Using Grammar Rule Clusters for Semantic Relation Classification

Emily Jamison

Independent Scholar

Los Alamos, NM 87544, USA

`jamison@ling.ohio-state.edu`

Abstract

Automatically-derived grammars, such as the split-and-merge model, have proven helpful in parsing (Petrov et al., 2006). As such grammars are refined, latent information is recovered which may be usable for linguistic tasks besides parsing. In this paper, we present and examine a new method of semantic relation classification: using automatically-derived grammar rule clusters as a robust knowledge source for semantic relation classification. We examine performance of this feature group on the SemEval 2010 Relation Classification corpus, and find that it improves performance over both more coarse-grained and more fine-grained syntactic and collocational features in semantic relation classification.

1 Introduction

In the process of discovering a refined grammar starting from rules in the original treebank grammar, latent-variable grammars recover latent information. Intuitively, the new split grammar states should reflect linguistic information that has been generalized from the lexical level but is not so general as the original syntactic level. While the intended use of this information is to improve syntactic parsing, the lexically-derived nature of the split grammar states suggests it may contain semantic information as well.

Petrov et al. (2006) note that while some of these split grammar states reflect true linguistic information, such as the clustering of verbs with similar de-

pendencies, other grammar states may reflect useless information, such as a split between rules that each terminate in a comma. However, it is the automatic nature of grammar splitting which shows potential for deriving semantic knowledge; such split grammar states may reflect statistical and linguistic observations not noticed by humans.

In this paper, we use this recovered latent information for the classification of semantic relations. Our goal is to determine whether recovered latent grammatical information is capable of contributing to the real-world linguistic task of relation classification. We will compare the feature performance of recovered latent information with that of other syntactic and collocational features to determine whether or not the recovered latent information is helpful in semantic relation classification.

2 Task Description

We performed the task of classifying semantic relations from SemEval 2010 Task 8: Multi-way Classification of Semantic Relations between Pairs of Nominals. Each instance consists of a sentence, marked with two nominals, e1 and e2. One of 19 possible direction-sensitive relations is annotated for each pair of nominals. Two examples are shown below.

- The <e1>author</e1> of a keygen uses a <e2>disassembler</e2> to look at the raw assembly code.
  Relation: **Instrument-Agency(e2,e1)**

- Their <e1>knowledge</e1> of the power and rank symbols of the Continental em-

pires was gained from the numerous Germanic
<e2>recruits</e2> in the Roman army, and
from the Roman practice of enfeoffing various
Germanic warrior groups with land in the imperial provinces.
Relation: **Entity-Origin(e1,e2)**

We classified the semantic relations using a
Maximum Entropy classifier. In our system,
classification was 19-way, direction-sensitive[1]
between the classifications: Entity-Origin, Entity-Destination, Cause-Effect, Product-Producer,
Content-Container, Instrument-Agency, Member-Collection, Component-Whole, Message-Topic,
and Other (non-directional). The model was trained
on the 8000-instance training section of the SemEval 2010 Task 8 Semantic Relations Corpus.
Distribution of the training data is shown in Table 1
(Hendrickx et al., 2010).

Class	count	% of Data
Other	1410	17.63%
Cause-Effect	1003	12.54%
Component-Whole	941	11.76%
Entity-Destination	845	10.56%
Product-Producer	717	8.96%
Entity-Origin	716	8.95%
Member-Collection	690	8.63%
Message-Topic	634	7.92%
Content-Container	540	6.75%
Instrument-Agency	504	6.30%

Table 1: Class distribution in the training section of SemEval 2010 Task 8 Semantic Relations Corpus.

We tested the model on the 2717-instance testing
section of the same corpus. For each instance, the
user was provided with a sentence containing two
marked entities, `e1` and `e2`. We structured the task
such that, for each instance, we chose the best semantic relation out of the 19 available.

In this paper, we use grammatical cluster information (i.e., recovered latent information) from the
Berkeley Parser (Petrov 2006) as semantic features
of syntactic origin to classify semantic relations in
the SemEval 2010 Semantic Relations corpus, in a

---

[1]i.e., with a Content-Container relation, the nominal that is
the container and the nominal that is the content cannot be reversed.

