

Combining Resources: Taxonomy Extraction from Multiple Dictionaries

Rogelio Nazar, Maarten Janssen

IULA - Universitat Pompeu Fabra
Roc Boronat 138 - Barcelona
Rogelio.Nazar@upf.edu, Maarten.Janssen@upf.edu

Abstract

The idea that dictionaries are a good source for (computational) information has been around for a long while, and the extraction of taxonomic information from them is something that has been attempted several times. However, such information extraction was typically based on the systematic analysis of the text of a single dictionary. In this paper, we demonstrate how it is possible to extract taxonomic information without any analysis of the specific text, by comparing the same lexical entry in a number of different dictionaries. Counting word frequencies in the dictionary entry for the same word in different dictionaries leads to a surprisingly good recovery of taxonomic information, without the need for any syntactic analysis of the entries in question nor any kind of language-specific treatment. As a case in point, we will show in this paper an experiment extracting hyperonymy relations from several Spanish dictionaries, measuring the effect that the different number of dictionaries have on the results.

1. Introduction

The internet provides access to an enormous array of resources, many of them hand crafted by hard labour. With the number of resources available, it becomes possible to generate new types of resources by the automatic combination of existing ones. As a case in point, we will show in this article how it is possible to compile a taxonomy database from the comparison of various online dictionaries of the same language. Extracting taxonomic (or semantic) information from dictionaries has a long-standing tradition, starting back in the seventies (see section 2.). The typical way of extracting taxonomies from dictionaries is by taking a single dictionary, and craft a dedicated parser for the semantic definitions in that dictionary, which in turn extracts taxonomic relations from each dictionary entry.

The approach presented in this paper is quite different: it does not use any analysis of the content of the dictionary, but is rather based on a comparison of entries for the same word in a number of different dictionaries, using frequency counts of the words in these various definitions to establish taxonomic relations. The idea behind this comparison is that whereas the exact definitional phrasing in different dictionaries will be distinct (even if only for legal reasons), the choice of genus term should be (relatively) constant. If we assume that most words have an “ideal” genus term, or at least a genus term that lexicographers tend to agree upon, then most if not all dictionaries should use that genus term in their definition. And therefore, we can expect that if we compare dictionary definitions, the (typical) genus term for a given word should be that term that reoccurs in the definition of that word in the majority of the dictionaries.

For the study presented here, we use a comparison of several Spanish dictionaries that can be directly accessed via the internet, partially extended with dictionaries from other sources. Many dictionary sites offer the possibility to look up a word directly, meaning that given a word, it is easy to automatically retrieve its definition.

One of the advantages of the method presented here is that it is largely language and dictionary independent: the only source of information is the frequency counts of the defini-

tions. Although this paper describes the extraction of a taxonomy for Spanish from a number of Spanish dictionaries, the method uses little or nothing that is specific to Spanish or the dictionaries used. The method presented here should work equally well for any other language for which a number of different dictionaries is available electronically.

2. State of the Art

Automatic extraction of taxonomies has been an active field of research since the early days in computational linguistics. Two different types of automatic extraction can be identified. The first is the attempt to extract taxonomies from dictionaries, which started soon after the availability of the first copies of machine readable dictionaries in the late seventies and eighties. The second trend is the attempt to extract taxonomies directly from corpus, which came up in the nineties during the arrival of corpus linguistics and gained a lot of momentum with the growing of the internet. Despite the fact that we use statistical counts that are more particular for corpus based methods, we will focus here only on the dictionary related approaches.

The attempts to extract taxonomies from dictionaries started with pioneering work by Calzolari (1977) and Amsler (1981). Calzolari was interested in the extraction of hyperonymy as well as synonymy relations. Her approach basically focused on the first noun occurring in the definition of a word. She found that often in the top of the hyperonymic chain, words tend to refer to each other in circles. The work by Amsler focused specifically on taxonomy extraction. For the construction of a proper taxonomy, he needed not only the word used as hyperonym, but also the meaning in which that word was used. Since the meaning could not be established automatically, the kernels of the definitions were manually disambiguated.

Chodorow et al. (1985) attempted to automatize the project of genus term disambiguation in order to expand existing taxonomies with dictionary data. They observed that the genus term is usually the head of the defining phrase, and that the head can be extracted not via a full syntactic parsing of the definition but via simple heuristics, such as taking the

first verb after the element *to* in the case of the definitions of verbs.

