Learning Fine-Grained Selectional Restrictions

Yongmei Tan
School of Computer Science
Beijing University of Posts and Telecommunications
Beijing 100876, China
ymtan@bupt.edu.cn

Eduard Hovy
Language Technologies Institute
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA
25213, USA
hovy@cmu.edu

Abstract

Selectional Restrictions are essential for advanced text understanding. Learning selectional restrictions is a challenge task due to the fact that there is no perfect solution for determining the appropriate level(s) of generalisation to use in specifying the restrictions on a given argument of a verb predicate and the data sparseness problem. This paper presents a novel WordNet-based framework to choose the appropriate level(s) of generalisation in the WordNet hierarchy and then estimate the probability of any word to fill a relation under a predicate from a large corpus. We use Mechanical Turk (MTurk) annotation to evaluate the performance of our proposed framework over NYT data set and empirical results show that our framework is effective.

1 Introduction

Selectional restrictions are limitations on the applicability of arguments such as nouns to predicates such as verbs (Resnik, 1993, 1996). For example, in the sentence 'Alex won't eat much bread or meat', the predicate 'eat' has 'bread' and 'meat' as its arguments. The predicate has restrictions for the semantic class membership of the arguments filling each role (in this case, direct object).

Selectional restrictions are very useful for a wide range of NLP tasks including semantic role labeling (Gildea and Jurafsky, 2002; Zapirain et al., 2009, 2010), pronoun resolution (Bergsma et al., 2008), textual inference (Pantel et al., 2007), word-sense disambiguation (Resnik, 1997; McCarthy and Carroll, 2003), and many more.

Traditionally, selectional restrictions on verb relation fillers have been specified using carefully selected general terms (concepts), as in

<ANIMATE> eat <FOOD>

(Wilks 1975; Fass and Wilks 1983). The principal problem in determining such selectional restrictions for a given verb sense and role is deciding which generalisation(s) (classes) are chosen. An overly specific generalisation may disallow some valid fillers, and an overly general one may allow some invalid ones. Furthermore, a restriction that holds for one language, such as ANIMATE as subject of eat in English, might need to be specialized, as in German [MENSCH (person) essen (eat)] vs. [TIER (animal) fressen (eat/graze)].

Rather than attempt to learn a specific generalisation for each position of each verb, the word-based approach instead attempts to determine the probability for any common noun to fill a given position of any verb. Suppose that the data available to us are triples automatically extracted from a corpus using existing techniques. By counting the frequency of occurrence of each noun for each relation with a predicate, the 'goodness' can be estimated. But its weakness lies in that the strong assumptions on sufficient observation data.

In this paper we develop a novel framework to learn fine-grained selectional restrictions, which can find the right level(s) in WordNet and well estimated plausibility for arguments that were not seen in corpus. We collect a list about the surface forms of arguments and record the frequency count for each predicate type (a predicate is corresponding to a pairing of a verb v and a relation r) and argument type (an argument is corresponding to a triple of a verb v, a relation r and a noun c). Using this list and count information, we can generate a candidate list for each predicate type. We leverage these resources to define the selection score for each candidate level. Subsequently, we choose all the appropriate

27th Pacific Asia Conference on Language, Information, and Computation pages 126－135

level(s) for each predicate type from candidate levels. Furthermore, we define the coverage ratio for each appropriate level to measure the coherence between the level and the whole corpus. And we can give a rank to the candidate list for each predicate type. In addition, we extend the probabilities to semantically related words that do not appear in the corpus, by smoothing the probability while attenuating it as semantic distance increases between nouns seen in the corpus and their unseen neighbors. To validate the effectiveness of our framework, we empirically evaluate it over a general triple store (GTS). The experimental results show that our framework is effective. The main contributions of this paper are summarized as follows.

