

Towards a Domain Independent Semantics: Enhancing Semantic Representation with Construction Grammar

Jena D. Hwang^{1,2}

Rodney D. Nielsen¹

Martha Palmer^{1,2}

¹Ctr. for Computational Language and Education Research
University of Colorado at Boulder
Boulder, CO 80302
{hwangd, rodney.nielsen, martha.palmer}@colorado.edu

²Department of Linguistics
University of Colorado at Boulder
Boulder, CO 80302

Abstract

In Construction Grammar, structurally patterned units called constructions are assigned meaning in the same way that words are – via convention rather than composition. That is, rather than piecing semantics together from individual lexical items, Construction Grammar proposes that semantics can be assigned at the construction level. In this paper, we investigate whether a classifier can be taught to identify these constructions and consider the hypothesis that identifying construction types can improve the semantic interpretation of previously unseen predicate uses. Our results show that not only can the constructions be automatically identified with high accuracy, but the classifier also performs just as well with out-of-vocabulary predicates.

1 Introduction

The root of many challenges in natural language processing applications is the fact that humans can convey a single piece of information in numerous and creative ways. Syntactic variations (e.g. *I gave him my book.* vs. *I gave my book to him.*), the use of synonyms (e.g. *She bought a used car.* vs. *She purchased a pre-owned automobile.*) and numerous other variations can complicate the semantic analysis and the automatic understanding of a text.

Consider the following sentence.

- (1) They hissed him out of the university

While (1) is clearly understandable for humans, to automatically discern the meaning of *hissed* in this

instance would take more than learning that the verb *hiss* is defined as “make a sharp hissing sound” (WordNet 3.0). Knowing that *hiss* can also mean “a show of contempt” is helpful. However, it would also require the understanding that the sentence describes a causative event if we are to interpret this sentence as meaning something like “*They caused him to leave the university by means of hissing or contempt*”.

The problem of novel words, expressions and usages are especially significant because discriminative learning methods used for automatic text classification do not perform as well when tested on text with a feature distribution that is different from what was seen in the training data. This is recognized to be a critical issue in domain adaptation (Ben-David et. al, 2006). Whether we seek to account for words or usages that are infrequent in the training data or to adapt a trained classifier to a new domain of text that includes new vocabulary or new forms of expressions, success in overcoming these challenges partly lies in the successful identification and the use of features that generalize over linguistic variation.

In this paper we borrow from the theories presented by Construction Grammar (CxG) to explore the development of general features that may help account for the linguistic variability and creativity we see in the data. Specifically, we investigate whether a classifier can be taught to identify constructions as described by CxG and gauge their value in interpreting novel words.

The development of approaches to effectively capture such novel semantics will enhance applications requiring richer representations of language understanding such as machine

translation, information retrieval, and text summarization. Consider, for instance, the following machine translation into Spanish by the Google translate (<http://translate.google.com/>):

They hissed him out of the university.

→ Silbaban fuera de la universidad.

Tr. They were whistling outside the university.¹

The translation has absolutely no implication that a group of people did something to cause another person to leave the university. However, when the verb is changed to a verb that is seen to frequently appear in a caused motion interpretation (e.g. *throw*), the results are correct:

They threw him out of the university.

→ Lo sacaron de la universidad.

Tr. They took him out of the university.

Thus, if we could facilitate a caused motion interpretation by bootstrapping semantics from constructions (e.g. “X ___ Y *out of* Z” implies caused motion), we could enable accurate translations that otherwise would not be possible.

2 Current Approaches

In natural language processing (NLP), the issue of semantic analysis in the presence of lexical and syntactic variability is often perceived as the purview of either word sense disambiguation (WSD) or semantic role labeling (SRL) or both. In the case of WSD, the above issue is often tackled through the use of large corpora tagged with sense information to train a classifier to recognize the different shades of meaning of a semantically ambiguous word (Ng and Lee, 2006; Agirre and Edmonds, 2006). In the case of SRL, the goal is to identify each of the arguments of the predicate and label them according to their semantic relationship to the predicate (Gildea and Jurafsky, 2002).