Several authors have attempted to combine information extracted from dictionaries with information extracted from corpus (Briscoe, 2001; Velardi et al., 2007; Granitzer et al., 2009), but these attempts are beyond the scope of this paper. With few exceptions (Ide and Véronis, 1993b; Sanfilippo and Poznanski, 1992), the vast majority of the revised literature does not take advantage of the use of more than one dictionary to increase certainty on the extracted hyperonymy links and no one, to our knowledge, has attempted a purely statistic and language independent approach.

After this initial work, many other projects have arisen to extract data from dictionaries, including the notable work by Fox et al. (1988), Alshawi (1989), Boguraev (1991), Barrière and Popowich (1996), Chang (1998), Renau & Battaner (2008), and many others. The methodology for extraction in these more recent projects is not significantly different from the method used in the first projects, the main innovation being that the newer projects often applied *lexico-syntactic patterns* (Hearst, 1992), and that apart from hyperonymy relations, they included other types of semantic information as well, such as agentive and meronymic information.

The initial enthusiasm was dampened by a series of critical reviews of the progress in the field by Véronis & Ide (1991; 1993a; 1995), who concluded that a fully automatic procedure for the extraction of taxonomies from dictionaries with an acceptable quality was beyond the reach of computational linguistics, at least at its state at the time. Their criticism is based on the fact that dictionaries are in practice too inconsistent and unsystematic to lead to a reliable knowledge base. It should be noted that a good number of the points of criticism by Ide & Véronis have to do with the more “advanced” semantic relations and not only to hyperonymy.

3. Experimental Design

As said in the introduction, our method for the extraction of hyperonym terms from dictionaries is to look for the word(s) in the definition of a given lexical entry that appears in the largest number of dictionaries. In the experiment to test the validity of this idea, we extracted hyperonymic relations for Spanish words from a number of dictionaries available online. The dictionaries that were used for this are:

1. Diccionario del Real Academia Española (DRAE - <http://buscon.rae.es/draeI/>)
2. Diccionario General de la Lengua Española (DGLE - via <http://www.diccionarios.com/>)
3. Diccionario Clave de Uso del Español Actual (Clave - <http://clave.librosvivos.net/>)
4. Diccionario del Español Usual en México (DEM - <http://mezcal.colmex.mx/dem/>)

The best demonstration of the idea behind the method is to use an example. Let us consider the Spanish word *ablandabrevas* (good-for-nothing). The definitions of this

word in the dictionaries listed above, with the exception of the DEM which does not include this word, are given in table 1.

DRAE	ablandabrevas <i>sustantivo común</i> 1 fig., fam. Persona inútil o perezosa. NOTA:Pl. blandabrevas.
DGLE	ablandabrevas 1. <i>com. coloq.</i> Persona de poco valer.
Clave	ablandabrevas (plural <i>ablandabrevas</i>) s. <i>com. col. desp.</i> Persona inútil y pusilánime

Table 1: Dictionary definitions for the word *ablandabrevas*

A quick look at these entries shows that each dictionary uses a different wording in their definition, but that nevertheless, there are two words that recur in these different definitions: *inútil* (useless 2x) and *persona* (person 3x). Our suggestion is that the reason why *persona* is the most frequent term in these definitions is that it is the hyperonym of *ablandabrevas*, and being the typical hyperonym term for that word, it is used in the definitions in each of these dictionaries.

We could in principle even go a step further: given that the word *inútil* always occurs to the right of the word *persona*, we can expand the n-gram into *persona inútil*. However, in this article, we will restrict ourselves to single word hyperonymic terms.

The algorithm used to predict that *persona*, as well as *inútil*, are hyperonymy candidates for *ablandabrevas* consists fundamentally of the following three simple steps:

1. For a given input word X, retrieve the definitions of that word from each of the dictionaries used.
2. Count for each word in these definitions in how many of the definitions it occurs.
3. Consider the words that occur in the largest number of dictionaries as candidate genus terms.

This algorithm works in spite of the fact that different dictionaries use different formulations in their definitions. In fact, it *relies* on the fact that the definitions are different, since otherwise, all the words of the definition would appear in each dictionary. Fortunately, dictionary definitions are copyright protected, which obliges lexicographers to use a phrasing for their definitions that is different from that in competing dictionaries.

