Searching Semantic Resources for Complex Selectional Restrictions to Support Verb Sense Disambiguation

Merwyn Taylor
The MITRE Corporation
McLean, VA USA
mgtaylor@mitre.org

Lynn Carlson, Stephanie Poisson,
Sun Fontaine, Ethan Cooper
Department of Defense
Ft. George G Meade, MD
USA
{} @tycho.ncsc.mil

***Abstract*— Natural language processing systems are increasingly integrating lexicons with ontologies for word sense disambiguation (WSD). Manually acquiring a lexicon that is integrated with a large ontology and other semantic resources can be difficult and inefficient in part due to the complexity of ontologies and inconsistency of entity extractors supporting WSD applications. A major contributing factor to the difficulty is the creation of selectional restrictions with respect to particular semantic resources. This paper presents a process for acquiring complex expressions for selectional restrictions via search through an ontology.**

I. Introduction

Supervised learning of lexicons for Verb Sense Disambiguation (VSD) is an active are of research in Natural Language Processing (NLP). Supervised learning is used in the semi-automated acquisition of verb lexicons to support automated information extraction. An individual entry in a lexicon expresses the meaning & structure of a verb sense via constraining the interpretations of verb arguments to concepts in an ontology. The constraints are commonly referred to as selectional restrictions [1].

We are developing METEOR event extraction system which implements a theory of VSD based on complex selectional restrictions of semantic parses. METEOR's VSD theory relies on entity extraction of nouns based on semantic resources such as ontologies and semantic networks mapped to ontologies. As a result, lexicographers creating lexicons for METEOR have to familiarize themselves with the contents of the semantic resource. Supervised learning is used to reduce the burden of navigating complex semantic resources.

The task of the lexicographer is to determine which concepts in the ontology best characterizes the types of entities that determine the senses of verbs included in the lexicon. The grammar of the METEOR lexicon allows lexicographers to create complex expressions involving disjunction, conjunction, and negation of ontological concepts subsuming the entity types extracted as argument fillers for the verbs of interest. For example, a "ConvergenceEvent" can be expressed as the following selectional restrictions on the verb "meet":
{meet.subject} → (Physical – CognitiveAgent) &
{meet.object} → (Physical – CongnitiveAgent).
This expression states that the filler of the subject(object) of a meet verb has to have an ontological interpretation that is subsumed by Physical but not subsumed by CognitiveAgent as defined in SUMO [2].

Lexicographers manually search semantic resources to identify concepts for selectional restrictions that represent terms in collections of sentences. The manual process is timing consuming and is not scalable. We use supervised learning to assist the manual lexicon acquisition task.

The benefits of supervised learning of lexicons to lexicographers include:

- reduced lexicon acquisition time
- efficient ontological traversal
- efficient lexicon update if the ontology is replaced/updated
- provide suggestions to existing lexicons

We are particularly interested in developing a supervised learning system that induces inclusionary selectional restrictions to cover the positive training examples of selected verb senses and exclusionary selectional restrictions to dismiss negative examples. In practice we believe that the system will be a tool for the lexicographer to seed a verb lexicon because the meaning of some senses may not be sufficiently conveyed via the training examples.

II. Related Research

The use of supervised learning for word sense disambiguation is an active area of research. Most of the work is based on the use of WordNet [3] as a semantic resource. Resnik discussed a probabilistic model that captures the co-occurrence behavior of predicates and conceptual classes in a taxonomy for noun sense disambiguation [4]. Ye presented an approach based on a semantic parse and demonstrated the use of arguments other than subjects and objects [5]. The work presented in this article adopts an approach similar to [5]. In [6], Scheffczyk discusses an approach to link FrameElements in FrameNet to SUMO. The approach discussed in this paper differs from (6) in that we search for more complex expressions that include classes to exclude. Dligach proposed an approach to supervised learning for WSD based on the lexical, syntactic, and semantic features [7]. Dligach demonstrated the utility of words surrounding the target verb, POS tags, and the path through parse tree connecting the target verb to its arguments.