There are several corpora available for training WSD classifiers such as WordNet’s SemCor (Miller 1995; Fellbaum 1998) and the GALE OntoNotes data (Hovy et. al., 2006). However, most, if not all, of these corpora include only a small fraction of all English predicates. Since WSD systems train separate classifiers for each

predicate, if a particular predicate does not exist in the sparse training data, a system cannot create an accurate semantic interpretation. Even if the predicate is present, the appropriate sense might not be. In such a case, the WSD will again be unable to contribute to a correct overall semantic interpretation. This is the case in example (1), where even the extremely fine-grained sense distinctions provided by WordNet do not include a sense of *hiss* that is consistent with the caused motion interpretation rendered in the example.

Available for SRL tasks are efforts such as PropBank (Palmer et al., 2005) and FrameNet (Fillmore et al., 2003) that have developed semantic role labels (based on differing approaches) and have labeled large corpora for training and testing of SRL systems. PropBank (PB) identifies and labels the semantic arguments of the verb on a verb-by-verb basis, creating a separate frameset that includes verb specific semantic roles to account for each subcategorization frame of the verb. Much like PB, FrameNet (FN) identifies and labels semantic roles, known as Frame Elements, around a relational target, usually a verb.² But unlike PB, Frame Elements less verb specific, but rather are defined in terms of semantic structures called *frames* evoked by the verb. That is, one or more verbs can be associated with a single semantic frame. Currently FN has over 2000 distinct Frame Elements.

The lexical resource VerbNet (Kipper-Schuler, 2005) details semantic classes of verbs, where a class is composed of verbs that have similar syntactic realizations, following work by Levin (1993). Verbs are grouped by their syntactic realization or *frames*, and each frame is associated with a meaning. For example, the verbs *loan* and *rent* are grouped together in class 13.1 with roughly a “give” meaning, and the verbs *deposit* and *situate* are grouped into 9.1 with roughly a “put” meaning.

Although differing in the nature of their tasks, WSD and SRL systems both treat lexical items as the source of meaning in a clause. In WSD, for every sense we need a new entry in our dictionary to be able to interpret the sentence. With SRL, we

¹ We have hand translated the Google translation back to English for comparison.

² PropBank labels Arg0 and Arg1, for the most part, correspond to Dowty’s Prototypical Agent and Prototypical Patient, respectively, providing important generalizations.

need the semantic role labels that describe the predicate argument relationships in order to extract the meaning.

In either case, we are still left with the same issue – if the meaning lies in the lexical items, how do we interpret unseen words and novel lexical usages? As shown in the CoNLL-2005 shared task (Carreras and Marquez, 2005), system performance numbers drop significantly when a classifier, trained on the Wall Street Journal (WSJ) corpus, is tested on the Brown corpus. This is largely due to the “highly ambiguous and unseen predicates (i.e. predicates that do not have training examples)” (Giuglea and Moschitti, 2006).

3 Construction Grammar

This issue of scalability and generalizability across genres could possibly be improved by linking semantics more directly with syntax, as theorized by Construction Grammar (CxG) (Fillmore et. al., 1988; Golberg, 1995; Kay, 2002; Michaelis, 2004; Goldberg, 2006). This theory suggests that the meaning of a sentence arises not only from the lexical items but also from the patterned structures or *constructions* they sit in. The meaning of a given phrase, a sentence, or an utterance, then, arises from the combination of lexical items and the syntactic structure in which they are found, including any patterned structural configurations (e.g. patterns of idiomatic expressions such as “The Xer, the Yer” – *The bigger, the better*) or recurring structural elements (e.g. function words such as determiners, particles, conjunctions, and prepositions). That is, instead of focusing solely on the semantic label of words, as is done in SRL and in many traditional theories in Linguistics, CxG brings more into focus the interplay of lexical items and syntactic forms or structural patterns as the source of meaning.

3.1 Application of Construction Grammar

Thus, rather than just assigning labels at the level of lexical items and predicate arguments as a way of piecing together the meaning of a sentence, we follow the central premise of CxG. Specifically, that semantics *can be* and *should be* interpreted at the level of the larger structural configuration.

Consider the following three sentences, each having the same syntactic structure, each taken

from different genres of writing available on the web.