- We propose a novel framework for learning fine-grained selectional restrictions that overcomes both the generalisation level(s) problem of the class-based method and the data sparseness problem of the word-based method.
- We have created a new huge GTS from the New York Times (NYT) for language processing and other data interpretation task.
- We extensively use Mechanical Turk (MTurk) to evaluate the framework's performance over GTS.

2 Related Work

There has been a substantial amount of research on selectional restrictions. For example: Resnik (1993) defines selectional association as an information-theoretic measure of semantic fit of a particular semantic class as an argument to a predicate. Li and Abe (1998) use the Minimum Description Length (MDL) principle to select the appropriate class. Dagan et al. (1999) propose probabilistic word association models based on distributional word similarity, and apply them to two tasks, language modeling and pseudo-word disambiguation. Rooth et al. (1999) generalize over seen headwords using EM-based clustering. Abney and Light (1999) propose Hidden Markov Models as a way of deriving selectional restrictions over words, senses, or even classes, whereas Ciaramita and Johnson (2000) use Bayesian Belief Networks to quantify selectional restrictions. Clark and Weir (2002) employ hypothesis testing. Pantel (2007) learns selectional restrictions with the semantic classes from CBC and WordNet respectively and uses these selectional restrictions in filtering out incorrect inferences. Erk (2007) shows the similarity-based method over Resnik's information-theoretic class-based method on a pseudo-disambiguation evaluation. Alan et al. (2010) present a LinkLDA approach to computing selectional restrictions, as evaluation is performed on pseudo-disambiguation and textual inference.

Much recent work described above has focused on purely distributional methods and do not use a predefined hierarchy but learn to make generalisations about predicates and arguments from corpus observations alone (Seaghdha and Korhonen, 2012). While there are many manual collections of semantic classes including hierarchies such as WordNet (Fellbaum, 1998), Levin verb classes (Levin, 1993) and FrameNet (Baker et al., 1998). The predefined hierarchy we adopt in this work is WordNet, an open domain ontology.

3 Learning Fine-Grained Selectional Restrictions

3.1 Overview

In this paper, we propose a novel framework to tackle learning fine-grained selectional restrictions with two modules as follows:

- Finding the right level(s) in WordNet

Normally, in WordNet hierarchy, the words at the higher hierarchical levels (the shorter distance between root and these words) are more general and then they provide no discriminatory power among their hyponyms. The words at the lower hierarchical levels (the longer distance between root and these words) are more specific and then offer little generalisation and be applied in only extremely few cases. Neither the words at the higher nor the words at the lower hierarchical levels could well represent the selectional restrictions for a relation r with a predicate v.

- Estimating plausibility for unseen arguments

To address the problem that data sparseness will result in estimating many probability parameters to be zero, we assign probability values to words encountered in the corpus and then use WordNet to provide an estimate of the semantic relatedness of unseen words to words that have nonzero probability. Based on the degree of semantic relatedness, unseen words can then be given some probability, smoothed or attenuated in some ways.

Those two modules are introduced in the following subsections in details.

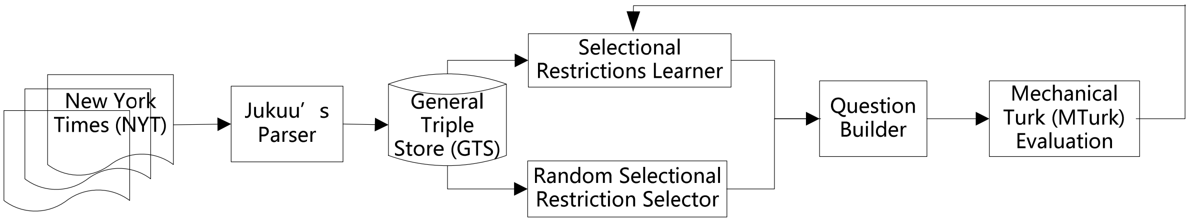

Figure 1. Overview of finding the right level(s) in WordNet

3.2 Finding the Right Level(s) in WordNet

Figure 1 illustrates the process of finding the right level(s) in WordNet. Our module learns selectional restrictions and randomly selects a list of potential selectional restrictions for a given relation *r* with a predicate (verb) *v*.