Maximum Entropy model. We conduct two sets of
experiments. In the first experiment, we examine the
effect of using Berkeley Parser latent cluster features
to enhance specificity over more general features
(POS tags and others), where the cluster features are
inherently more closely tuned with the data than the
other features, and more likely to lead to an overfitted model. In the second experiment, we examine
the effect of using cluster features to enhance generalizability over more specific features (the words of
the cluster features' terminal nodes), in which case
the cluster features generalize over othe more specific features, but are more likely to miss detailed
patterns.

2.1 Previous Work

The classification of semantic relations has been
proposed to help NLP tasks ranging from word sense
disambiguation, language modelling, paraphrasing,
and recognising textual entailment (Hendrickx et al.,
2010).

Semantic world knowledge is crucial for accurate
semantic classification of many types, and sources
range from the hand-crafted-yet sparse (such as
WordNet) to the robust-yet-noisy (such as the Internet). For this community task, teams proposed
a variety of knowledge sources and other features for their relation classification, from knowledge databases (Tymoshenko and Giuliano, 2010),
WordNet (Rink and Harabagiu, 2010), Wikipedia
(Szarvas and Gurevych, 2010), to formal linguistic
Levin classes (Rink and Harabagiu, 2010), to collocational metrics (Rink and Harabagiu, 2010) and
stems (Chen et al., 2010).

Syntactic features present special benefits to any
semantic classification task: they can generalize
over the local context in ways that collocational metrics cannot, and unlike knowledge database sources
which assign the most common word sense to a
word, syntactic features are sensitive to the word's
sense, as determined by the local context of the
word. Several teams in SemEval 2010 Task 8 used
syntactic features for semantic relation classification. Chen et al. (2010) use a feature set of the
syntactic parent node held in common by the two
nominals. Rink and Harabagiu (2010) use a feature
set of dependency paths of length 1 or 2 from the
dependency tree around the two nominals.

47

2.2 Grammatical cluster information

For our investigations, we used the Berkeley Parser (Petrov et al 2006, Petrov and Klein 2007) as a source of grammar rule clusters. We used the eng_sm6.gr off-the-shelf model.

The Berkeley Parser starts with an initial grammar extracted from Wall Street Journal corpus sections 2-22. The parser then tries to learn a set of rule probabilities over latent annotations to maximize the likelihood of the training trees using Expectation-Maximization (EM).

Consider a sentence w and its unannotated tree T, a non-terminal A spanning (r, t), and its children B and C spanning (r, s) and (s, t). A_x is a subsymbol of A, B_x of B, and C_x of C. We calculate the posterior probability of all annotated rules and positions for each training set tree T in the Expectation step (Petrov et al., 2006):

$$(1)$$

$$P((r, s, t, A_x \to B_y C_z) \mid w, T) \propto \mathbf{P}_{\text{OUT}}(r, t, A_x)$$
$$\times \beta(A_x \to B_y C_z) \mathbf{P}_{\text{IN}}(r, s, B_y) \mathbf{P}_{\text{IN}}(s, t, C_z)$$

The probabilities from the Expectation step act as weighted observations to update the rule probabilities in the Maximization step:

$$\beta(A_x \to B_y C_z) := \frac{\#\{A_x \to B_y C_z\}}{\Sigma_{y', z'} \#\{A_x \to B_{y'} C_{z'}\}} \quad (2)$$

In each cycle of EM, the grammar is split randomly in halves, and some halves are merged back together. The grammar is retrained, and the results are used to initialize the next round of EM.

In the splitting step, all grammatical nodes are split in two. Although the grammar grows more finely fitted to the training data with each splitting step, its size quickly becomes unmanageable, its rules become overfitted, and because the splits are not a result of likelihood calculation, many unhelpful rules are produced. The merging step functions to remove unhelpful rules. In the merging step, each split is examined for the loss of likelihood removing it would cause; splits whose likelihood contribution is below a cutoff are re-combined.