III. Learning Algorithm

A. Basic algorithm

The METEOR system uses a complex lexicon consisting of selectional restrictions with references to an ontology, cardinality constraints, and argument quantifiers that influence the semantics of the selectional restrictions. The goal of this research is to automatically learn lexicons that are consistent with the aforementioned features. The learning algorithm has to accomplish the following:

- Find a minimal set of ontological concepts that subsume terms in positive examples
- Find a minimal set of ontological concepts that subsume terms in negative examples but not terms in the positive examples
- Approximate the saliency of arguments

Before describing the lexicon learning algorithm, we introduce the following notations:

- $A = \{A_{sub}, A_{obj}, A_{with}, \ldots\}$ is set of attribute names denoting attributes that we expect from the parser. A_{sns} is the special attribute denoting the sense of a training example. A_{cls} denotes the set of senses to which an example has been classified.
- $D = \{I_1, I_2, \ldots I_n\}$ denotes the set of training examples. I_i is a vector containing values for all attributes of A, A_{sns}, and A_{cls}. $I_i.A_a = T$ denotes that T is the entity type for the filler of argument A_a in training example I_i or T is the filler if the entity type is unknown.
- $V = \{V_1, V_2, \ldots V_n\}$ denotes the set of senses for a particular verb. V contains the set of senses that we want to automatically learn.
- $V_i.A_a.range = \{IC1, IC2,..\} - \{EC1,EC2,..\}$ denotes an ontological range constraint for sense V_i on argument A_a. This is a selectional restriction. IC_i denotes inclusionary constraints while EC_i denotes exclusionary constraints.
- Vi.Aa.quantifier = {„C+’|’C-„|’S’} is an annotation denoting the saliency of a selectional restriction. „C+’ denotes that the argument is required and the fillers must satisfy the range constraint. „C-„ denotes that the argument is restricted such that its fillers do not satisfy the range constraint. „S’ denotes that the argument is optional.
- Dist(T,C) denotes the semantic distance from T to C. If T is a term then Dist(T,C) is the longest distance from all interpretations of T to C based on the WordNet2SUMO [8] mappings. If T is a concept then Dist(T,C) is the longest distance from T to C.

The algorithm has to learn a selectional restriction for verb sense V_i and argument A_a in the form

$$V_i.A_a.range = \{IC_i\} - \{EC_j\}$$

as described above. For example the selectional restriction for an argument can be expressed as

$$range = \textbf{Physical} - \textbf{CognitiveAgent} \quad \textbf{(1)}$$

which includes all concepts subsumed by Physical except those subsumed by CognitiveAgent. An alternative selectional restriction would be

$$\mathbf{range} = \{\mathbf{C}|\mathbf{C} \rightarrow \mathbf{Physical}\ \&\ \neg\mathbf{CognitiveAgent} \rightarrow \mathbf{C}\} \cup \{\mathbf{C}|\mathbf{C\ is\ a\ sibling\ of\ CognitiveAgent}\} \quad \mathbf{(2)}$$

This expression is logically equivalent to (1) but it is not a robust expression. If the ontology is updated to include new descendants of Physical, (2), and all similar equations would have to be updated. For this reason we seek concise and robust expressions.

The algorithm also has to assign quantifiers to all selectional restrictions in form

$V_i.A_a$.quantifier = {„C+'|'C-„|S}. These quantifiers influence the semantics of the selectional restriction and are thus vital to the learning process.

The basic algorithm is listed in Figure 1.

The IncludeConcpets() function searches for a set of concepts in the ontology that subsume the distinct values for all arguments partitioned by sense. The function does not adjust for negative examples because the lexicon grammar allows two degrees of freedom to exclude concepts and concepts are excluded in the ExcludeConcepts function.

The criteria for selecting a concept C_a from CT_a in the IncludeConcepts function is based on three factors:

- SemDist: average semantic distance from values in $I.A_a$ to C_a
- DistVals: proportion of distinct values in $I.A_a$ that C_a subsumes
- RelFrq: the probability that C_a subsumes values in $I.A_a$

A subsumption score is calculated for every C_a in CT_a. The subsumption score equation is

$$\frac{\mathbf{1-SemDist + DistVals*0.5 + RelFrq*0.5}}{\mathbf{3}}\,. \quad \mathbf{(3)}$$

$$SemDist = \frac{\mathbf{log2}(\sum_{\mathbf{t} \in \mathbf{DV_a}} \mathbf{Dist(t,C_a)})}{\mathbf{log2}(\sum_{\mathbf{t} \in \mathbf{DV_a}} \mathbf{Dist(t,Entity)})} \quad \mathbf{(4)}$$