Blogger arrested - blog him out of jail! [Blog]
Someone mind controlled me off the cliff. [Gaming]
He clocked the first pitch into center field. [Baseball]

Each of these sentences makes use of words, especially the verb, in ways particular to their genre. Even if we are unfamiliar with the specific jargon used, as a human we can infer the general meaning intended by each of the three sentences: *a person X causes an entity Y to move in the path specified by the prepositional phrase* (e.g. third sentence: “A player causes something to land in the center field.”).

In a similar way, if we can assign a meaning of caused motion at the sentence level and an automatic learner can be trained to accurately identify the construction, then even when presented with an unseen word, a useful semantic analysis is still possible.

3.2 Caused-Motion Construction

For this effort, we focused on the *caused-motion construction*, which can be defined as having the coarse-grained syntactic structure of Subject Noun Phrase followed by a verb that takes both a Noun Phrase Object and a Prepositional Phrase: (NP-SBJ (V NP PP)); and the semantic meaning ‘*the agent, NP-SBJ, directly causes the patient, NP, to move along the path specified by the PP*’ (Goldberg 1995). This construction is exemplified by the following sentences from (Goldberg 1995):

- (2) Frank sneezed the tissue off the table.
- (3) Mary urged Bill into the house.
- (4) Fred stuffed the papers in the envelope.
- (5) Sally threw a ball to him.

However, not all syntactic structures of the form (NP-SBJ (V NP PP)) belong to the caused-motion construction. Consider the following sentences.

- (6) I considered Ben as one of my brothers.
- (7) Jen took the highway into Pennsylvania.
- (8) We saw the bird in the shopping mall.
- (9) Mary kicked the ball to my relief.

In (6) and (9), the PPs do not specify a location, a direction or a path. In (8), the PP is a location;

however, the PP indicates the location in which the “seeing” event happened, not a path along which “we” caused “the bird” to move. Though the PP in (7) expresses a path, it is not a path in which Jen causes “the highway” to move.

3.3 Goals

As an initial step in determining the usefulness of construction grammar for interpreting semantics in computational linguistics, we present the results of our study aimed at ascertaining if a classifier can be taught to identify caused-motion constructions. We also report on our investigations into which features were most useful in the classification of caused-motion constructions.

4 Data & Experiments

The data for this study was pulled from the WSJ part of Penn Treebank II (Marcus et al., 1994). From this corpus, all sentences with the syntactic form (NP-SBJ (V NP PP)) were selected. The selection allowed for intervening adverbial phrases (e.g. “Sally threw a ball *quickly* to him”) and additional prepositional phrases (e.g. “Sally threw a ball to him *on Tuesday*” or “Sally threw a ball *in anger* into the scorer’s table”). A total of 14.7k instances³ were identified in this manner.

To reduce the size of the corpus to be labeled to a target of 1800 instances, we removed, firstly, instances containing traces as parsed by the TreeBank. These included passive usages (e.g. “Coffee was shipped from Colombia by Gracie”) and instances with traces in the object NP or PP including questions and relative clauses (e.g. “What did Gracie ship from Colombia?”). In construction grammar, however, traces do not exist, since grammar is a set of patterns of varying degrees of complexity. Thus CxG would characterize passives, questions structures, and relative clauses as having their own respective phrasal constructions, which combine with the caused-motion construction. In order to ensure sufficient training data with the standard form of the caused-motion construction as defined in Goldberg 1995 and 2006 (see Section 3.2), we

³ We use the term *instances* over *sentences* since a sentence can have more than one instance. For example, the sentence “I gave the ball to Bill, and he kicked it to the wall.” is composed of 2 instances.

chose to remove these usages.

Secondly, we removed the instances of sentences that can be deterministically categorized as non-caused motion constructions: instances containing ADV, EXT, PRD, VOC, or TMP type object NPs (e.g. “Cindy drove *five hours* from Dallas”, “You listen, *boy*, to what I say!”). Because we can automatically identify this category, keeping these examples in our data would have resulted in even higher performance.