We ran this algorithm on a random sample of 100 Spanish nouns, and the results will be presented in section 4. However, before turning to the results of the experiment we will present some refinements we made to this basic algorithm.

3.1. Refinements

The hyperonym in the (random) example in table 1 stands out very nicely. But the results are not always equally clear-cut. In fact, when the methodology explained above is applied directly, the results are rather disappointing. Therefore, we implemented several adjustments to improve the

performance. The most relevant of these refinements are mentioned in the remainder of this section. An important point is that none of these improvements made use of the internal structure of the definitions nor did they include any kind of linguistic processing.

3.1.1. Definition Selection

What is retrieved by the harvesting script is the lexical entry for a given word in each dictionary. However, the hyperonym should only appear in the semantic definitions, and not in the rest of the text of the entry – the example sentences, the inflectional information, etc. Some of the webpages of dictionaries use specific, recognizable HTML tags to indicate where the definitions are to be found.

For dictionaries that have such indicative labels, all text except for the definitions themselves was stripped off. In principle, it might be possible to parse the entries of dictionaries that do not have such tags to extract the definitions. However, retrieving the definitions from the entry is a messy task given the inconsistent markup of many dictionaries. For dictionaries in which the definitions could be identified, we furthermore only took the first 6 words of the definition into account, since the hyperonym is in practice almost always used near the beginning of the definition.

3.1.2. Raw Frequency

In theory, the comparison analysis described in the article relies not on raw frequencies, but on the total number of dictionaries in which a word appears. However, in practice, it is better to take raw frequency into account too, because the most likely word to be re-used a number of times in a single definition is the hyperonym. Therefore, we used a weighted frequency score instead of just counting the number of dictionaries. The ranking score for each word was calculated as in (1), where X is number of dictionaries and Y is the overall frequency of the word:

$$\text{Score} = X + \max(X/2, Y - X) \quad (1)$$

This scoring takes the raw frequency into account, but always weighs the number of dictionaries heavier than the raw frequency count.

3.1.3. Filters

We applied a number of filters to the candidate list that boost the precision without losing any significant recall. The main filters are described below.

A genus term for a word should itself also be a word, and therefore should appear in the dictionary. Much of the noise in the candidate list is cleared by filtering out all genus term candidates that are not themselves in the dictionary. This is done simply by looking up each (high ranking) candidate in the list in one of the dictionaries in which it appeared. If a definition for the candidate is not found, the candidate is discarded. Furthermore, the genus term should always be of the same word-class as the word itself, so we also filtered out words that do occur in the dictionary, but do not belong to the same word class. Given that genus terms always appear in citation form, it is not necessary to lemmatize the candidates first.

We used a stoplist to discard high-frequency words from the candidate list. Since dictionary texts have their own special

wordings, the use of a standard stoplist is not working ideally well in this case. Therefore, we created a dictionary specific stoplist for each dictionary. These stoplists were created automatically in the following way. From a sample of 1000 entries from each dictionary, words that occurred in more than 5% of the definitions were put on the stoplist for that dictionary. For instance, in the Clave dictionary, 20% of all the entries in our sample contained the word *latín* (Latin) because it is common to indicate the etymology of a word. For the same reason, 38% of the entries contained the word *etimología* (etymology).

We did manually modify this stoplist afterwards, for the following reason: there is a small number of hyperonym terms that are used so commonly that they actually appear in more than 5% of the definitions, such as the word *persona* (person). Therefore, we manually removed all those words from the dictionary stoplists that looked like they might be common genus terms.

3.1.4. More Dictionaries

It is obvious that for extracting hyperonymy candidates by dictionary comparison, the more dictionaries are used, the better the probability of finding a correct hyperonym. As part of the current experiment, we extended the number of dictionaries consulted for each word by including some more dictionaries for which we had a partial electronic version available that was collected in a different project:

- Diccionario del Uso del Español (María Moliner)
- Diccionario del Español Actual (Manuel Seco)
- Diccionario Salamanca de la Lengua Española

We will discuss the influence that these additional dictionaries have in section 4.4.