The experiments we perform in this paper are a gamble on the possibility that the saved splits are picking up semantic information from the rule structure they reflect in the increased likelihood. We use the final split cluster ID's (PP-5, PP-8, etc.) as features in our experiments.[2]

2.3 Features

We used several sets of features in our experiments. All POS-tags, syntactic structure, and Cluster ID features come from the Berkeley Parser. The lemmatization comes from Morpha (Minnen et al., 2001). All features occurring less than two times in the training data were discarded, for ease of processing. A sample sentence and the resulting features are shown in Table 2. Note that all features, collocational and syntactic, were used for discovering semantic knowledge.

	The Crayola `<e1>box</e1>` *contained two* `<e2>pencils</e2>`.
SW	the-dt, crayola-jj, contain-vbd, two-cd, pencil-nns, box-nn
IBW	contained, two, contained∧two
OCW	crayola-jj, box-nn, contain-vbd, pencil-nns
POS-tags	vbd, cd, vbd∧cd
ID's	vbd6, cd1, vbd6∧cd1

Table 2: A sample sentence and its accompanying features.

Collocational Features:

- **Surrounding Words (SW):** From Ye and Baldwin's (2007) preposition sense disambiguation system, this set of features consists of lemmas of all of the words within a window of seven words before and after each of **e1** and **e2**. Features are not, however, marked with relative location, as we found that this reduced accuracy.

- **In-Between Words (IBW):** This bag of features consists of the string of words occuring in the sentence in between **e1** and **e2**, exclusive, as well as all the substrings of those consecutive words. We tried marking each feature with its relative location, but we found that results improved without location marking, and so we do not use location marking in these experiments.

---

[2]Note that cluster ID's are only meaningful when compared to other cluster ID's split from the same parent node.

Syntactic Features:

- **Open Class Words (OCW):** from Ye and Baldwin's (2007) preposition sense disambiguation system, this set of features consists of the lemmas of all of the open-class words in the sentence (i.e., NP, VP, ADJP, ADVP).

- **POS-tags:** The POS tags of the words (i.e., terminal nodes) and all consecutive strings of POS tags in between **e1** and **e2**, exclusive. Tags are from the Berkeley Parser.

- **Cluster ID's:** The Berkeley Parser syntactic rule cluster ID's and POS-tags of the terminal nodes in between **e1** and **e2**. ID numbers are only relevant when comparing ID's with the same POS tag.

3 Experiment: Cluster ID's as more spcific features

In our first experiment, we compared two systems of Surrounding Words, Open-Class Words, and In-Between Terminal Tags, with and without In-Between Terminal Cluster ID's. The results are shown in Table 3.

3.1 Results and Analysis

Table 3 shows the results of adding more specific Cluster ID features to the more general POS-tag, Open-Class, and Surrounding-Words features. While this could have led to an over-fitted model, apparently it did not. Overall precision increased from 66.60% to 68.62%, an increase of 2.02%, yet recall also increased, from 64.26% to 65.33%, an increase of 1.07%. The more precise, more closely-fitted features did not harm performance, but actually enhanced it. The Maximum Entropy learner itself preferred the Cluster ID features: Table 4 shows per-class POS-tag and Cluster-ID features with a lambda value over 0.25, comparing when both POS-tag features and Cluster ID tags are available, versus just POS-tags (all among other features used in Experiment 1). When given the opportunity, the MaxEnt learner considered the Cluster ID features more important than the POS-tag features.

As shown in Table 3, we can see that adding the Cluster ID's did mildly increase F-measure

(by 1.41 %, from 65.01% to 66.42%[3]. However, when viewed on a class-by-class basis, some classes show great improvement with the addition of Cluster ID's while others remain unchanged. The classes *Cause-Effect*, *Component-Whole*, *Content-Container*, *Instrument-Agency*, and *Message-Topic* all gained significantly with the addition of cluster ID features. We investigated important features of these classes more carefully.