We also considered the possibility of reducing the size by removing certain classes of verbs such as verbs of communication (e.g. *reply*, *bark*), psychological state (e.g. *amuse*, *admire*), or existence (e.g. *be*, *exist*). While it is reasonable to say that these verb types are highly unlikely to appear in a caused-motion construction, if we were to remove sets of verbs based on their likely behavior, we would also be excluding interesting usages such as “The stand-up comedian amused me into a state of total enjoyment.” or “The leader barked a command into a radio.”

After filtering these sentences, 8700 remained. From the remaining instances, we selected 1800 instances at random for the experiments presented.

4.1 Labels and Classifier

The 1800 instances were hand-labeled with one of the following two labels:

- Caused-Motion (CM)
- Non Caused-Motion (NON-CM)

The CM label included both literal usages (e.g. “Well-wishers stuck little ANC flags in their hair.”) and non-literal usages (e.g. “Producers shepherded ‘Flashdance’ through several scripts.”) of caused-motion.

After the annotation, the corpus was randomly divided into two sets: 75% for training data and 25% for testing data. The distribution of the labels in the test data is 33.3% CM and 66.7% NON-CM. The distribution in the training set is 31.8% CM and 68.2% NON-CM. For our experiments, we used a Support Vector Machine (SVM) classifier with a linear kernel. In particular we made use of the LIBSVM (Chang and Lin, 2001) as training and testing software.

4.2 Baseline Features

The baseline consisted of a single conceptual feature - the lemmatized, case-normalized verb. We chose the verb as a baseline feature because it is generally accepted to be the core lexical item in a sentence, which governs the syntactic structure and semantic constituents around it. This is especially evidenced in the Penn Treebank where NP nodes are assigned with syntactic labels according to the position in the tree relative to the verb (e.g. Subject). In VerbNet and PropBank, the semantic labels are assigned to the constituents around the verb, each according to its semantic relationship with the verb.

This verb feature was encoded as 478 binary features (one for each unique verb in the dataset), where the feature value corresponding to the instance's verb was 1 and all others were 0.

4.3 Additional Features

In the present experiments, we utilize gold-standard values for two of the *PP features* for a proof of feasibility. Future work will evaluate the effect of automatically extracting these features. In addition to the baseline verb feature (feature 1), our full feature set consisted of 8 additional types for a total of 334 features. Examples used in the feature descriptions are pulled from our data.

PP features:

2. Preposition (76 features) The preposition heading the prepositional phrase (e.g. “*Producers shepherded ‘Flashdance’* *[[through]_P several scripts]_{PP}*.”) was encoded as 76 binary features, one per preposition type in the training data. For instances with multiple PPs, preposition features were extracted from each of the PPs.
3. Function Tag on PP (11 features) Penn Treebank encodes grammatical, adverbial, and other related information on the PP's POS tag (e.g. “PP-LOC”). The function tag on the prepositional phrase was encoded as 10 binary features plus an extra feature for PPs without function tags. Again, for instances with multiple PPs, each corresponding function tag feature was set to 1.
4. Complement Category to P (19 features) Normally a PP node consists of a P and a NP.

However, there are some cases where the complement of the P can be of a different syntactic category (e.g. “*So, view permanent insurance* *[[for]_P [what it is]_{SBAR}]_{PP}*.”). Thus, the phrasal category tags (e.g. NP, SBAR) of the preposition's sister nodes were encoded as 19 binary features. For instances with multiple PPs, all sister nodes of the prepositions were collected.

VerbNet features: The following features were automatically extracted from VerbNet classes with frames matching the target syntactic structure, namely “NP V NP PP”.

5. VerbNet Classes (123 features) The verbs in the data were associated with one or more of the above VerbNet classes according to their membership. The VerbNet classes were then encoded as 122 binary features with one additional feature for verbs that were not found to be members of any of these classes. If a verb belongs to multiple matching classes, each corresponding feature was set.
6. VerbNet PP Type (27 features) VerbNet frames associate the PP with a description (e.g. “NP V NP PP.*location*”). The types were encoded as 26 binary features, plus an extra feature for PPs without a description. The features represented the union of all PP types (i.e. if a VerbNet class included multiple PPs, each of the corresponding features was assigned a value of 1). If a verb was associated with multiple VerbNet classes, the features were set according to the union over both the corresponding classes and their set of PP types.