First, we parse 31+ million sentences from the NYT (years 1994 - 2002) using Jukuu's parser. We extract all tuples from each sentence parse and obtain 270 million dependency triples that we call the GTS.

Second, as described above, the learner learns selectional restrictions from the GTS, while the selector randomly chooses exemplar words from WordNet.

Third, the Question Builder uses these selectional restrictions and examplar words to create tasks to be evaluated by Mechanical Turk.

Considering that these words in the higher or lower levels have the fewer direct hyponyms, we combine this information into the *selection score* formula to distinguish them from others.

Inspired by maximum entropy model, we also prefer the most uniform model that satisfies any given constraints. If a word has many direct hyponyms and its frequency is large and the frequency distributions are even among its direct hyponyms, this word represents the selectional restrictions better than others for the same relation under a predicate.

We investigate a model that finds suitable abstractions that can better represent the selectional restrictions for the relation *r* with a predicate *v*. It works as follows.

1) Extract all hyponyms in WordNet for the words that complete the triples $< v, r, \_ >$ and all hyponyms of all possible senses of these words are considered.

2) Sum all hyponym frequencies for each word as its frequency.

3) Use the *selection score* to choose the suitable abstraction level.

Given an argument type, we define its *entropy measure* $EnMea(< v, r, c >)$ as follows:

$$EnMea(< v, r, c >) = \sum_{x \in dihypo(c)} - \frac{count(<v,r,x>)}{\sum_{y \in dihypo(c)} count(<v,r,y>)} \times log \frac{count(<v,r,x>)}{\sum_{y \in dihypo(c)} count(<v,r,y>)}$$

where $count(< v, r, x >)$ is the frequency of the triple $< v, r, c >$ in corpus, and $dihypo(c)$ is the set of all direct hyponyms of word c in WordNet. Therefore, this definition gives higher value to the word when its direct hyponym frequencies distribution is more even and its frequency is large.

Next, we can calculate the *selection score* $sc(c|v, r)$ the probability of a predicate *v* taking argument *c* under a relation *r*.

$$sc(c|v,r) = \frac{\sum_{x \in dihyp(c)} count(<v,r,x>)^{l} \times EnMea(<v,r,c>)}{|\{x|x \in dihypo(c)\}|}$$

where l is the parameter tuned in experiments.

To keep the candidate distribution as same as the real corpus, we use *coverage probability* $CP(c|v, r)$ to rank the top *N* candidates listed by *selection score* $sc(c|v, r)$ finally. The $CP(c|v, r)$ is defined as.

$$CP(c|v,r) = \frac{|\{x| \sum_{x \in hypo(c)} < v, r, x > \text{ in GTS}\}|}{|\{x|x \in hypo(c)\}|}$$

where $hypo(c)$ is the set of all hyponyms of synset *c*.

Verb	#sub (triples)	#sub (types)	sub relation filler example	#dobj (triples)	#dobj (types)	dobj relation filler example
Announce	68336	16818	Republican_Governor announce	864	300	announce endowment
Eat	16810	2665	Rawley eat	274	155	eat fish_stick
Enter	32199	8083	Mussina enter	698	323	enter Afghanistan
Request	8204	2404	Susan_Vaughn request	108	82	request dressing
Pay	86134	12161	tenant pay	1953	713	pay 5.7_percent
Provide	74664	15606	krona provide	2184	952	provide ideal

Table 1. Experimental data for the predicate type 'sub : verb' and 'verb : dobj'

3.3 Estimating plausibility for unseen arguments

We investigate the second module that assigns probabilities to unseen words, attenuated by the distance between them and words in WordNet that have been seen in the corpus and hence have nonzero probability. The process is described as follows.

