pt - uma rede léxico-conceptual do português online. In: XXI Encontro da Associação Portuguesa de Linguística, 28-30 de Setembro, Porto, Portugal (2005)
3. Pianta, E., Bentivogli, L., Girardi, C.: Multiwordnet: developing an aligned multilingual database. In: Proceedings of the First International Conference on Global WordNet, Mysore, India, January 21-25 (2002)
4. Oliveira, H., Santos, D., Gomes, P., Seco, N.: Papel: A dictionary-based lexical ontology for portuguese. In: Teixeira, A., de Lima, V.L.S., de Oliveira, L.C., Quaresma, P. (eds.) PROPOR 2008. LNCS (LNAI), vol. 5190, pp. 31–40. Springer, Heidelberg (2008)
5. Velardi, P., Pazienza, M.T., Fasolo, M.: How to Encode Semantic Knowledge: A Method for Meaning Representation and Computer-Aided Acquisition. Computational Linguistics 17, 153–170 (1991)
6. Oliveira, H., Santos, D., Gomes, P.: Avaliação da extracção de relações semânticas entre palavras portuguesas a partir de um dicionário. In: 7th Brazilian Symposium in Information and Human Language Technology, STIL 2009 (2009)
7. Navarro, E., Sajous, F., Gaume, B.: Wiktionary and nlp: Improving synonymy networks. In: Proceedings for the Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources (2009)
8. Zesch, T., Müller, C., Gurevych, I.: Using wiktionary for computing semantic relatedness. In: Proceedings of AAAI, pp. 861–867 (2008)
9. Muller, C., Gurevych, I.: A study on the semantic relatedness of query and document terms in information retrieval. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (2009)
10. Santos, D., Barreiro, A., Costa, L.: O papel das relações semânticas em português: Comparando o tep, o mwn.pt e o papel. To appears in XXV Encontro Nacional da Associação Portuguesa de Linguistíca, Lisboa, Portugal (2009)
11. Naber, D.: Openthesaurus: Building a thesaurus with a web community. Technical report (2004)