OntoNotes: The 90% Solution

Abstract

We describe the OntoNotes methodology and its result, a large multilingual richly-annotated corpus constructed at 90% interannotator agreement. An initial 300K portion will be made available to the community during 2007.

1 Introduction

Many natural language processing applications could benefit from a richer model of text meaning than the bag-of-words and n-gram models that currently predominate. Until now, however, no such model has been identified that can be annotated dependably and rapidly. We have developed a methodology for producing such a corpus at 90% inter-annotator agreement, and will release completed segments beginning in early 2007.

The OntoNotes project focuses on a domain independent representation of literal meaning that includes predicate structure, word sense, ontology linking, and coreference. Pilot studies have shown that these can all be annotated rapidly and with better than 90% consistency. Once a substantial and accurate training corpus is available, trained algorithms can be developed to predict these structures in new documents.

This process begins with parse (TreeBank) and propositional (PropBank) structures, which provide normalization over predicates and their arguments. Word sense ambiguities are then resolved, with each word sense also linked to the appropriate node in the Omega ontology. Coreference is also annotated, allowing the entity mentions that are propositional arguments to be resolved in context.

Annotation will cover multiple genres to create a resource that is broadly applicable, and multiple languages, since semantic mapping between languages should benefit translation performance.

2 Treebanking

The Penn Treebank (Marcus *et al.*, 1993) is annotated with information to make predicate-argument structure easy to decode, including function tags and markers of "empty" categories that represent displaced constituents. To expedite later stages of annotation, we have developed a parsing system (<anonymous>, 2006) that recovers both of these latter annotations, the first we know of. A first-stage parser matches the Collins (2003) parser on which it is based on the Parseval metric, while simultaneously achieving near state-of-the-art performance on recovering function tags (F-measure 89.0). A second stage, a seven stage pipeline of maximum entropy learners and voted perceptrons, achieves state-of-the-art performance (F-measure 74.7) on the recovery of empty categories by combining a linguistically-informed architecture and a rich feature set with the power of modern machine learning methods.

3 PropBanking

The Penn Proposition Bank, funded by ACE (DOD), focuses on the argument structure of verbs, and provides a corpus annotated with semantic roles, including participants traditionally viewed as arguments and adjuncts. The 1M word Penn Treebank II Wall Street Journal corpus has been suc-

cessfully annotated with semantic argument structures for verbs and is now available via the Penn Linguistic Data Consortium as PropBank I (Palmer *et al.*, 2005). Links from the argument labels in the Frames Files to FrameNet frame elements and VerbNet thematic roles are being added. This style of annotation has also been successfully applied to other genres and languages.

4   Word Sense

Word sense ambiguity is a continuing major obstacle to accurate information extraction, summarization and machine translation. The subtle fine-grained sense distinctions in WordNet have not lent themselves to high agreement between human annotators or high automatic tagging performance. Building on results in grouping fine-grained WordNet senses into more coarse-grained senses that led to improved inter-annotator agreement (ITA) and system performance (Palmer *et al.,* 2004; Palmer *et al.,* 2006), we have developed a process for rapid sense inventory creation and annotation that includes critical links between the grouped word senses and the Omega ontology.

This process is based on recognizing that sense distinctions can be represented by linguists in an hierarchical structure, similar to a decision tree, that is rooted in very coarse-grained distinctions which become increasingly fine-grained until reaching WordNet senses at the leaves. Sets of senses under specific nodes of the tree are grouped together into single entries, along with the syntactic and semantic criteria for their groupings, to be presented to the annotators.

As shown in Figure 1, a 50-sentence sample of

instances is annotated and immediately checked for inter-annotator agreement. ITA scores below 90% lead to a revision and clarification of the groupings by the linguist. It is only after the groupings have passed the ITA hurdle that each individual group is linked to a conceptual node in the ontology. In addition to higher accuracy, we find at least a three-fold increase in annotator productivity.
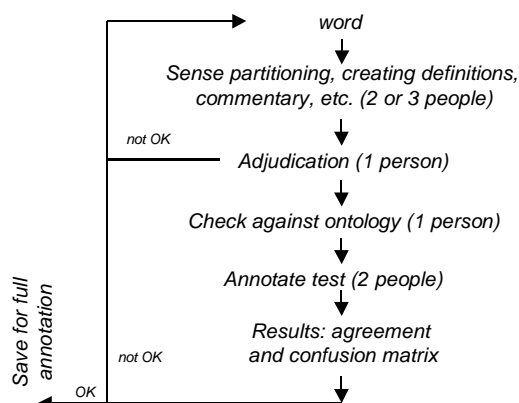

Figure 1. Annotation Procedure

As part of OntoNotes we are annotating the most frequent noun and verb senses in a 300K subset of the PropBank, and will have this data available for release in early 2007.