Classes that significantly improved with Cluster ID's:

- **Cause-Effect**: Cluster ID features that correlated highly with **Cause-Effect**, besides keyword-type single word clusters (*from*, *that*), were a cluster of certain occurances of the prepositions *by*, *from*, *of*, *in*; and a cluster of cause-type verbs (shown in Table 5) plus the phrase *by*.

- **Content-Container**: Features positively correlated with **Content-Container**, besides some keywords and phrases such as *full of*, *was*, *in*, and *the/a*, included a Cluster ID feature with a number of verbs commonly used to refer to containers and the processes of filling and emptying them, such as *leaked*, *contained*, *poured*, *stuffed*, *took*, *injected*, *inserted*, and *found*. The verbs from this feature are listed in Table 6.

- **Instrument-Agency**: Several Cluster ID features of verbs correlate with this class. Although it is not as obvious as the verb list with **Cause-Effect**, Table 7 compares several verb clusters that did have a noticeable positive correlation with **Instrument-Agency** with several verb clusters of the same POS-tags that did not correlate.

- **Component-Whole**: Notable keyword and key phrase features include *of the/a/an*, *has a*, and *has*. One Cluster ID feature is a cluster of third-person, possessive, and reflexive pronouns. Al-

---

[3]An F-measure of 66.42% would have put our system in the middle of the pack on Task performance if it had participated in the actual SemEval 2010 Task 8. Task results for the entire dataset ranged from 82.18% F-m with a carefully-design knowledge database, to 52.16% with parse features, NE's, and semantic seed lists, and 26.67% using punctuation, prepositional patterns, and context words. Our goal, however, is to determine whether recovered latent grammatical information is capable of contributing to relation classification at all.

	Precision			Recall			F-measure		
Class	**no ID**	**w/ ID**	**diff**	**no ID**	**w/ ID**	**diff**	**no ID**	**w/ ID**	**diff**
Cause-Effect	79.43	82.62	**3.19**	76.52	76.83	0.31	77.95	79.62	**1.67**
Component-Whole	56.58	61.86	**5.28**	55.13	57.69	**2.56**	55.84	59.70	**3.86**
Content-Container	74.37	77.04	**2.67**	77.08	78.65	1.57	75.70	77.84	**2.14**
Entity-Destination	73.30	72.60	-0.70	88.36	88.01	-0.35	80.12	79.57	-0.55
Entity-Origin	67.42	66.67	-0.75	69.77	69.77	0.00	68.57	68.18	-0.35
Instrument-Agency	55.73	56.30	0.57	46.79	48.72	1.93	50.87	52.23	**1.36**
Member-Collection	68.40	67.57	-0.83	73.39	75.11	1.72	70.81	71.14	0.33
Message-Topic	65.95	70.47	**4.52**	46.74	52.11	**5.37**	54.71	59.91	**5.20**
Product-Producer	58.19	62.50	**4.31**	44.59	41.13	-3.46	50.49	49.61	-0.88
Other	30.97	28.65	**-2.32**	36.56	35.46	-1.10	33.54	31.69	-1.85
Total, Macro-Avg	66.60	68.62	2.02	64.26	65.33	1.07	65.01	66.42	1.41

Table 3: Comparison of Open-Class Words, Surrounding Words, and POS-tags, with and without Cluster ID features. Per SemEval2010 task standards, total does not include 'Other'. Directionality is evaluated, but results are combined for viewability. Bold-font differences are most notable.

	POS only	POS & Cluster ID	
Class	**POS**	**POS**	**ID's**
Cause-Effect	5	0	6
Component-Whole	4	1	6
Content-Container	4	0	7
Entity-Destination	4	0	4
Entity-Origin	2	1	6
Instrument-Agency	7	0	6
Member-Collection	3	0	2
Message-Topic	3	0	6
Product-Producer	5	1	5
Other	1	1	0

Table 4: Number of POS-tag and Cluster ID features with a lambda value over 0.25, with and without Cluster ID features being available. High lambda values are assigned when a classifier finds the features has a high positive correlation with correct examples in the training data.

though it is a somewhat rare feature, when it occurs it is positively-correlated. An example of a **Component-Whole** pronoun is below:

