Barbecued Opakapaka: Using Semantic Preferences for Ontology Population

Ismail El Maarouf **Georgiana Marsic** **Constantin Orăsan**
University of Wolverhampton
{i.el-maarouf, georgie, c.orasan}@wlv.ac.uk

Abstract

This paper investigates the use of semantic preferences for ontology population. It draws on a new resource, the Pattern Dictionary of English Verbs, which lists semantic categories expected in each syntactic slot of a verb pattern. Knowledge of semantic preferences is used to drive and control bootstrapped pattern extraction techniques on the EnClueWeb09 corpus with the aim of identifying common nouns belonging to twelve semantic types. Evaluation reveals that syntactic patterns perform better than lexical and surface patterns, at the same time raising issues about assessing ontology population candidates out of context.

1 Introduction

This paper investigates the use of weakly supervised techniques driven by semantic preferences from the Pattern Dictionary of English Verbs (PDEV)[1] on the task of ontology population.

PDEV is the output of Corpus Pattern Analysis (CPA; Hanks, 2004), a technique in corpus lexicography for mapping meaning onto words in text. PDEV (Hanks and Pustejovsky, 2005; Hanks, 2013; El Maarouf et al., 2014) is a new resource which organizes the description of a verb entry according to its main patterns of use. Its major features are (1) that it only accounts for uses found in a corpus in a bottom-up data-driven approach, and (2) that the analysis focuses on the accurate description of word patterns, rather than on the analysis of word meanings in isolation.

Ontology population is defined as the automatic identification of the nouns classed under a semantic category in the CPA ontology[2].

[1] http://pdev.org.uk/

[2] http://pdev.org.uk/#onto

This paper describes ontology population techniques driven by PDEV semantic preferences applied to a web-scale corpus. The next section describes the resources used in this paper, section 3, the ontology population techniques, and section 4, the evaluation, before concluding in section 5.

2 Resources

2.1 The CPA Ontology

PDEV aims to provide a well-founded corpus-driven account of verb meaning, using semantic types (STs) to stand as prototypes for collocational clusters occurring in each clause role. Current CPA practice has shown that the scientific concepts from WordNet (Fellbaum, 1998), the most widely used semantic repository in NLP, do not map well onto words as they are actually used. This is partly because folk concepts, and not scientific concepts, form the foundation of meaning in natural language (Wierzbicka, 1984). For this reason, the CPA Ontology has been developed for PDEV, and it contrasts with WordNet in the following key aspects: (1) WordNet considers each synset (sense) as a node in the ontology while the CPA Ontology connects STs which cover multiple senses; thus WordNet synsets are either STs or word senses. (2) WordNet is intuition-based whereas the CPA Ontology is ‘corpus-driven’.

The CPA ontology is inspired from the Brandeis Semantic Ontology (Pustejovsky et al., 2006), but has been gradually populated with STs based on the need to capture a verb’s set of collocates. Each of the 220 STs currently included in the CPA Ontology is connected to at least one verb pattern, as can be observed on the public PDEV website.

2.2 Unambiguous PDEV Verb Patterns

PDEV uses STs to characterize the set of collocates found in the slots of a verb pattern. For example, the verb *barbecue* has only one pattern, as

Proceedings of Recent Advances in Natural Language Processing, pages 153–159,
Hissar, Bulgaria, Sep 7–9 2015.

Pattern	[[Human]] barbecues [[Food]]
Implicature	Human cooks Food on a rack over an open fire in the open air
Example	The South African environment department has refused permission to fishermen in Struisbaai to catch and barbecue a whale belonging to a species recognised as endangered.

Table 1: PDEV entry for *barbecue*

illustrated in Table 1. This suggests that *barbecue* is only used in this meaning, and that the subject can only be a Human, while the object can only be of type Food. In other words, one can unambiguously collect Food instances by looking at the nouns that occur as objects of the verb *barbecue*.

Out of the 9,200 subject and object slots included in the current version of PDEV that totals over 4,600 patterns, we identified 741 unambiguous slots. An unambiguous slot can either be the subject or the object slot of a verb that is characterized by no semantic alternation (i.e., only one ST) in that particular slot across all patterns of the verb which take the slot. We found that these 741 instances of unambiguous slots account for 66 different STs. We selected 12 of the most productive STs for our experiments. The experiments described in this paper focus on identifying common nouns that can populate the following target STs: Activity, Body_Part, Document, Eventuality, Food, Human_Group, Inanimate, Institution, Liquid, Location, Proposition, and State_of_Affairs. In total there are 70 verbs that take the STs above unambiguously as subject or object.