Named Entity features: These features were automatically annotated using BBN's *Identifinder* (Bikel, 1999). The feature counts for the subject NP and object NP differ strictly due to what entities were represented in the data. For example, the entity type “DISEASE” was found in an object NP position but not in a subject NP.

7. NEs for Subject NP (23 features) The union of all named entities under the NP-SBJ node was encoded as 23 binary features.
8. NEs for Object NP (27 features) The union of all named entities under the object NP node was encoded as 27 binary features.
9. NEs for PP's Object (28 features) The union

of all named entities under the NP under the PP node was encoded as 28 binary features.

5 Results

For the baseline system, the model was built from the training data using a linear kernel and a cost parameter of $C=1$ (LIBSVM default value). When using the full feature set, the model was also built from the training data using a linear kernel, but the cost parameter was $C=0.5$, the best value from 10-fold cross validation on the training data.

In Table 1, we report the precision (P), recall (R), F_1 score, and accuracy (A) for identifying caused-motion constructions⁴.

Features	P%	R%	F	A%
Baseline* Set	78.0	52.0	0.624	79.1
Full Set	87.2	86.0	0.866	91.1

Table 1: System Performance (*verb feature baseline)

The results show that the addition of the features presented in section 4.3 resulted in a significant increase in both precision and recall, which in turn boosted the F score from 0.624 to 0.857, an increase of 0.233.

6 Feature Performance

In order to determine the usefulness of the individual features in the classification of caused-motion, we evaluated the features in two ways. In one (Table 2), we compared the performance of each of the features to a majority class baseline (i.e. 66.7% accuracy). A useful feature was expected to show an increase over this baseline with statistical significance. Significance of each feature’s performance was evaluated via a chi-squared test ($p<0.05$).

Our results show that the features 3, 1, 2 and 5 performed significantly better over the majority class baseline. The features 4, 7 and 8 were unable to distinguish between the caused-motion constructions and the non caused-motion usages.

⁴ As we can see in Table 1, the accuracy is higher than precision or recall. This is because precision and recall are calculated with regard to identifying caused-motion constructions, whereas accuracy is based on identifying both caused-motion and non-caused motion constructions. Since it’s easier to get better performance on the majority class (NON-CM), the overall accuracy is higher.

#	Included Feature	P%	R%	F	A%	
3	Preposition	82.4	65.3	0.729	83.8	*
1	Verb	78.0	52.0	0.624	79.1	*
2	Function Tag on PP	82.6	38.0	0.521	76.7	*
5	VerbNet Classes	73.5	33.3	0.459	73.8	*
6	VerbNet PP Type	59.6	33.3	0.427	70.2	
9	NEs for PP’s Object	71.4	6.7	0.122	68.0	
4	Comp. Cat. of P		0.0		66.7	
7	NEs for Subject NP		0.0		66.7	
8	NEs for Object NP		0.0		66.7	

Table 2: Effect of each feature on the performance in classification of the caused-motion construction, in the order of decreasing F-score. Features that performed statistically higher than the majority class baseline are marked with an * in the last column.

Their precision values could not be calculated due to the fact that these features resulted in zero positive (CM) classification.

In a second study, we evaluated the performance of the system when each feature was removed individually from the full set of features (Table 3). The removal of a useful feature was expected to show a statistically significant drop in performance compared to that of the full feature set. Significance in this performance degradation when compared against the full set of features was evaluated via chi-squared test ($p<0.05$). Here, features 3, 8 and 1, when removed, showed a statistically significant performance drop. The rest of the features were not shown to have a statistically significant effect on the performance.