$$DistVals = \frac{|\{\mathbf{t}|\mathbf{t}\in\mathbf{DV_a}\ \&\ t \uparrow \mathbf{C_a}\}|}{|\{\mathbf{t}|\mathbf{t}\in\mathbf{DV_a}\ \&\ t \uparrow \mathbf{Entity}\}} \quad \mathbf{(5)}$$

```
IncludeConcepts()
foreach subsense Vi in V
  foreach argument Aa in A
     1. collect distinct values DVa from I.Aa for
        all I in D where I.Asns = Vi
     2. build coverage tree CTa for DVa
     3. select concepts from CTa that subsume
        all values in DVa .
     4. Vi.Aa.range = CTa
     5. compute Vi.Aa.quantifier

ExcludeConcepts()
foreach subsense Vi
   compute error for Vi
   foreach argument Aa in A
     1. collect distinct values DVa from I.Aa for
        all I in D where I.Asns = Li
     2. collect distinct values DCVa from I.Aa
        for all I in D where I is incorrectly
        labeled as Vi
     3. build coverage tree CCTa from DCVa
     4. select concepts from CCTa that subsume
        values in DCVa but do not subsume
        values in DVa
     5. Vi.Aa.range = Vi.Aa.range – CCTa
```

Figure 1: Selectional Restriction Algorithm

$$RelFrq = \frac{|\{\mathbf{I}|\mathbf{I}\in\mathbf{D}\ \&\ \mathbf{V_i} \in \mathbf{D.I.A_{sns}}\ \&\ D.I.\mathbf{A_a} \uparrow \mathbf{C_a}\}|}{|\{\mathbf{I}|\mathbf{I}\in\mathbf{D}\ \&\ V_\mathbf{i} \in \mathbf{D.I.A_{sns}}\ \&\ D.I.\mathbf{A_a} \uparrow \mathbf{Entity}\}|} \quad \mathbf{(6)}$$

SemDist measures the generality of concepts with respect to the values that they subsume in the training data. Concepts that are semantically closer to the instance data are preferred. The Dist(t,C) function uses the longest path from term t to concept C to determine the distance between t and C. Zhong reported the use of the longest path to measuring semantic distance that performed well [9]. DistVals measures the total number of distinct values subsumed by C_a. Higher values are preferred since a single concept may subsume many examples. RelFrq measures the percentage of training examples containing values for argument A_a that are subsumed by C_a. Higher values are preferred because viewer concepts will be required to model the training examples. DistVals and RelFrq are weighted with 0.5 because in practice the average semantic distance, SemDist, tended to be more significant with respect to selecting the desired concepts for selectional restrictions. Other weights for DistVals and RelFrq yielded concepts that were too specific and resulted in poor METEOR performance.

Searching semantic resources is often based on semantic distance measures. Onyshkevych introduced a distance measure based on weighted properties, along the path connecting concepts, to determine the semantic distance [10]. Zuber presented a similarity measure derived from the number of descendants for each concept and common features [11].

In this research, we focus on finding suitable expressions for selectional restrictions. In many cases, the concept that best minimizes the semantic distance between terms is not always the best concept for an expression. For this reason, we use a simple edge-counting approximation to semantic distance.

Step 3 in IncludeConcepts is performed iteratively until all values in DVa are covered by the selectional restrictions. This searches for concepts in the ontology that cover values that cluster well together as opposed to searching for a single concept that covers all values as in [12].

The ExcludeConcepts function attempts to find a set of concepts in the ontology that subsume values from negative examples while preserving values in the positive examples. This phase restricts existing C+ selectional restrictions or creates new C- selectional restrictions. Both approaches exploit the expressivity of METEOR's lexicon grammar.

An existing selectional restriction with a C+ quantifier is refined by searching for concepts in the ontology that subsume values from the negative examples while not subsuming values in the positive examples. To achieve this goal we use (3) as well as entropy measures. We select concepts with low entropies and are predictive of values stored in the negative examples.