2.3 Web Corpus Data

For our experiments, we use the EnClueWeb09 corpus (Pomikalek et al., 2012; Kilgarriff et al., 2014), a large web-scale corpus (70 billion words) from which we extract for a given ST up to 50,000 concordances for each of the verbs that unambiguously take that particular ST in a subject or object slot. The resulting corpus, named Web-12, includes 3.6m sentences and 97m words, and has been parsed using the Stanford Parser (Klein and Manning, 2003).

2.4 Gold Standard for Automatic Evaluation

This paper proposes two different evaluations, an automatic one to evaluate system recall, and a manual one to evaluate system precision. The automatic evaluation is based on a gold standard ST lexicon, named WN, based on a mapping between WordNet synsets and the 12 STs, manually prepared by a CPA lexicographer[3]. Proper nouns and multi-word expressions were filtered out, as the techniques presented in this paper target single-word common nouns.

Two other gold standards were produced out of WN: WN-web containing nouns from WN that are also present in the Web-12 corpus, and WN-web-dep which contains nouns from WN which also occur in a dependency relation to one of the ST-indicative verbs according to the Stanford parser.

3 Ontology Population Techniques

This section describes the ontology population techniques implemented to automatically extract new instances belonging to each target ST.

3.1 Lexical Patterns

Hearst's patterns (Hearst, 1992) consist of regular expressions made up of lexical clues to collect hypernymy relations. Each pattern contains two slots: one for the hypernym (in our case the ST, e.g., Food), and one for the hyponym (in our case the ST instance, e.g., *fish*).

These patterns generally yield nouns with a satisfying precision. For this reason they are used as a starting point of more complex ontology population systems (Snow et al., 2005; Kozareva et al., 2008; Kozareva et al., 2009). In this paper, we use the patterns listed in (Etzioni et al., 2005). We evaluate two setups for our set of 12 STs: the first is applied to our Web-12 corpus (System S1), and the other is applied to the whole EnClueWeb09 repository (System S1+). Table 2 lists the most productive patterns used by System S1 together with the number of extractions and unique nouns identified across all 12 STs.

3.2 Surface Patterns

Another popular ontology population method is to automatically extract patterns that can reliably identify ST members. A pattern extraction technique particularly used for relation extraction relies on identifying sequences of words between

[3]The lexicographer identified links based on the gloss of a WordNet synset, and on the overlap between its hyponyms and the ST in the CPA ontology

Pattern	Extr.	Nouns
System S1 (12 STs)		
ST,? (such as|especially|including) N	353525	26062
ST,? (and|or) other N	225455	14226
such ST as N	3302	1873
N is a ST	1117367	58023
System S2 (Food)		
V ing N	82918	9675
V ed N	77005	7106
V d N	49553	9045
System S2+ (Food)		
, V ing N and	3056	157
, V ing N or	1187	30
, V d N ,	814	183
System S3+ (Food)		
||VBN| ||NN|nsubjpass |be||auxpass	7068	3484
||VBN| ||NN|dobj_enum ||NN|dobj_enum	6568	1247
||VBN| ||NNS|nsubjpass |be||auxpass	4090	2180

Table 2: Examples of patterns for each system

two entities of interest (Ravichandran and Hovy, 2002; Pantel and Pennacchiotti, 2006).

System 2 adapts this method by considering as relation boundaries the verb and the ST instance. Given a category and a set of seed words, it first extracts any string occurring between the verb and each seed. Patterns are built using the extracted strings and the verbs that unambiguously combine with an ST, and then applied to Web-12 to extract new instances of a given ST. The words extracted by these patterns are ordered by frequency.

This approach is evaluated in two setups, one deriving the pattern from only the string linking the verb to the noun (System S2), and another one that also includes a context word to the left and to the right of the verb and noun pair (System S2+). Table 2 provides examples of the most productive patterns for both setups applied to the Food ST. One may notice a dramatic drop in the number of extractions when including the outward context (System S2+), but also the fact that the most frequent patterns mostly capture suffix variation, determiners, prepositions, or punctuation. Clearly those patterns are applicable to many verbs, but specifically capture Food items due to the semantic preferences of the verbs they combine with.