Our results show that the preposition feature is the single most predictive feature and the feature that has the most significant effect in the full feature set. These results are encouraging: unlike the purely lexical features like the named entity features (6, 7, and 8) that are dependent on the particular expression used in the sentence,

#	Removed Feature	P%	R%	F	A%	
3	Preposition	76.9	73.3	0.751	83.8	*
8	NEs for Object NP	84.6	80.7	0.826	88.7	*
1	Verb	85.9	81.3	0.836	89.3	*
2	Function Tag on PP	85.2	84.7	0.849	90.0	
9	NEs for PP’s Object	87.5	84.0	0.857	90.7	
7	NEs for Subject NP	87.0	84.7	0.858	90.7	
5	VerbNet Classes	86.0	86.0	0.860	90.7	
4	Comp. Cat. of P	86.7	86.7	0.867	91.1	
6	VerbNet PP Type	87.8	86.0	0.869	91.3	

Table 3: System performance when the specified feature is removed from the full set of features, in the order of increasing F-score. Significant performance degradation, when compared against the full feature set performance (Table 1) was labeled with an * in the last column.

prepositions are function words. Like syntactic elements, these function words also contribute to the patterned structures of a construction as discussed in Section 3. Furthermore, unlike the semantics of features that are dependent on content words that are subject to lexical variability, prepositions are limited in their lexical variability, which make them good general features that scale well across different semantic domains.

In addition to the preposition feature, the verb feature was found to affect performance at a statistically significant level in both cases. Based on the numerous studies in the past that have shown the usefulness of the verb as a feature, this is not an unexpected result. Interestingly, our results seem to indicate interactions between features. This can be seen in two different instances. First, while feature 8 (NEs for Object NP) alone was not found to be a predictive feature, when removed, it resulted in a statistically significant drop in performance compared to that of the full feature set. The opposite effect can be seen with the VerbNet Classes feature. While it showed a statistically significant boost in performance when introduced into the system by itself, when dropped from the full feature set, the drop in the system performance was not found to be significant. This seems to indicate that NEs for Object NP and the VerbNet Classes features have strong interactions with one or more of the other features. We will continue investigating these interactions in future work.

7 Out-of-Vocabulary Verbs

Additionally, we separately examined the performance on the test set verbs that were not seen in the training data (i.e. out-of-vocabulary/OOV items). Just over a fifth of the instances (92 out of 450 constructions) in the test data had unseen verbs, with a total of 83 unique verb types. The results show that there was no decrease in the accuracy or F-score. In fact, there was a chance increase, not statistically significant, in a two-sample t-test ($t=1.13$; $p>0.2$).

We carried out the same feature studies for the OOV verbs, as detailed in section 6 (Tables 4 and 5). The performance in both of the studies reflected the results seen in Tables 2 and 3, with one expected exception. The verb feature was, of course, found to be of no value to the predictor.

#	Removed Feature	P%	R%	F	A%
3	Preposition	63	76	0.69	90
2	Function Tag on PP	83	80	0.82	82
6	VerbNet PP Type	84	84	0.84	67
5	VerbNet Classes	84	84	0.84	73
9	NEs for PP's Object	84	84	0.84	74
1	Verb		0		73
4	Comp. Cat. of P		0		73
7	NEs for Subject NP		0		73
8	NEs for Object NP		0		73

Table 4: Effect of each feature on the performance in classification of the caused-motion construction with OOV verbs, in the order of decreasing F-score. The precision values could not be calculated for the performance of the features 1,4,7, and 8 due to the fact that these features resulted in zero positive classifications.

What is interesting here is that the verb feature did perform at a significant level for the full test data. By this observation, it would be expected that the overall performance on the OOV verbs would be negatively affected since there is no available verb information. However, this was not the case.

8 Discussion and Conclusion

The results presented show that a classifier can be trained to automatically identify the semantics of constructions; at least for the caused-motion construction, and that it can do this with high accuracy. Furthermore, we have determined that the preposition feature is the most useful feature when identifying caused-motion constructions. Moreover, in considering our results in light of the performance of the SRL systems (Gildea and Jurafsky, 2002; Carreras and Marquez, 2005), where unseen predicates result in significant performance degradation, we found in contrast that using CxG to inform semantics resulted in equally high performance on the out-of-vocabulary predicates. This serves as evidence that semantic

#	Removed Feature	P%	R%	F	A%
3	Preposition	63	76	0.69	82
8	NEs for Object NP	83	80	0.82	90
2	Function Tag on PP	84	84	0.84	91
5	VerbNet Classes	84	84	0.84	91
7	NEs for Subject NP	84	84	0.84	91
1	Verb	88	88	0.88	93
4	Comp. Cat. of P	88	88	0.88	93
6	VerbNet PP Type	92	88	0.90	95
9	NEs for PP's Object	92	88	0.90	95

Table 5: System performance when the specified feature is removed from the full set of features in the classification of constructions with OOV items, in the order of increasing F-score.

analysis of novel lexical combinations and unseen verbs can be improved by enriching semantics with a construction-level analysis.