1) From WordNet we collect all hyponyms for the suitable level(s) for fillers of the relation r under predicate v. For example, the direct object filler of the predicate *eat* can be the synset *food*, so all hyponyms of that synset are collected.

2) We calculate *selectional preferability* $SP(c|v,r)$, the probability of each level c based on the frequencies of its neighbors within a certain distance, under the assumption that a level will be acceptable for the relation r with the predicate v if its neighbors are acceptable too.

The *selectional preferability* $SP(c|v,r)$ of a concept c appearing as a selectional restriction for r with predicate v is to be interpreted as the probability that some nouns in *synset(c)* appear for r with predicate v. For example, $SP(food|eat,dobj)$ is the probability that some nouns in the synset of $food$ appear as the direct object for the verb eat[1].

The intuition is that a word w, one of the hyponyms of level c, will be an acceptable filler for relation r for predicate v if it is frequent enough in corpus and their distance is not 'too distant'.

[1]Note that no distinction is made between the different senses of a verb and that each use of a noun is assumed to correspond to exactly one concept. Differentiating senses is future work.

The following is the definition of the *selectional preferability* $SP(c|v,r)$ in the corpus:

$$SP(c|v,r) = N(v,r)\sum_{\{w||w-c|\leq Dis\}}\frac{count(w)}{Dis^k}$$

where $N(v,r)$ is a normalization constant such that the probabilities sum to 1 (i.e., $\sum_c P(c|v,r) = 1$), and Dis is the shortest path between two concepts in WordNet. w is the word that the distance between itself and c is less than or equal to Dis. $count(w)$ denotes the frequency of word w in the corpus. K is the power for the distance Dis.

4 Experiments

4.1 Data Preparation

We evaluate our framework over GTS. In GTS each word (words are stemmed; multi-word phrases are fused) corresponds to a specific node in the dependency tree. It is associated with a part-of-speech tag, and is linked to its head using the directed syntactic relation in the tree. We calculate counts for the patterns N-V (5,592,223 triple instances, corresponding to 2,242,626 types); N-V-N (6,604,005 instances; 5,311,954 types), V-P-N (11,867,014 instances; 4,775,263 types), N-V-N-P-N (531, 427 instances; 509, 804 types), and other 10 types involving adjectives, relations, etc.

To explore the influence of different verbs' usage on selection restrictions, we select 2 most common verbs, 2 more common verbs and 2 less common verbs. So there are 6 verbs used in our experiments: announce, eat, enter, pay, provide, and request. In this paper we only focus on the subject and direct object relations; see Table 1.

Pattern	P	R	F
sub : verb			
Announce	0.6	1	0.75
Eat	0.5	0.77	0.61
Enter	0.5	0.77	0.61
Pay	0.65	0.72	0.68
Provide	0.7	0.82	0.76
Request	0.65	0.93	0.76
verb : dobj			
Announce	0.7	0.93	0.8
Eat	0.55	0.92	0.69
Enter	0.55	0.92	0.69
Pay	0.3	0.86	0.44
Provide	0.5	0.67	0.57
Request	0.75	0.79	0.77

Table 3. Experimental results over GTS

4.2 Evaluation Method: Annotation

MTurk is an online service that we often use for evaluation. The basic unit of work on MTurk is called a Human Intelligence Task (HIT). Using a majority vote with the local-view, HIT is an easy way of taking advantage of the "wisdom of crowd" principle, and the majority output of MTurk workers provides a gold standard. The system's results can be compared to the answers provided by the MTurk workers.

Firstly, we select the 40 candidates for each pattern, 20 from the top-scoring relation filler candidates by learner, and 20 from WordNet by random selector. Question Builder randomly mixes these candidates and provides each possible verb-filler combination to 4 different MTurk workers who only need select one among 4 options, thereby obtaining 4 different judgments for each example. Given "announce" (v) and "direct object" (r), our framework produce an example as the following Table 2.