3.3 Syntax-driven Techniques

Syntactic dependencies offer an attractive representation of the context of a verb which allows to abstract away from undesirable variation, such as word order, or insertion of modifiers or appositions. For example, the same direct object relation between *opakapaka* and *barbecue* holds in the following two sentences: *"he barbecued opakapakas"* and *"he barbecued several times opakapakas"*. Thus syntactic relations such as direct object can be used to retrieve instances of a ST in the predicted slot extracted from PDEV. System S3 relies on this assumption and populates a ST with all the nouns that occur in the unambiguous slots of the verbs that are indicative of that ST (e.g., for the Food ST, S3 will extract all the nouns that are direct objects of the verbs *barbecue*, *brown*, *fry*, *masticate*, *overcook*, *scoff*, *vomit*, and *wolf*).

Apart from this setting, we have experimented with learning syntactic patterns from the Web-12 corpus parsed with the Stanford parser. For each ST and each verb unambiguously taking the ST as subject/object, all verb occurrences were extracted together with their direct syntactic dependents, as well as dependents indirectly connected to the verb via coordination with a direct dependent. Each verb context is a combination of tokens represented as WORD|LEMMA|POS|DEPREL, where WORD, LEMMA, POS and DEPREL correspond, respectively, to the word, lemma, part of speech and dependency relation associated to the word. Patterns are then learned by System S3+, and examples of the most frequent patterns learned by S3+ are shown in Table 2.

3.4 Bootstrapped Learning and Ranking

Pattern-based approaches for ontology population are commonly used as part of a bootstrapping algorithm (Hearst, 1992; Ravichandran and Hovy, 2002; Etzioni et al., 2005; Pantel and Pennacchiotti, 2006). For comparison purposes, we apply an iterative ranking method inspired from the work of Thelen and Riloff (2002) to the output of the pattern-driven techniques presented above. At each iteration, the learned patterns are ranked according to their tendency to extract ST members and only the best patterns drive the extraction of new ST candidates which also undergo a ranking process to enable the selection of a fixed number of top nouns to be added to the ST lexicon. This method uses at each iteration the latest ST lexicon to rank and select a pattern pool. The bootstrapping process starts with the same set of 10 seeds which was used by the pattern extraction techniques, and the process is repeated until a certain number of extractions (in our case 500) is reached.

A pool of patterns is extracted from the whole set of patterns following a pattern ranking process that relies on scores calculated using Formula 1.

$$Score(pat_i) = \frac{F_i}{N_i} \times \log_2(F_i) \quad (1)$$

where F_i is the number of ST members extracted by pat_i, and N_i is the total number of nouns extracted by pat_i. This formula captures the insight that good patterns are those that capture a large portion of known category members at time t. The top $nP + i$ patterns are placed in the pattern pool, where nP is a fixed value, and i starts from 0 and is incremented at each iteration, to ensure constant addition of new patterns and renewal of the pattern pool. All the nouns extracted by patterns from the pattern pool are scored according to Formula 2.

$$S1(noun_i) = \frac{\sum_{j_i}^{P_i} \log_2(F_j + 1)}{P_i} \quad (2)$$

where P_i is the number of patterns that extract $word_i$, and F_j is the number of distinct category members extracted by pattern j. This formula captures the intuition that a good candidate is extracted by patterns that extract a large number of category members. The top nN candidates, where nN is a fixed number, are added to the ST lexicon which will be used in the next iteration.

4 Evaluation

4.1 Automatic Evaluation

The bootstrapping process described in Section 3.4 is applied in turn to each technique described in Sections 3.1, 3.2, and 3.3, with the exception of S1 which extracts very few nouns, and S3 which does not use patterns. A grid search is performed to obtain the best parameters for the number of patterns (nP) to be included in the pattern pool, and for the number of top nouns (nN) to be added to the lexicon at each iteration, using values from the set $5, 10, 20, 50$. The best systems were those which had the best macro-average precision at 500 extractions, specifically nN=5 and nP=50 for the lexical system S1+, nN=50 and nP=50 for both surface systems S2 and S2+, and nP=5 and nN=20 for the syntactic system S3+. Table 3 shows the results as averages over the 12 ST against WN-web-dep and can be compared to Table 4, which provides the results of each technique being applied only once on the Web-12 corpus and having its extractions ranked according to frequency of extraction. The results are somewhat surprising as the bootstrapped learning and ranking method has a particular negative effect on lexical and surface systems. This suggests that this bootstrapping method is better suited to syntactic patterns than to other systems. If we consider S1+ and S2, one reason might be that these systems extract patterns which have a large number of extractions (see table 2), and are therefore not sufficiently constrained. S2+, on the contrary, extracts more precise patterns in comparison with S2, but the trade-off is a lower number of extractions. Finally, syntactic patterns produce patterns which, on average, have a number of extractions only twice as much as noun types (see table 2), whereas lexical systems have a much larger discrepancy between the number of extractions and distinct noun types. We will explore this issue in future work, and investigate ranking methods which are more generic.