9 Future Work

There are several directions to go from here. First, in this paper we have kept our study within the scope of caused-motion constructions. We intend to introduce more types of constructions and include more syntactic variation in our data. We will also add more annotated instances. Secondly, we examine the impact of the introduction of additional features, such as a bag-of-words feature. In particular, we will include semantic features based on FrameNet to the VerbNet semantic features we are already using. This will be more feasible once the SemLink semantic role labeler for FrameNet becomes available (Palmer, 2009). Finally, we plan to include a more detailed analysis of the feature interactions, and examine the benefit that a construction grammar perspective might add to our semantic analysis.

Acknowledgements

We gratefully acknowledge the support of the Defense Advanced Research Projects Agency (DARPA/IPTO) under the GALE program, DARPA/CMO Contract No. HR0011-06-C-0022, subcontract from BBN, Inc. We are also grateful to Laura Michaelis for helpful discussions and comments.

References

Agirre, Eneko and Philip Edmonds. 2006. Introduction. In *Word Sense Disambiguation: Algorithms and Applications*, Agirre and Edmonds (eds.), Springer.

Ben-David, Shai, Blitzer, John, Crammer, Koby Pereira, Fernando. 2006. 'Analysis of representations for domain adaptation', in *NIPS*.

Bikel, D., Schwartz, R., Weischedel, R. 1999. An algorithm that learns what's in a name. *Machine Learning: Special Issue on NL Learning*, 34, 1-3.

Carreras, Xavier and Lluís Marquez. 2005. Introduction to the CoNLL- 2005 shared task: Semantic role labeling. *Procs of CoNLL- 2005*.

Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

Gildea, Daniel and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28:3, 245-288.

Fillmore, Charles J., Christopher R. Johnson and Miriam R.L. Petruck (2003) Background to Framenet, *International Journal of Lexicography*, Vol 16.3: 235-250.

Fillmore, Charles, Paul Kay and Catherine O'Connor (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of *let alone*. *Language* 64: 501-38.

Giuglea, Ana-Maria and Alessandro Moschitti. 2006. Shallow semantic parsing based on FrameNet, VerbNet and PropBank. In *Proceedings of the 17th European Conference on Artificial Intelligence*, Riva del Garda, Italy.

Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press

Goldberg, Adele. E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Hovy, Edward H., Mitch Marcus, Martha Palmer, Sameer Pradhan, Lance Ramshaw, and Ralph M. Weischedel. 2006. OntoNotes: The 90% Solution. Short paper. *Proceedings of the Human Language Technology / North American Association of Computational Linguistics conference (HLT-NAACL 2006)*. pp. 57-60, New York, NY.

Kay, Paul. 2002. English Subjectless Tag Sentences. *Language* 78: 453-81.

Kipper-Schuler, Karin. 2005. *VerbNet: A broad coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press, Chicago, IL.

Michaelis, Laura A. (2004). *Type Shifting in Construction Grammar: An Integrated Approach to Aspectual Coercion*. *Cognitive Linguistics* 15: 1-67.

Ng, Hwee Tou and Hian Beng Lee. 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, California, 40-47.

Marcus, Mitchell P, Santorini, Beatrice, Marcinkiewicz, Mary A. (1994) "Building a large annotated corpus of English: the Penn Treebank" *Computational Linguistics* 19: 313-330.

Palmer, Martha. "Semlink: Linking PropBank, VerbNet and FrameNet." *Proceedings of the Generative Lexicon Conference*. Sept. 2009, Pisa, Italy: GenLex-09, 2009.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. *The Proposition Bank: An Annotated Corpus of Semantic Roles*. *Computational Linguistics*, 31(1):71-106.