	Announce : dobj
Question	An **American_state** (such as or, California, Florida, Texas...) announces something.
Answer 1	Always Reasonable
Answer 2	Sometimes Reasonable
Answer 3	Never Reasonable
Answer 4	I Don't Know

Table 2. An example for <announce : dobj>

Secondly, we evaluate whether annotators agreed that the high-scoring word fillers are in fact appropriate as selectional restrictions while the low-scoring ones are not. We select the 40 candidates by learner for Question Builder, 20 from the top-scoring (most likely) relation filler candidates, and 20 from the bottom-scoring relation filler candidates. We provide each possible verb-filler combination to 4 different MTurk workers, thereby obtaining 4 different judgments for each example.

4.3 Experimental Results

There are 3 metrics used to assess our framework performance: precision P, recall R and F-measure F1, which are used in most work about selectional restrictions. Using majority agreement, we obtain a gold standard from annotators' feedbacks. Table 3 shows the results for the different predicate type: 'subj : verb' and 'verb : dobj' in our work.

We also use cohen's kappa κ to measure the agreement between each annotator and framework showed by Table 4.

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where Pr(a) is the relative observed agreement among raters, and Pr(e) is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category.

	sub : verb					verb : dobj				
Verb	A1	A2	A3	A4	Avg	A1	A2	A3	A4	Avg.
Announce	0.75	0.6	0.45	0.45	0.55	0.65	0.7	0.5	0.65	0.625
Eat	0.55	0.4	0.25	0.15	0.34	0.75	0.35	0.35	0.35	0.45
Enter	0.2	0.5	0.35	0.35	0.35	0.5	0.4	0.45	0.45	0.45
Pay	0.55	0.35	0.5	0.15	0.39	0.25	0.35	0.35	0.55	0.375
Provide	0.3	0.1	0.3	0.5	0.3	0.4	0.1	0.4	0.4	0.33
Request	0.8	0.7	0.5	0.75	0.69	0.75	0.5	0.35	0.15	0.44

Table 4. Each annotator agreement vs. framework

4.4 Comparison with Modified MDL and MDL

For comparison, we also implemented MDL method (Li and Abe, 1998). The noun taxonomy of WordNet has a structure of directed acyclic graph (DAG), and its nodes stand for a word sense and often contain several words having the same word sense. So we could not use the MDL based method directly and then we follow up their modification. Firstly, we copy each subgraph multiple parents (and its associated data) so that DAG is transformed to a tree structure. Secondly, we equally divide the observed frequency of a noun between all the nodes containing that noun. Finally, when an internal node contained nouns actually occurring in the data, we assigned the frequencies of all the nodes below it to that internal node, and excised the whole subtree (subgraph) below it.

The search space is too large with the original MDL method, so that we modify it by removing the single branches where each superclass has only one subclass.

Table 5 shows the comparison of selectional restrictions learned by our framework, modified MDL and MDL.

The selection restrictions got by modified MDL or original MDL are in the lower hierarchy within WordNet and their meanings are too specific to offer little generalization so that they are only applied in extremely few cases. While the seletional restrictions learned by our framework are suitable to better represent the selectional preferences for a relation r with a predicate v.

4.5 Comparison with Link LDA

We compare the results of our framework against the LinkLDA-SP (Ritter et. al., 2010). After obtaining a set of topics, we map the inferred topics to an equivalent class in WordNet. For each predicate type, we pick the top 6 levels. See Table 6 for comparative results.