4.2 Manual Evaluation

In order to get a clear idea of systems' precision, a manual evaluation process focused on four STs (Document, Food, Liquid, and Location) and an annotation of the top 500 nouns extracted by bootstrapped learning and ranking with syntactic patterns[4] (S3+) was performed for each of the four STs. Each ST noun set was manually annotated by a different pair of 4 annotators. As the system extracts the nouns from the web, the extractions often yield knowledge unfamiliar to the annotator, and therefore, to be fair with the system, it is important to allow annotators access to encyclopaedia and dictionaries to learn what a word means (e.g. "opakapaka is a fish"), and if an established word use exists (e.g., *report* is not only a Speech Act: *His report of the conference was bleak.*, but also a Document: *He printed the report.*)

Human annotators had to assess whether a noun can or cannot be interpreted as a member of a given ST (i.e., provide a "yes"/"no" annotation for every noun in the top 500 extracted by the system), but at the same time the annotators had the option to provide a less categorical decision for nouns that they were unable to decide on (i.e., assign "maybe" to nouns they were unsure about). The annotation process consisted of two rounds. As the first round produced low agreement due to unforeseen difficulties, the guidelines were revised and clarified, and a second round was performed. The issues causing disagreement mainly concerned:

[4]This was the best performing ontology population technique, and was thus chosen as target for manual evaluation.

	Precision				Recall			
topN	S1	S2	S2+	S3+	S1	S2	S2+	S3+
100	0.037	0.070	0.047	0.312	0.002	0.020	0.011	0.058
200	0.036	0.068	0.047	0.285	0.005	0.035	0.017	0.107
300	0.036	0.057	0.043	0.254	0.010	0.044	0.021	0.141
400	0.033	0.050	0.044	0.236	0.012	0.053	0.023	0.167
500	0.036	0.045	0.044	0.218	0.017	0.059	0.023	0.188

Table 3: Bootstrapped ranking: precision and recall against WN-web-dep at 500 extractions

	Precision				Recall			
topN	S1+	S2	S2+	S3+	S1	S2	S2+	S3+
100	0.106	0.164	0.247	0.337	0.020	0.038	0.062	0.066
200	0.090	0.156	0.193	0.273	0.032	0.070	0.088	0.099
300	0.080	0.146	0.170	0.245	0.043	0.092	0.113	0.136
400	0.077	0.141	0.155	0.223	0.056	0.118	0.134	0.156
500	0.072	0.137	0.146	0.207	0.063	0.146	0.156	0.175

Table 4: Frequency ranking: precision and recall against WN-web-dep at 500 extractions

1. the difficulty in evaluating a noun out of context ('slice', 'course' for Food): the revised guidelines specified clearly that these cases should be marked as "maybe";
2. general nouns that are not prototypically ST instances, but can be used in a context to refer to an ST member without making the sentence semantically anomalous (e.g., *thing* standing for a Food item): these nouns should be marked as "maybe";
3. regular category shifts, e.g. the Food category includes Dishes (*pudding*), but also Animals, Vegetables, Insects, Fruits, etc.: these nouns should be assigned "yes".

Tables 5 and 6 report inter-annotator agreement for each annotation round. The output of the second annotation round shows a good/very good agreement and was used to build two gold standard sets for each ST. The instances considered by both annotators as true ST members ("yes") are included in the gold standard HUM-STRICT. To this set we add all potential ST members ("maybe") agreed on by both annotators to obtain the second gold standard HUM-RELAXED.

4.3 Manual Evaluation Results

The evaluations results of S3+ are presented in Tables 7 and 8: one strict evaluation against HUM-STRICT, and another relaxed evaluation against HUM-RELAXED, respectively. The difference between the results obtained on the two gold standards is less than 0.1 in precision, therefore the potential ST members have limited impact. Precision drops as more candidates are extracted, in agreement with the so-called 'semantic drift' tendency also observed by other authors (Komachi et al., 2008). We can also observe that precision drops more significantly for some categories such as Liquid and Document.