> He stopped rowing when the boat was opposite to the paddle wheel of the steamer, and the `<e1>steamer</e1>` stopped **her** `<e2>engine</e2>` at the same time.

accompanied, affected, built, caused, completed, composed, contained, cooked, covered, created, derived, developed, discovered, distilled, driven, enclosed, fabricated, followed, founded, generated, given, known, led, made, manufactured, obtained, offered, produced, published, raised, represented, run, shared, supported, transmitted, triggered, used, wrapped, written

Table 5: Contents of the VBN-12 Cluster that occurred 3 or more times in the Relational Semantics corpus training data. Many of the verbs denote a cause-effect relationship.

- **Message-Topic**: Some helpful keyword features were *in*, *to*, and *that*. A helpful Cluster ID feature was a cluster of the prepositions *about*, *over*, *upon*, *around*, and *between*. The model also ranked highly a Cluster ID feature containing a number of 'discussion' and 'document' verbs, shown in Table 8.

Classes that did not significantly improve with Cluster ID's:

- **Entity-Origin**: This class is suspected to have been plagued by faulty annotation. 7 out of the first 24 training examples are incorrectly marked as **Entity-Origin**, according to the corpus's definitions. This noise in the data likely prevents effective comparison of features. Despite the noise, some clusters with a high correlation to the class include: a cluster of verbs

adjusted, applied, became, brought, built, caused, contained, created, described, did, established, examined, featured, followed, formed, found, gave, included, injected, inserted, introduced, involved, joined, leaked, made, marked, posted, poured, produced, released, reported, saw, sent, spotted, stuffed, took, used, was, were, won, wrote, wrapped, written

Table 6: Contents of the VBD-5 Cluster that occurred 3 or more times in the Relational Semantics corpus training data. A number of the verbs can refer to actions involving containers and their contents.

consisting of mostly *made*, *kept*, and *left*; and a cluster of verbs consisting mostly of *made*, *left*, *kept*, *departed*, *arrived*, *travelled*, and *consisted*.

- **Entity-Destination**: The Cluster ID features that correlated highly with this class mark individual key words and phrases: *for the*, *on the*, *to the*, and *to*. Clusters of words were not helpful for this class.

- **Product-Producer**: The Cluster ID features that strongly correlated with this class consisted mostly of the words *who/whom* and *by*; individual key word features would have been just as good. Several clusters of verbs were highly correlated as well, but apparently there was too much noise in the clusters for them to be effective.

- **Member-Collection**: This class used the 'jj-24' POS cluster, which contains the word *other* among other adjectives. This is probably from the classic **Member-Collection** phrase "Y and other X's". However, this features, along with a feature mostly consisting of *of*, was not enough to make much difference (0.33%) over POS features.

- **Other**: The class *Other* decreased in F-measure with the addition of cluster ID features. Combined with the overall F-measure increase for all the regular classes, we interpret this decrease in F-measure as an increase in entropy, as more examples with identifiable

useful features are removed from the *Other* category, and the MaxEnt learner has fewer accurate patterns with which to cluster this diverse group of examples. In order words, we actually desire to see a decrease in *Other* F-measure, as the examples in *Other* have almost nothing in common with eachother and should be hard to identify.

Overall, some of the semantic relation classes were correlated with features of syntactic clusters, and the clusters boosted scores, while other classes weren't, and their scores remained roughly the same. The results of this experiment show that syntactic clusters did not lead to overtraining of data, and were helpful with semantic relation classification.

4 Experiment: Cluster ID's as more general features

In our second experiment, using the same experimental set-up but different features, we compared In-Between Words (IBW), IBW Plus POS-tags, and IBW Plus POS-tags Plus Cluster ID features. The results are shown in Table 9. The goal of this experiment is to compare Cluster ID features to an even more fine-grained feature, the words themselves. The words, POS-tags, and Cluster ID tags all concern the same nodes in the sentence.