Figure 2 illustrates the problem of generating the selectional restriction for the subject of a

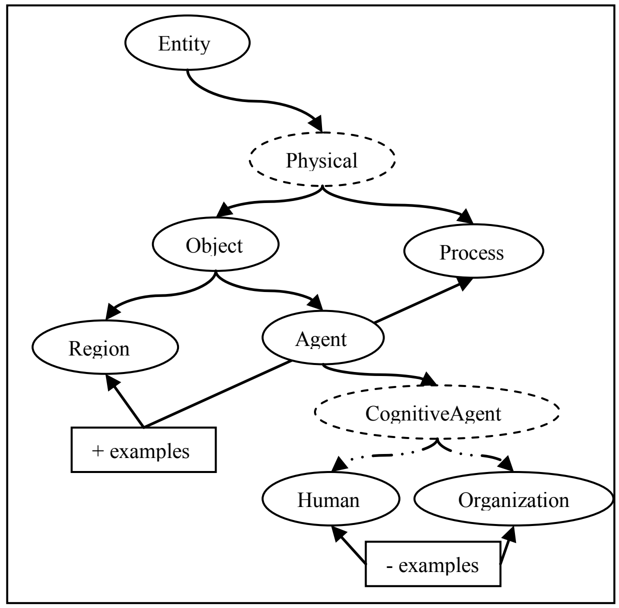

Figure 2 Distribution of positive and negative examples leading to the expression Physical – CognitiveAgent.

"ConvergenceEvent" involving physical objects. The desired expression is

$$V_i.A_{sub}.\text{range} = \text{Physical} - \text{CognitiveAgent}.$$

The positive training examples are expressions having subjects which are instances of Region, Process, etc. but not CognitiveAgent. The negative examples have subjects which are instances of Human and Organization. The positive examples are generalized to Physical while the negative examples are generalized to CognitiveAgent. In this example, Physical is selected because it is the concept that has the maximum value for (1) among the positive examples and CognitiveAgent was selected because it had the maximum value for (1) among the negative examples.

New selectional restrictions with C- quantifiers are created by identifying arguments that are prevalent in the negative examples. Let A_a denote an argument that has fillers in negative examples but is not used in selectional restrictions for a given verb sense V_i. A new selectional restriction,

$$V_i.A_a.\text{range} = \{CCT_{a1}, CCT_{a2}\ldots\},$$

is created by identifying concepts CCT_{ai} in the ontology where entropy(CCT_{ai}) < 0.01 and maximizes (3). Concepts with low entropies are preferred because they tend to partition the training examples better.

B. Argument Quantifier

In addition to selecting range constraints for selectional restrictions, the algorithm also attempts to quantify selection restrictions with semantics that influence the interpretations of the selectional restrictions. The algorithm currently generates quantifiers in the set {C+, C-, S}. A selectional restriction $V_i.A_a$ is annotated with 'S' if the argument A_a is used in less than 74% of the training examples where $D.I.A_{sns} = V_i$. A selectional restriction $V_i.A_a$ is annotated with 'C+' if the argument A_a is used in more 74% of the training examples where $D.I.A_{sns} = V_i$. A selectional restriction $V_i.A_a$ is annotated with „C-„ if it is created as described above.

IV. Performance

The lexicon learning system was applied to a corpus of sentences used to manually create lexicons for METEOR. We compared the automatically generated lexicon to the manual created lexicon to determine the amount of labor that would be required to generate the final lexicon for a collection of verbs. We measured the differences in recall and precision and used these measurements to gauge the amount of labor required to finalize a lexical items.

Table 1 contains a comparison of the recall and precision between the lexicon that was automatically generated and the manually created lexicon for a subset of the verbs. Some expressions were classified as multiple senses. In these cases all erroneous senses are counted individually. The difference in the average recall was +0.09 and the difference in the average precision was +0.09.

Verb	Recall	Precision
join	+0.14	+0.25
enter	+0.03	-0.08
tell	-0.01	-0.30
travel	+0.10	+0.10
purchase	0	0
meet	-0.08	+0.33

Table 1 Lexicon learner performance compared to manually acquired lexicon

There were fewer misclassifications reported when the automatically generated lexicon was applied to the training corpus. The lexicon learner identified exclusionary ontological restrictions that lexicographers can use when finalizing lexical items. These exclusionary restrictions were in the form of „C-„ selectional restrictions for lexical items as well as concepts to exclude in „C+' selectional restrictions. The results indicate that the automated approach can be used to reduce the burden of manually navigating semantic resources in an attempt to find concepts that are appropriate for selectional restrictions for a given sense of a verb.

To illustrate the utility of automated semantic search we discuss a lexical entry for the verb "enter". The automatically generated entry included restrictions on collections of prepositions that were not included in the manually created entry. Consequently, the automatically generated entry had a higher precision. These additional expressions could be manually added to the final lexicon at very little cost.