Category	Pairwise	Cohen K	Fleiss K
Document	66.7%	0.433	0.407
Food	89%	0.758	0.758
Liquid	87.2%	0.717	0.716
Location	73.3%	0.486	0.473

Table 5: Inter-annotator agreement, round 1

Category	Pairwise	Cohen K	Fleiss K
Document	85%	0.739	0.738
Food	92.6%	0.84	0.84
Liquid	96.8%	0.932	0.932
Location	88.2%	0.674	0.674

Table 6: Inter-annotator agreement, round 2

topN	Document	Liquid	Location	Food	Average
100	0.84	0.65	0.96	0.89	0.835
200	0.675	0.475	0.88	0.79	0.705
300	0.543	0.393	0.83	0.763	0.632
400	0.445	0.372	0.785	0.72	0.581
500	0.414	0.332	0.73	0.652	0.532

Table 7: Precision for S3+ on HUM-STRICT

topN	Document	Liquid	Location	Food	Average
100	0.92	0.67	0.96	0.9	0.863
200	0.745	0.49	0.925	0.8	0.74
300	0.627	0.41	0.89	0.78	0.677
400	0.525	0.39	0.85	0.748	0.628
500	0.498	0.352	0.798	0.678	0.582

Table 8: Precision for S3+ on HUM-RELAXED

However, when compared to results presented in Section 4.1, we can see a clear improvement, possibly due to a non-optimal mapping between the CPA Ontology and WordNet, but also explainable by ST members correctly extracted from the web, but absent from WordNet. The next subsection looks into this in more detail.

4.4 Comparison Between Gold Standards

Results on the manual reference have shown that a large portion of true candidates (HUM-STRICT) are not in WN, the resource built by mapping CPA STs to WordNet synsets and extracting all their hyponyms. An analysis of the nouns marked by annotators as true members of an ST (HUM-STRICT), but not included in WN, has revealed the following across the four target STs (Document, Food, Liquid, and Location). Out of the total number of 2,000 manually annotated nouns corresponding to the four STs, there are 623 nouns present in HUM-STRICT, but absent from WN. A percentage of 12% of these nouns are not in WordNet. They include foreign words used in English texts (e.g., Document: fiche <French for *index card* or *form*>, Food: pancetta <Italian for *bacon*> and *kielbasa* <Polish for *sausage*>), trademarks used as common nouns (e.g., Liquid: *frappuccino*, Food: *mcmuffin*), English common nouns absent from WordNet (e.g., Location: *forestland*), collapsed multiword expressions appearing as two-word expressions in WordNet (e.g., Food: *fastfood*, Liquid: *potlikker*), and obvious misspellings (e.g., Food: *vegtable*, *buritto*).

The remaining 88% of the nouns are present in WordNet, but are not included in WN due to two main reasons. Firstly, the mapping between the CPA Ontology and WordNet is not optimal and other WordNet subtrees can be added to each ST. The Food ST for example was populated with nouns found in the subtree corresponding to the synset *food#2*. An analysis of the nouns marked as food items by the annotators, but missing from the WN Food ST has revealed that the WordNet subtrees headed by *dish#2* and *course#7* can also be added to this ST. Secondly, there are cases when one would have to add many WordNet leaf synsets that are not grouped into a higher-level subhierarchy mappable to a CPA ST. In the case of the Liquid ST for example, there are many instances of liquid sauces (e.g., *vinegar*, *salsa*, *ketchup*) that are subsumed by *condiment#1*, but since many condiments come as powders, one cannot add the subtree headed by *condiment#1* to the Liquid ST, but should instead add individual synsets scattered across WordNet. Future work will address these issues in order to better align these resources.

5 Conclusions and Perspectives

Three types of ontology population techniques have been experimented in this paper: a lexical approach that draws on Hearst's patterns, a surface approach that looks at surface strings joining an ST-preferring verb with a candidate noun, and a syntactic approach that relies on patterns drawn from dependency relations connecting an ST-indicative verb with a candidate noun. A bootstrapped learning and ranking approach is then applied to each pattern-driven technique. These techniques are applied to a web corpus built by extracting a high number of concordance lines for 70 verbs unambiguously associated with 12 target STs via their semantic preferences extracted from PDEV, and then evaluated by ranking their outputs both frequency-wise and using the bootstrapped learning and ranking approach. The best 500 extractions yielded by each technique are assessed against a resource derived as a result of mapping each CPA ST to WordNet sub-hierarchies.

A manual annotation of the top 500 nouns extracted by the best system for four STs, namely Document, Food, Liquid and Location is then performed. All experiments indicate that the syntactic approach is superior to employing lexical patterns and surface patterns for ontology population.