4. Results

Although the methodology of the experiment presented in this paper is rather straightforward, the evaluation is less trivial. The problem is that it is not obvious when to consider a candidate correct or incorrect. Firstly, because there are several different ways of checking whether a candidate is a genus term or not, and these different methods lead to different results. Secondly, because there are several ways of measuring the quality of the result. Given that we select more than one candidate, there are often several candidates with the same score in the ranking and, moreover, there is often more than one correct hyperonym. Therefore, we did three different evaluation tests:

1. We compared the generated genus term candidates against the hyperonyms listed in WordNet.
2. We evaluated a set of genus term candidates by hand.
3. We conducted an evaluation with a group of 10 human judges.

4.1. WordNet Evaluation

To (roughly) estimate the quality of the hyperonymy candidates, we used the Spanish WordNet taxonomy as a reference, despite the fact that it has several drawbacks for the experiment at hand. The comparison was done as follows: we took a sample of words from the Spanish part of EuroWordNet, consisting of 100 randomly selected single-word nouns. For each of these words, we looked up the direct hyperonym listed in WordNet, and also the two hyperonym levels above that. We then ran the word against our algorithm, looking up the dictionary definitions and extracting the candidate hyperonym terms. We considered the result correct if one of the first five candidates matched one of the hyperonyms listed in WordNet.

We use not only the direct hyperonym but a hyperonymic chain because dictionaries tend to use less intermediate levels in their taxonomy than WordNet does. We considered the result correct even if the highest ranking candidate was not a match because in many cases, the different candidates represent genus terms of different subsenses of a word, and the dictionary tends to include more word senses than the Spanish WordNet. It should be noted that in many cases, the algorithm only yields one or two candidates.

All cases in which the word for WordNet was found in less than three of the dictionaries were ignored, since the comparison does not yield reliable results when there are insufficient dictionaries to compare. Given the differences in nature between the Spanish WordNet and the dictionaries used, this is rather often the case: about 1/3 of the words in WordNet does not appear in (sufficiently many) dictionaries. The main reason for this large mismatch is that, although the dictionaries are all substantially larger than the Spanish WordNet, WordNet contains things that are typically not included in dictionaries, such as proper names. We also ignored all the cases in which there was no hyperonym for the word in WordNet itself.

From the remaining trials, the evaluation showed a match between the hyperonym candidate list and the WordNet hyperonymic chain in about 50% of the words. Although this number is rather low, it does not actually mean that the other 50% of the cases yielded an incorrect hyperonym. For several reasons, the WordNet comparison yield a significant number of false negatives, which is why we manually verified the status of the remaining candidates, as explained in the next section.

4.2. Manual Post-Evaluation

For one of the trial sets, we manually evaluated the results of the algorithm. This manual verification showed many cases in which the hyperonymy candidate was actually correct, but simply different from what is found in WordNet. There are two main reasons for this: either the dictionary provides an alternative hyperonym, or the dictionary contains a different word-sense from what is listed in WordNet. A case of an alternative correct hyperonym is the following: for the word *devaneo* (idle pursuit) the algorithm correctly finds the words used as hyperonyms in the dictionary, including *pasatiempo* (pastime) and *distracción* (distraction; amusement). WordNet, however, does not list any of these as possible hyperonyms, but rather list *diversión* (fun) as a

hyperonym. These kinds of mismatches account for more than a third of the mis-hits.

To give an example where the algorithm provided a hyperonym for a different word-sense: the word *muestra* means either “exhibition” or “sample”, but only the first meaning is listed in WordNet. The algorithm encountered a hyperonym for the second meaning: *porción/cantidad* (portion/quantity). But since that hyperonym does not relate to a word-sense listed in WordNet, it is hence also not listed in WordNet as a hyperonym. There were five cases in our trial run in which such a correct hyperonym was found belonging to a non-WordNet word sense.

The detailed results of the manually verified trial-run are given in table 2. If we consider all the cases where the algorithm obtains a correct result, the accuracy in this trial is 71%.