4.1 Results and Analysis

Table 9 shows the results of adding coarser-grained Cluster ID features to the more specific In-Between-Words features, as well as to the POS-tag features. The addition of Cluster ID features improved classification over IBW plus POS-tags, as well as IBW alone. While the previous experiment showed that Cluster ID features were not too specific to be helpful, this experiment shows that they are also not too general as to blur lexical patterns. While overall F-measure increased 2.13% from IBW with the addition of POS-tag features, from 63.35% to 65.48%, F-measure also increased further by the addition of Cluster ID features to IBW plus POS-tags, with a total increase of 2.72% over IBW features alone, from 63.35% to 66.07%.

Table 10 breaks down results into Precision and Recall for the different groups of features. Since this experiment was starting with a more precise base-

Cluster	Words
	Instrument-Agency Positively-correlated Clusters:
VBD-7	approached, arrived, attached, bought, built, carried, caught, changed, chose, clicked, contained, covered, deposited, described, directed, donated, dragged, dropped, entered, erected, established, explained, fetched, fired, fled, gave, grabbed, hit, inserted, joined, kept, killed, knew, left, lived, lost, made, moved, noticed, observed, opened, organized, packed, passed, performed, placed, poured, prepared, presented, pressed, pulled, pushed, put, removed, rescheduled, saw, scaled, searched, sent, sold, spent, stirred, struck, stuffed, threw, took, tore, turned, used, was, wrote
VBZ-9	applies, assists, brings, builds, changes, comprises, considers, contains, converts, covers, creates, cuts, describes, emits, encloses, enters, gets, hits, holds, joins, keeps, leaves, makes, needs, offers, plays, portrays, prepares, provides, removes, s, spreads, stirs, studies, teaches, uses, writes
	Non-positively-correlated Clusters:
VBD-4	became, bought, carried, caused, completed, contained, created, developed, dug, filled, formed, got, had, held, issued, killed, made, presented, produced, reached, received, required, saw, showed, stopped, took, triggered
VBD-9	began, kept, started, stopped
VBD-8	continued, decided, had, happened, managed, needed, seemed, tried, used, wanted
VBD-2	found, learned, noted, noticed, read, revealed, saw
VBZ-8	arrives, brings, comes, comprises, consists, contains, contributes, copes, departs, extends, falls, feels, flows, focuses, goes, grows, hangs, leads, looks, moves, originates, passes, pulls, refers, relates, rests, results, returns, runs, s, sits, speaks, starts, stops, talks, travels, uprises

Table 7: Some positively-correlating and non-correlating verb clusters for **Instrument-Agency**. Verbs occurred at least 3 times in the Relational Semantics corpus training data. Many verbs from positively-correlating Cluster ID features may occur with mention of a tool or object to be used to carry out the action.

attaches, builds, carries, causes, combines, comprises, contains, creates, describes, discusses, encloses, gives, holds, includes, keeps, makes, manipulates, means, needs, offers, performs, presents, processes, provides, represents, requires, s, shows, takes, wears, writes

Table 8: Contents of the VBZ-11 Cluster that occurred 3 or more times in the Relational Semantics corpus training data. Many of the verbs are associated with documents or speaking.

line, IBW features, and adding coarser grained features, POS-tags and Cluster IDs, we might expect to see a simultaneous decrease in precision and increase in recall from the baseline IBW to the enhanced, POS-tag and Cluster ID versions. As can be seen in Table 10, this is exactly what happens. However, Cluster ID features are found to be helpful to the overall goal of semantic relation classification, because they increase recall by much more (4.44%) than they decrease precision (-0.44%).

Classes for which the Cluster ID plus POS-tag plus IBW combo was highest include **Content-Container**, **Entity-Destination**, **Member-Collection**, **Message-Topic**, and **Other**. **Component-Whole**, **Instrument-Agency**, and **Product-Producer** all showed gains over just IBW, but had lower scores than IBW plus just POS-tags. Only **Cause-Effect** and **Entity-Origin** failed to show any improvement with POS-tags or Cluster ID's over the baseline IBW features.