	Our Framework		Modified MDL		MDL	
Verb	sub : verb	verb : dobj	sub : verb	verb: dobj	sub : verb	verb : dobj
Annouce	executive_department	appraisal	company	plan	company	plan
	independent_agency	increase	it	activity	it	agreement
	Gregorian_calendar_m	change_of_magnitude	government	choice	government	earnings
	digit	accomplishment	who	object	who	it
	American_state	payment	he	measure	he	result
	university	thinking	official	agreement	official	intention
Eat	large_integer	nutriment	who	substance	who	meal
	President_of_the_US	dish	I	food	I	it
	film_maker	cake	people	artifact	people	meal
	small_indefinite_qu	indefi-	he	measure	he	fish

	antity	nite_quantity				
	chemist	starches	it	it	it	sandwich
		food	artifact	relation	family	food
Enter	President_of_the_US	world_organization	who	artifact	he	market
	American_state	Asian_country	It	market	radon	game
	European_country	building	He	time_period	he	business
	large_integer	show	company	communication	I	race
	state_capital	attempt	team	cognition	metallic_element	country
	athlete	room	state	attribute	metal	United_States
Pay	executive_department	regular_payment	it	attention	it	attention
	American_state	large_integer	company	interest	pharmacy	interest
	digit	servant	who	price	surgery	price
	independent_agency	possession	people	tax	medicine	tax
	weekday	associate	government	bill	ballet	bill
	European_country	assets	I	fee	singing	fee
Provide	Gregorian_calendar_month	force	it	artifact	it	service
	executive_department	situation	company	service	computing	information
	independent_agency	assets	who	group	tactics	detail
	European_country	message	government	location	pharmacy	evidence
	Asian_country	representation	bill	information	surgery	support
		defender	program	causal_agent	medicine	access
Request	executive_department	pause	who	entity[2]	who	information
	South_American_country	support	it		it	anonymity
	independent_agency	choice	he		he	pause
	Asian_country	gift	agency		agency	recipe
	perceiver	document	administration		administration	change
		transaction	company		company	choice

Table 5. Comparison of selectional restrictions learned by our framework, modified MDL and MDL

Our Framework	LinkLDA

[2] Note that no distinction is made between the different senses of a verb and that each use of a noun is assumed to correspond to exactly one concept. Differentiating senses is future work.

Verb	sub : verb	verb : dobj	sub : verb	verb : dobj
an-nounce	executive_department	appraisal	entity	entity
	independent_agency	increase	physical_entity	abstraction
	Gregorian_calendar_m	change_of_magn	abstraction	psychologi-cal_feat
	digit	accomplishment	object	event
	American_state	payment	group	act
	university	thinking	whole	physical_entity
eat	large_integer	nutriment	entity	entity
	President_of_the_US	dish	abstraction	physical_entity
	film_maker	cake	communication	abstraction
	small_indefinite_quantity	indefinite_quantity	written_symbol	matter
	chemist	starches	symbol	object
		food	signal	whole
enter	President_of_the_US	world_organizatio n	entity	entity
	American_state	Asian_country	abstraction	abstraction
	European_country	building	physical_entity	physical_entity
	large_integer	show	communication	psychologi-cal_feat
	state_capital	attempt	symbol	object
	athlete	room	signal	event
pay	executive_department	regular_payment	entity	entity
	American_state	large_integer	abstraction	abstraction
	digit	servant	physical_entity	psychologi-cal_feat
	independent_agency	possession	communication	relation
	weekday	associate	object	possession
	European_country	assets	signal	cognition
provide	Gregori-an_calendar_month	force	entity	entity
	executive_department	situation	abstraction	abstraction
	independent_agency	assets	physical_entity	psychologi-cal_feat
	European_country	message	psychologi-cal_feat	event
	Asian_country	representation	object	physical_entity
		defender	physical_entity	act
request	executive_department	pause	entity	entity
	South_American_country	support	physical_entity	abstraction
	independent_agency	choice	abstraction	psychologi-cal_feat
	Asian_country	gift	object	event
	perceiver	document	whole	act
		transaction	causal_agent	communication

Table 6. Comparison of selectional restrictions learned by our framework and LinkLDA

In LinkLDA, we find that the subjects and the direct objects are at the higher hierarchical levels and their meanings are too general. While the subject levels and the direct object levels learned by our framework are suitable to better represent the selectional restrictions for a relation *r* with a predicate *v*.