The results of this article point to the difficulty in evaluating pattern-driven ontology population methods. The main reasons are that existing resources have limited coverage of nouns in a given usage, which is contextual. Intrinsic categorization of nouns offers a limited appreciation of system performance.

This work is the first to use semantic preferences from PDEV for ontology population from the web, therefore it is still work in progress. Particularly important is to investigate the best use of the ontology structure as part of pattern extraction algorithms. Bootstrapped learning and ranking has had limited impact on system precision, and we believe this is one place where future efforts should be concentrated. Since the present paper only investigates semantic preferences of PDEV verbs for 12 STs, it is important to extend this work to other categories. Another specific area of interest is the use of extractions from unambiguous semantic preferences data to disambiguate ambiguous contexts and verbs.

Acknowledgements

We are very grateful to anonymous reviewers for suggestions to improve this work. We would like to thank Emma Franklin for taking part in the annotation. This work was partly supported by a AHRC grant [DVC, AH/J005940/1, 2012-2015] and the EXPERT project (People Programme (Marie Curie Actions) of the EU FP7 (REA grant agreement 317471)).

References

Guy Aston and Lou Burnard. 1998. *The BNC handbook*. Edinburgh University Press, Edinburgh.

Ismaïl El Maarouf, Jane Bradbury, Vít Baisa and Patrick Hanks. 2014. *Disambiguating Verbs by Collocation: Corpus Lexicography meets Natural Language Processing*. *Proceedings of LREC*, 1001–1006.

Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, Alexander Yates. 2005. *Unsupervised named-entity extraction from the web: An experimental study*. *Artificial Intelligence*,165(1):91134.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Patrick Hanks and James Pustejovsky. 2005. *A Pattern Dictionary for Natural Language Processing*. *Revue Française de linguistique applique*, 10:2.

Patrick Hanks. 2004. *Corpus Pattern Analysis*. G. Williams and S. Vessier (eds.), *Euralex Proceedings*, Vol. 1.

Patrick Hanks. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press, Cambridge, MA.

Marti Hearst. 1992. *Automatic acquisition of hyponyms from large text corpora*. *Proceedings of COLING-92*, 539–545.

Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. *The Sketch Engine: ten years on*. *Lexicography*, 1(1):7–36.

Dan Klein and Christopher D. Manning. 2003. *Accurate Unlexicalized Parsing*. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, 423–430.

Komachi, Mamoru and Kudo, Taku and Shimbo, Masashi and Matsumoto, Yuji. 2008. *Graph-based Analysis of Semantic Drift in Espresso-like Bootstrapping Algorithms*. *Proceedings of EMNLP*, 1011–1020.

Zornitsa Kozareva, Ellen Riloff and Eduard H. Hovy. 2008. *Semantic Class Learning from the Web with Hyponym Pattern Linkage Graphs*. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, 1048–1056.

Zornitsa Kozareva and Eduard H. Hovy and Ellen Riloff. 2009. *Learning and Evaluating the Content and Structure of a Term Taxonomy*. *Learning by Reading and Learning to Read, Papers from the 2009 AAAI Spring Symposium, Technical Report SS-09-07*, 50–57.

Patrick Pantel and Marco Pennacchiotti. 2006. *Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations*. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, 113–120.

Jan Pomikalek, Miloš Jakubíček and Pavel Rychlý. 2012. *Building a 70 billion word corpus of English from ClueWeb*. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC12)*.

James Pustejovsky, Catherine Havasi, Jessica Littman, Anna Rumshisky and Marc Verhagen. 2006. *Towards a Generative Lexical Resource: The Brandeis Semantic Ontology*. *Proceedings of LREC 2006*.

Deepak Ravichandran and Eduard Hovy. 2002. *Learning surface text patterns for a question answering system*. *Proceedings of ACL-2002*, 41–47.

Philip Resnik. 1997. *Selectional Preferences and Sense Disambiguation*. *Proceedings of the ANLP Workshop "Tagging Text with Lexical Semantics: Why What and How?"*.

Rion Snow, Daniel Jurafsky and Andrew Y. Ng. 2005. *Learning syntactic patterns for automatic hypernym discovery*. *Advances in Neural Information Processing Systems 18 (NIPS 2005)*.

Michael Thelen and Ellen Riloff. 2002. *A Bootstrapping Method for Learning Semantic Lexicons Using Extraction Pattern Contexts*. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Volume 10, 214–221.

Anna Wierzbicka. 1984. *Apples are not a kind of fruit: the semantics of human categorization*. *American Ethnologist*, Vol. 11, No. 2, 313–328.