An unintended consequence of this research was the detection of a sense substructure that was hidden in example training data. The substructure was manually identified as the result of an unusual selectional restriction. The unusual selectional restriction was created to accommodate training data that were structurally different from examples in the sense to which they belonged. The offending examples were re-labeled and a new substructure was added to the lexical item.

V. Problems

A. Word Sense Disambiguation of Nouns

When multiple interpretations for a term exist, the algorithm tends to select the interpretation that covers the highest number instances in other relevant training examples. If the coverage statistics for the training examples are not sufficient for selecting an interpretation, the algorithm may randomly select an inappropriate interpretation that does not sufficiently characterize the sense conveyed by a verb sense. Selecting all interpretations actually degrades the accuracy and thus the performance of the lexicon.

B. Selectional Restriction Size

It is desirable to create selectional restrictions that are as concise as possible. This requires the right balance of concept generality and the number of concepts. More specific concepts yield lexicons that are too fined grained and contain a higher amount of inclusionary concepts in selectional restrictions. More general concepts yield fewer inclusionary concepts but require more exclusionary concepts. The system currently doesn't consistently generate concise selectional restrictions. It sometimes generates large disjuncts of specific concepts which could be reduced to fewer disjuncts of general concepts with a few exclusions.

VI. Future Work

METEOR's lexicon grammar allows a lexicographer to create a disjunction over 2 or more selectional restriction where the semantics of the disjunction are that at least one of the selectional restrictions has to be satisfied. An example is a sense/structure of the travel verb which has to have a source or a destination but both are not required in an expression. The learning algorithm currently does not group selectional restrictions in this manner. This is a very expressive feature of METEOR's lexicon that we wish to automatically learn in the future.

The system tends of create too many C- selectional restrictions. This over generation of C- selectional restrictions negatively impacts the applicability of a lexicon. We want to research techniques for identifying the minimal amount of such selectional restrictions based principle component analysis.

VII. Bibliography

[1] Framis, Francesc Ribas, "On learning more appropriate selectional restrictions," 1995. Proceedings of 7th Conference of the European Chapter of the Association for Computational Linguistics.

[2] Niles, I and Pease, Adam, "Towards a standard upper ontology," [ed.] Chris Welty and Barry Smith. Ogunquit, Maine : s.n., 2001. Proceedings of the 2nd Internaltional Conference on Formal Ontology in Information Systems (FOIS-2001).

[3] Fellbaum, Christiane. "WordNet: An electronic lexical database," s.l. : MIT Press, 1998.

[4] Resnik, Phillip, "Selectional preferences and sense disambiguation." Washington DC : s.n., 1997. Proceedings of ACL SIGLEX Workdhop on Tagging Text with Lexical Semantics, Why, What and How?

[5] Ye, Patrick and Baldwin, Timothy, "Verb sense disambiguation using selectional preferences extracted with a state-of-the-art semantic role labeler," Sydnet, Australia : s.n., 2006. Proceedings of the Austrlasian Language Technology Workshop. pp. 141-148.

[6] Scheffcyzk, Jan, Pease, Adam and Ellsworth, Michael, "Linking FrameNet to the Suggested Upper Merged Ontology," Baltimore, Maryland : s.n., 2006. Proceedings of the 2006 International Conference on Formal Ontology in Information Systems (FOIS 2006).

[7] Dligach, Dmitriy and Palmer, Martha, "Novel semantic features for verb sense disambiguation," Columbus, Ohio : s.n., 2008. Proceedings of Association for Computational Linguistics (ACL-08). pp. 29-32.

[8] Niles, I and Pease, Adam, "Linking lexicons and pntologies: Mapping WordNet to the Suggested Upper Merged Ontology," Las Vegas, Nevada : s.n., 2003. Proceedings of the 2003 International Conference on Information and Knowledge Engineering.

[9] Zhong, Jiwei, "Conceptual graph matching for semantic search," London, UK : s.n., 2002. Proceedings of the 10th International Conference on Conceptual Structures. pp. 92-96.

[10] Onyshkevych, Boyan, "OntoSearch: Using an ontology as a search space for knowledge based text processing," Carnegie Mellon University. 1997. PhD Thesis.

[11] Schicket-Zuber, Vincent, "OSS: A semantic similarity function based on hierarchical ontologies," 2007. Proceedings of the 2007 International Joint Conference on Artificial Intelligence IJCAI'07.

[12] Resnik, Philip, "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural langauge," 1999, Journal of Artificial Intelligence Research, Vol. 11, pp. 95-130.