5 Error Analysis

As a final analysis we manually inspect several incorrect selectional restrictions learned by our framework. Idiom is one common source of errors such as 'eat one's heart [out]' and 'break [the] ice'. An erroneous selectional restriction 'person eat person' is learned from our data because several animal names also have human connotations in WordNet ('snake' also means 'a deceitful or treacherous person', 'turkey' also means 'a person who does something thoughtless or annoying', 'pet' also means 'a special loved one', etc.).

Another problem for selectional restrictions learning is the noise in the data: problematic dependency triples can lead induction astray. Noise arises from errors in part of speech tagging or syntactic analysis, or due to metaphorical usage. Typically, however, 'good' examples will appear with much greater frequency.

These kinds of examples show the importance of word sense disambiguation to identify idiomatic usage and unusual senses of animal names. Simply using WordNet without further sense disambiguation is probably not advised.

6 Conclusion

Learning selectional restrictions is very important for many tasks such as semantic role labeling, pronoun resolution, textual inference, and word-sense disambiguation. Our framework determines a preferred set of nodes in WordNet, computes probabilities for them, and then propagates the probabilities to all other nodes.

The significance of this work is three-fold. First, we propose a novel framework for learning selectional restrictions that overcomes both the data sparseness problem of the word-based method and the level choice problem of the class-based method. Second, we have created a new huge GTS from the NYT for language processing and other data interpretation tasks. Third, we use MTurk to evaluate the framework's performance against human annotations and ratings.

In future, we will provide MTurk workers the option to indicate problems with the given selectional restrictions or the listed options. Workers could write in the correct selectional restrictions if they determine that it isn't present in the list of options, or the correct selectional restrictions if the one they are presented with is malformed. This allows them to correct errors made by the selectional restrictions identifier. We will also carry out a study to test the annotation interface and experiment with different ways of presenting the selectional restrictions options to workers.

Acknowledgments

References

1. Abney, S. and M. Light. Hiding a Semantic Class Hierarchy in a Markov Model. In A. Kehler and A. Stolcke (eds.), Proceedings of ACL Workshop on Unsupervised Learning in Natural Language Processing, pp. 1-8. 1999.
2. Baker C., Fillmore C. & Lowe J. 1998, The BerkeleyFrameNet project, In Proceedings of COLING/ACL pages 86–90, Montreal, Canada.
3. Berland, M. and E. Charniak, 1999. Finding Parts in Very Large Corpora. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics* (ACL).
4. Bergsma, S. D. Lin, and Randy Goebel. 2008. Discriminative learning of selectional preference from unlabeled text. EMNLP.
5. Blei, D. M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. J. Mach. Learn. Res.
6. Ciaramita, M. and M. Johnson. 2000. Explaining Away Ambiguity: Learning Verb Selectional Preference with Bayesian Networks. In Proceedings of the 18th International Conference on Computational Linguistics (COLING'00).
7. Clark, S. and D. Weir. 2002. Class-based probability estimation using a semantic hierarchy. Computational Linguistics.
8. Dagan, I., L. Lee, and F. C. N. Pereira. 1999. Similarity-based models of word cooccur- rence probabilities. In Machine Learning.
9. Erosheva, E., S. Fienberg, and J. Lafferty. 2004. Mixed-membership models of scientific pub- lications. Proceedings of the National Academy of Sciences of the United States of America.
10. Erk, K. 2007. A simple, similarity-based model for selectional preferences. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics.
11. Etzioni, O., M. Banko, and M.J. Cafarella. 2006. Machine Reading. *Proceedings of the 21st National Conference on Artificial Intelligence* (AAAI).
12. Fass, D. and Y. Wilks. 1983. Role restrictions 111: Formedness and Metaphor. *Journal of Computational Linguistics* 9, pp. 178–187.
13. Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
14. Galton, F. (1892). Finger Prints *Macmillan*, London.