Number of Words	100
Correct according to WordNet	34
Ignored	34
Not in WordNet	Alternative
	Other meaning
	Misses
	8
	5
	19

Table 2: Scoring of the Trial Run

Many of the cases where the algorithm completely fails are cases in which the dictionaries do not give a genus term in the definition: As already noticed by Calzolari (1977), the hyperonym can be missing in the case of definitions in terms of remission, when only a synonym is given, or because the definition excludes a hyperonym: when the hyperonym is *person* or *thing*, the dictionary often simply states *who* or *what* instead of giving the noun explicitly. However, there is also a small number of cases in which the dictionaries do give a hyperonym, but the algorithm fails to pick it up. This is mostly the case when the different dictionaries do not agree upon the hyperonym: only consolidated hyperonyms are detected by the algorithm.

4.3. Human Evaluation

To have an independent scoring that avoids the above mentioned problems in the WordNet comparison, we conducted a human evaluation as well. Evaluators were asked to think of a word, introduce it in a web interface of the algorithm and check if there is a correct hyperonym among the first five hyperonym candidates. They provided a satisfaction score according the following scale: 0 (no hyperonym on the list), 1 (hyperonym found) or 0.5 (dubious). Dubious cases are for instance those where the hyperonym retrieved by the algorithm is correct but it is not the prototypical hyperonym of the word in question. Table 3 shows the results of this second evaluation, with its average of 77% satisfaction.

We believe that the reason for this slight increase in the proportion of correct trials is the fact that human judges had to come up with the input word by themselves, and this of course guarantees that the majority of them will not propose very rare words such as those that will be retrieved by random sampling, as in the comparison against WordNet. If a word is relatively common, there is a strong prob-

Number of Observers		10
Number of Words		69
Scoring	Correct	51
	Incorrect	14
	Dubious	4
Average satisfaction		0,77

Table 3: Human Evaluation results

ability that the hyperonym will be common too, resulting in a more likely repetition of the hyperonym word in the dictionaries. The most frequent hyperonyms retrieved after this human evaluation were *mamífero* (mammal) and *mueble* (furniture) with four occurrences; followed by *recipiente* (container), *árbol* (tree), *asiento* (seat), and *instrumento* (instrument) with three occurrences.

4.4. More Dictionaries

As could be expected, the number of dictionaries used in the comparison has a significant influence on the accuracy of the results of our algorithm. However, the effect was less than we expected it to be. To demonstrate exactly what the effect is, we ran a set of words against different numbers of dictionaries, and calculated the accuracy scores for each set, the results of which are given in figure 1.

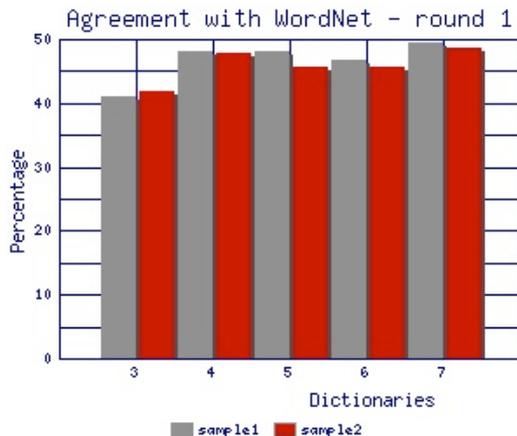


Figure 1: Agreement with WordNet and Dictionary count

As before, we selected a set of 100 random nouns, and calculated in how many cases the correct hyperonym was retrieved. Given the amount of work it would require to manually verify this much data, we only used the automated comparison against the WordNet taxonomy described in section 4.1. Therefore, the numbers indicated do not really represent the precision of the algorithm, but merely the amount of agreement with the WordNet taxonomy. However, since the same methodology was used for each set of dictionaries, figure 1 gives a good indication of the effect of the number of dictionaries.

To show the effect of the number of dictionaries, we started with a set of three dictionaries, and ran the experiment using a set of 100 nouns. Then we added one more dictionary to the mix and ran the experiment again, continuing to do so until all seven dictionaries were included. To counter the effect of the particular choice of dictionaries and the par-

ticular choice of nouns, we ran the entire experiment four times. We ran it first with one sample of 100 nouns, and then with another set of 100 nouns. And then we ran the entire experiment again starting with a different set of three dictionaries. So in total, we ran the experiment 20 times, the detailed results of which are given in table 4.