A comparison of features between Experiments 1 and 2 showed that nearly all of the significantly helpful positive-corellated Cluster ID features (with lambda greater than 0.25) in Experiment 2 were also important in Experiment 1. Some cluster ID features in Experiment 1 that isolated out a single word were replaced in Experiment 2 by a more-accurate individual IBW word feature.

5 Conclusion

In this paper, we presented a new method of semantic relation classification: using automatically-derived grammar rule clusters as a semantic knowledge source for relation classification. We tested performance of the feature on the SemEval 2010 Relation Classification corpus, and found that it improved performance over both more coarse-grained

	F-measure					
Class	**IBW**	**+POS**	**IBW+POS diff**	**+ID**	**IBW+ID diff**
Cause-Effect	**83.39**	80.19	-3.20	81.55	-1.84
Component-Whole	52.50	56.07	**3.57**	55.06	2.56
Content-Container	73.79	73.27	-0.52	75.19	**1.40**
Entity-Destination	77.98	80.06	2.08	81.49	**3.51**
Entity-Origin	**68.56**	67.21	-1.35	67.33	-1.23
Instrument-Agency	54.29	56.43	**2.14**	55.63	1.34
Member-Collection	73.22	75.30	2.08	75.50	**2.28**
Message-Topic	39.59	47.06	7.47	49.45	**9.86**
Product-Producer	46.88	53.77	**6.89**	53.46	6.58
Other	27.69	30.08	2.39	30.63	**2.94**
Total, Macro-Avg	63.35	65.48	2.13	66.07	**2.72**

Table 9: F-measure comparison of In-Between Words, IBW plus POS-tags, and IBW plus POS-tags plus Cluster ID features. Per SemEval2010 task standards, total does not include **Other**. Bold-font differences are the highest improvements (or baseline, whichever is higher).

Analysis	**iBW**	**+POS**	**iBW +POS diff**	**+ID**	**iBW +ID diff**
Precision	**67.03**	66.28	-0.75	66.59	-0.44
Recall	62.10	66.12	4.02	**66.54**	4.44

Table 10: Comparison of IBW, IBW plus POS-tags, and IBW plus POS-tags plus Cluster ID features. Per SemEval2010 Task 8 standards, total does not include **Other**. Bold-font differences are the highest improvements (or baseline).

and more fine-grained syntactic and collocational features in semantic relation classification.

Acknowledgments

The author wishes to thank William Schuler and Yannick Versley for their advice and support on this project.

References

Yuan Chen , Man Lan , Jian Su , Zhi Min Zhou , Yu Xu 2010. ECNU: Effective Semantic Relations Classification without Complicated Features or Multiple External Corpora. *Proceedings of the 5th International Workshop on Semantic Evaluation.*

Iris Hendrickx , Su Nam Kim , Zornitsa Kozareva , Preslav Nakov , Diarmuid O Seaghdha , Sebastian Pado, Marco Pennacchiotti , Lorenza Romano, Stan Szpakowicz. 2010. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. *Proceedings of the 5th International Workshop on Semantic Evaluation.*

Guido Minnen, John Carroll, and Darren Pearce. 2001. Applied morphological processing of English. *Natural Language Engineering*, 7(3):207223.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning Accurate, Compact, and Interpretable Tree Annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association of Computational Linguistics.*

Slav Petrov and Dan Klein. 2007. Improved Inference for Unlexicalized Parsing. *Proceedings of NAACL 2007.*

Bryan Rink and Sanda Harabagiu 2010. UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources. *Proceedings of the 5th International Workshop on Semantic Evaluation.*

Gyorgy Szarvas and Iryna Gurevych 2010. TUD: semantic relatedness for relation classification *Proceedings of the 5th International Workshop on Semantic Evaluation.*

Kateryna Tymoshenko and Claudio Giuliano. 2010. FBK-IRST: Semantic Relation Extraction using Cyc. *Proceedings of the 5th International Workshop on Semantic Evaluation.*

Patrick Ye and Timothy Baldwin. 2007. MELB-YB: Preposition Sense Disambiguation Using Rich Semantic Features. *Proceedings of SemEval 2007.*