4.1   Verbs

Our initial goal is to annotate the 700 most frequently occurring verbs in our data; so far 300 verbs have been grouped and 150 double annotated. Subcategorization frames and semantic classes of arguments play major roles in determining the groupings, as illustrated by the grouping for

GI: operating or traveling via a vehicle *NP (Agent) drive NP, NP drive PP*	WN1: "Can you drive a truck?", WN2: "drive to school,", WN3: "drive her to school,", WN12: "this truck drives well," WN13: "he drives a taxi,",WN14: "The car drove around the corner,", WN:16: "drive the turnpike to work,"
G2: force to a position or stance *NP drive NP/PP/infinitival*	WN4: "He drives me mad.," WN6: "drive back the invaders," WN7: "She finally drove him to change jobs," WN8: "drive a nail," WN15: "drive the herd," WN22: "drive the game."
G3: to exert energy on behalf of something *NP drive NP/infinitival*	WN5: "Her passion drives her," WN10: "He is driving away at his thesis."
G4: cause object to move rapidly by striking it *NP drive NP*	WN9: "drive the ball into the outfield ," WN17 "drive a golf ball," WN18 "drive a ball"
G5: a directed course of conversation	WN11: "What are you driving at?"
G6: excavate horizontally, as in mining	WN19: "drive a tunnel through the mountain"
G7: cause to function or operate	WN20: "steam drives the engine"

Figure 2. Grouping of WordNet Senses for "drive"

the 22 WN 2.1 senses for *drive* in Figure 2. In addition to improved annotator productivity and accuracy, we predict a corresponding improvement in system performance. Training on this new data, Chen and Palmer (2005) report 80.7% accuracy for verbs using a smoothed maximum entropy model and rich linguistic features. They also report state-of-the art performance on fine-grained senses, but the results are more than 10% lower.

4.2 Nouns

We follow a similar procedure for the annotation of nouns. The same individual who groups Word-Net verb senses also creates noun senses, starting with WordNet and other dictionaries. We aim to double-annotate the 1100 most frequent polysemous nouns in the initial corpus by the end of 2006, while maximizing overlap with the sentences containing annotated verbs.

Certain nouns carry predicate structure; these include nominalizations (whose structure obviously is derived from their verbal form) and various types of relational nouns (like *father*, *President*, and *believer*, that express relations between entities, often stated using *of*). We have identified a limited set of these whose structural relations can be semi-automatically annotated with high accuracy.

5 Ontology

In standard dictionaries, the senses for each word simply listed. In order to allow access to additional useful information, such as subsumption, property inheritance, predicate frames from other sources, links to instances, and so on, our goal is to link the senses to an ontology. This requires decomposing the hierarchical structure into subtrees which can then be inserted at the appropriate conceptual node in the ontology.

The OntoNotes terms are represented in the 110,000-node Omega ontology (Philpot *et al.*, 2005), under continued construction and extension at ISI. Omega has been assembled semi-automatically by merging a variety of sources, including Princeton's WordNet, New Mexico State University's Mikrokosmos, and a variety of Upper Models, including DOLCE (Gangemi et al., 2002), SUMO (Niles and Pease, 2001), and ISI's Upper Model, which are in the process of being reconciled. The verb frames from PropBank, FrameNet,

WordNet, and Lexical Conceptual Structures (Dorr and Habash, 2001) have all been included and cross-linked.

In work planned for later this year, verb and noun sense groupings will be manually inserted into Omega, replacing the current (primarily WordNet-derived) contents. For example, of the six verb groups G1–G6 for *drive* in the table above, G1 and G4 will be placed into the area of "controlled motion," while G2 will go with "attitudes."

6 Coreference

The coreference annotation in OntoNotes connects coreferring instances of specific referring expressions, meaning primarily NPs that introduce or access a discourse entity. For example, "Elco Industries, Inc.", "the Rockford, Ill. Maker of fasteners", and "it" could all corefer. (Non-specific references like "officials" in "Later, officials reported…" are not included, since coreference for them is frequently unclear.) In addition, proper premodifiers and verb phrases can be marked when coreferent with an NP, such as linking, "when the company withdrew from the bidding" to "the withdrawal of New England Electric".