# Dicts	Round 1		Round 2	
	Sample 1	Sample 2	Sample 1	Sample 2
3	41.09	41.79	32.43	33.82
4	48.05	47.82	42.85	37.68
5	48.05	45.71	44.15	42.85
6	46.75	45.71	46.75	47.14
7	49.35	48.57	48.05	50

Table 4: Agreement with WordNet and Dictionary count

5. Issues and Comparisons

Although the overall results of the experiment were presented in the previous section, there are some issues related with the taxonomy extraction that merit a more detailed explanation. Many of the problems listed here are well-known problems in the literature, and we merely point out how our algorithm deals with them; others are problems that are specific to the type of methodology we used. We do not provide an extensive comparison with similar projects in this section, because such a comparison is complicated by a number of factors, most importantly by the fact that there are no other experiments as far as we know that start from comparable premises. However, we do provide some comparison where possible.

5.1. Polysemy

The traditional problem of polysemy in taxonomies is that if the genus term used in a definition has more than one meaning, it will be the hyperonym only in one of its meanings. Determining the correct reading of the genus term was one of the major concerns of Amsler (1981). The problem is more complicated than just finding out the right meaning, since there are cases in which there is no hyperonymic meaning to be found. Take for instance the word *porthole*, which is a type of *window*. But it is a type of window in two meanings of the word that are listed in the dictionary: in both its sense of an opening in a wall, and its meaning of the frame surrounding that opening (Janssen, 2002).

On top of the traditional problem, algorithms that combine different dictionaries, such as ours, have an additional problem with polysemy: each word-sense has its own hyperonym, and hence a dictionary entry should yield as many hyperonyms as it has word-senses (although several word-senses may have the same hyperonym). In the comparison of the dictionaries, the word-senses should then be compared, and not the words themselves. However, there is nothing to indicate which meaning in one dictionary corresponds to which meaning in another. Automatically lining up the senses is a complicated task, if possible at all. On top of that, dictionaries in general do not correspond nicely in terms of the selection and number of meanings they list for

polysemous words, so there even is no alignment between the different word-senses across dictionaries.

In our approach, these two problems are not resolved: the algorithm merely provides hyperonymy relations between words, not a full-blown taxonomy of word-senses. In principle, the algorithm returns all the hyperonyms for all the word-senses. Of course, the more hyperonyms a word has, the less reliable the results become. For a word like *mano* (hand), there is even a hard cut-off rule in the algorithm that discards many of the hyperonyms: we select the five best candidates as candidate hyperonym terms, whereas the DRAE lists 36 different word senses for this word.

5.2. Synonymy

As noted already by for instance Ide & Véronis (1993b), dictionaries often differ in their choice of hyperonym. If the hyperonym is not the same in each dictionary, our algorithm quickly fails: it picks up not just any hyperonym listed, but only those hyperonyms that are commonly agreed upon (by lexicographers). In a sense, as also pointed out by Ide & Véronis, this is an advantage, since the resulting taxonomy is typically better balanced. But of course it does mean that the algorithm fails in such cases. If there are only two alternative hyperonym, they are typically both picked up by the algorithm, but the reliability clearly goes down.

When the dictionaries do not agree on the hyperonym, there is not a real problem with the fact that the algorithm does not pick it up. But one would want to say that there is a consolidated hyperonym when two dictionaries provide two hyperonyms that are synonyms or, even stronger, orthographic variants of each other. For instance, there are several types of dances, such as the *cachucha*, which is defined in the Moliner dictionary as a *danza* (dance), but in the DRAE as a *baile* (dance). Although we did experiment with using synonymy data to retrieve such variants in the choice of hyperonym, there is no (or at least no easy) way of retrieving the hyperonym with our method when dictionaries do not use exactly the same word as the genus term.

6. Conclusion

In this article, we have shown how it is possible to extract hyperonymy relations directly from online dictionaries without going through the process of writing dedicated parsing scripts. Although the experiment shown here was with Spanish dictionaries, there is nothing language specific in the set-up, and the design of the experiment should work equally well for other languages, provided that there are at least three dictionaries available. Moreover, the same methodology should be exploitable for other purposes as well, taking advantage of the same idea: the use of frequency as a factor in the process of learning from heterogeneous and even noisy sources.

It should be noted that the results of the extraction method explained in this article do generate more noise than typically desirable for a fully automatically construed taxonomy. But especially as a method amongst others, the results are very fruitful and easy to obtain.