15. Gayral, F., N. Pernelle, P. Saint-Dizier. 2000. On Verb Selectional Restrictions: Advantages and Limitations, pp. 57–68.
16. Gildea, D. and D. Jurafsky. 2002. Automatic labeling of semantic roles. Comput. Linguist.
17. Girju, R. 2003. Automatic Detection of Causal Relations for Question Answering. *Proceedings of the ACL Workshop on Multi-lingual Summarization and Question Answering*.
18. Hearst, M. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. *Proceedings of the 14th Conference on Computational Linguistics* (COLING), pp. 539–545.
19. Kozareva, Z., E. Riloff and E. 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: Human Language Technologies* (NAACL-HLT).
20. Levin, B. 1993. English Verb Classes and Alternations: A preliminary investigation. The University of Chicago Press.
21. Li, H. and N. Abe. Generalizing Case Frames Using a Thesaurus and the MDL Principle. Computational Linguistics, 24(2): 217-244, MIT Press, 1998.
22. Pado S. and Pado U. and Erk K.: Flexible, Corpus-Based Modelling of Human Plausibility Judgements (details). Proceedings of EMNLP-CoNLL 2007. Prague, Czech Republic.
23. Pantel, P. and M. Pennacchiotti. 2006. Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations. *Proceedings of Conference on Computational Linguistics / Association for Computational Linguistics* (COLING/ACL). pp. 113–120.
24. Pantel P., R. Bhagat, B. Coppola, T. Chklovski, and E. Hovy. 2007. Isp: Learning inferential selectional preferences. In HLT-NAACL.
25. Ravichandran, D. and E. Hovy. 2002. Learning Surface Text Patterns for a Question Answering System, pp. 41–47.
26. McCarthy, D. and J. Carroll (2003) `Disambiguating nouns, verbs and adjectives using automatically acquired selectional preferences', Computational Linguistics, 29(4). 639-654.
27. Peñas, A. and E. Hovy. 2010. Semantic Enrichment of Text with Background Knowledge. *First International Workshop on Formalisms and Methodology for Learning by Reading* (FAM-LbR), NAACL.
28. Pennacchiotti, M. and P. Pantel. 2006. Ontologizing Semantic Relations. *Proceedings of Conference on Computational Linguistics / Association for Computational Linguistics* (COLING/ACL), pp. 793–800.
29. Resnik, P. (1993). Selection and information: A class-based approach to lexical relationships. Ph.D. thesis, University of Pennsylvania.
30. Resnik, P. 1996. Selectional Constraints: An Information-Theoretic Model and its Computational Realization. *Cognition* 6, pp. 127–159.
31. Ritter, A., Mausam and O. Etzioni. 2010. A Latent Dirichlet Allocation Method for Selectional Preferences. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL 2010).
32. Rooth, M., S. Riezler, D. Prescher, G. Carroll, and F. Beil. 1999. Inducing a semantically annotated lexicon via embased clustering. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics.
33. Schubert, L. 2002. Can We Derive General World Knowledge from Texts? In M. Marcus (ed.), *Proceedings of the 2nd Int. Conference on Human Language Technology Research* (HLT 2002).
34. Séaghdha, D. and A. Korhonen. 2012. Modelling selectional preferences in a lexical hierarchy. In Proceedings of the 1st Joint Conference on Lexical and Computational Semantics (*SEM 2012). Montreal, QC.
35. Smeeton, N.C. (1985) "Early History of the Kappa Statistic" in Biometrics. Vol. 41, p.795.
36. Wilks, Y. 1975. Role Restriction. In E.L. Keenan (ed), *Formal Semantics of Natural Language*. Cambridge University Press, pp. 329–348.
37. Zapirain, B., E. Agirre, and L. Ma`rquez. 2009. Generalizing over lexical features: Selectional preferences for semantic role classification. In Proc. of ACL-IJCNLP.