In the near future, we will try to combine the information from the hyperonymic links extracted from dictionaries

with the method of the induction of taxonomies from corpus, as it is described in Nazar et al. (submitted). It is to be expected that the quality of the final taxonomy extracted from the combination of these two independent methods will supersede the results of each of them in isolation.

7. References

- H. Alshawi, 1989. *Computational lexicography for natural language processing*, chapter Analysing the dictionary definitions, pages 153–169. Longman Publishing Group, White Plains, NY, USA.
- R. Amsler. 1981. A taxonomy for english nouns and verbs. In *Proceedings of the 19th annual meeting on Association for Computational Linguistics*, pages 133–138, Morristown, NJ, USA.
- C. Barrière and F. Popowich. 1996. Building a noun taxonomy from a children’s dictionary. In *Proceedings of Euralex’96*, pages 65–70.
- B. Boguraev. 1991. Building a lexicon: The contribution of computers. *International Journal of Lexicography*, 4(3):227–260.
- T. Briscoe. 2001. From dictionary to corpus to self-organizing dictionary: Learning valency associations in the face of variation and change. In *Proceedings of Corpus Linguistics 2001*, pages 79–89, Lancaster University.
- N. Calzolari. 1977. An empirical approach to circularity in dictionary definitions. *Cahiers de Lexicologie*, 31(2):118–128.
- J. S. Chang, S. J. Ker, and M. Chen. 1998. Taxonomy and lexical semantics - from the perspective of machine readable dictionaries. In *AMTA ’98: Proceedings of the Third Conference of the Association for Machine Translation in the Americas on Machine Translation and the Information Soup*, pages 199–212, London, UK. Springer-Verlag.
- M. S. Chodorow, R. J. Byrd, and G. E. Heidorn. 1985. Extracting semantic hierarchies from a large on-line dictionary. In *University of Chicago*, pages 299–304.
- E. A. Fox, J. T. Nutter, T. Ahlswede, M. Evens, and J. Markowitz. 1988. Building a large thesaurus for information retrieval. In *Proceedings of the second conference on Applied natural language processing*, Austin, Texas.
- M. Granitzer, A. Augustin, W. Kienreich, and V. Sabol. 2009. Taxonomy extraction from german encyclopedic texts. In *Proceedings of the Malaysian Joint Conference on Artificial Intelligence 2009*, Kuala Lumpur, Malaysia.
- M. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING)*, pages 539–545, Nantes, France.
- N. Ide and J. Véronis. 1993a. Extracting knowledge bases from machine-readable dictionaries : have we wasted our time? In *KB&KS’93 Workshop*, pages 257–266, Tokyo.
- N. Ide and J. Véronis. 1993b. Refining taxonomies extracted from machine readable dictionaries. In *Research in Humanities Computing 2*. Oxford University Press.
- N. Ide and J. Véronis. 1995. Knowledge extraction from machine-readable dictionaries: An evaluation. In *Ma-*

- chine Translation and the Lexicon*, volume 898, pages 19–34, Berlin. Springer Verlag.
- M. Janssen. 2002. *SIMuLLDA: a multilingual lexical database application using a structural interlingua*. Ph.D. thesis, Utrecht University.
- R. Nazar, L. Wanner, and J. Vivaldi. submitted. Automatic taxonomy extraction using distributional semantics.
- I. Renau and P. Battaner. 2008. Agrupación semántica de verbos de un daele a través de sus relaciones hiperonímicas. In *III Congreso Internacional de Lexicografía Hispánica*, Málaga.
- A. Sanfilippo and V. Poznanski. 1992. The acquisition of lexical knowledge from combined machine-readable dictionary sources. In *Proceedings of the third conference on Applied natural language processing*, pages 80–87, Morristown, NJ, USA. Association for Computational Linguistics.
- P. Velardi, A. Cucchiarelli, and M. Petit. 2007. A taxonomy learning method and its application to characterize a scientific web community. *IEEE Trans. on Knowl. and Data Eng.*, 19(2):180–191.
- J. Véronis and N. Ide. 1991. An assessment of semantic information automatically extracted from machine readable dictionaries. In *Fifth Conference of the European Chapter of the Association for Computational Linguistics*, pages 227–233, Berlin.