Unlike the coreference task as defined in the ACE program, attributives are not generally marked. For example, the "veterinarian" NP would not be marked in "Baxter Black is a large animal veterinarian". Adjectival modifiers like "American" in "the American embassy" are also not subject to coreference.

Appositives are annotated as a special kind of coreference, so that later processing will be able to supply and interpret the implicit copula link.

All of the coreference annotation is being doubly annotated and adjudicated. In our initial English batch, the average agreement scores between each annotator and the adjudicated results were 91.8% for normal coreference and 94.2% for appositives.

7 Related and Future Work

PropBank I (Palmer *et al*., 2005), developed at UPenn, captures predicate argument structure for verbs; NomBank provides predicate argument structure for nominalizations and other noun predicates (Meyers *et al*., 2004). PropBank II annota-

tion (eventuality ID's, coarse-grained sense tags, nominal coreference and selected discourse connectives) is being applied to a small (100K) parallel Chinese/English corpus (Babko-Malaya *et al.*, 2004). The OntoNotes representation extends these annotations, and allows eventual inclusion of additional shallow semantic representations for other phenomena, including temporal and spatial relations, numerical expressions, deixis, etc. One of the principal aims of OntoNotes is to enable automated semantic analysis. The best current algorithm for semantic role labeling for PropBank style annotation (Pradhan *et al.*, 2005) achieves an F-measure of 81.0 using an SVM. OntoNotes will provide a large amount of new training data for similar efforts.

Existing work in the same realm falls into two classes: the development of resources for specific phenomena or the annotation of corpora. An example of the former is Berkeley's FrameNet project (Baker *et al.*, 1998), which produces rich semantic frames, annotating a set of examples for each predicator (including verbs, nouns and adjectives), and describing the network of relations among the semantic frames. An example of the latter type is the Salsa project (Burchardt *et al.*, 2004), which produced a German lexicon based on the FrameNet semantic frames and annotated a large German newswire corpus. A second example, the Prague Dependency Treebank (Hajic *et al.*, 2001), has annotated a large Czech corpus with several levels of (tectogrammatical) representation, including parts of speech, syntax, and topic/focus information structure. Finally, the IL-Annotation project (Reeder et al., 2004) focused on the representations required to support a series of increasingly semantic phenomena across seven languages (Arabic, Hindi, English, Spanish, Korean, Japanese and French). In intent and in many details, OntoNotes is compatible with all these efforts, which may one day all participate in a larger multilingual corpus integration effort.

References

O. Babko-Malaya, M. Palmer, N. Xue, A. Joshi, and S. Kulick. 2004. Proposition Bank II: Delving Deeper, *Frontiers in Corpus Annotation, Workshop, HLT/NAACL*

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of COLING/ACL*, pages 86-90.

J. Chen and M. Palmer. 2005. Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features In *Proceedings of IJCNLP2005*, pp. 933-944.

B. Dorr and N. Habash. 2001. Lexical Conceptual Structure Lexicons. In Calzolari et al. ISLE-IST-1999-10647-WP2-WP3, *Survey of Major Approaches Towards Bilingual/Multilingual Lexicons.*

A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Pado, and M. Pinkal. 2006. Consistency and Coverage: Challenges for exhaustive semantic annotation. In Proceedings of DGfS-06.

C. Fellbaum (ed.). 1998. *WordNet: An On-line Lexical Database and Some of its Applications*. MIT Press.

A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. 2002. Sweetening Ontologies with DOLCE. In *Proceedings of EKAW* pp. 166-181.

J. Hajic, B. Vidová-Hladká, and P. Pajas. 2001: The Prague Dependency Treebank: Annotation Structure and Support. *Proceeding of the IRCS Workshop on Linguistic Databases*, pp. 105–114.

M. Marcus, B. Santorini, and M. A. Marcinkiewicz. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics* 19: 313-330.

A. Meyers, R. Reeves, C Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. 2004. The NomBank Project: An Interim Report. *Frontiers in Corpus Annotation*, *Workshop in conjunction with HLT/NAACL.*

I. Niles and A. Pease. 2001. Towards a Standard Upper Ontology. *Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS-2001).*

M. Palmer, O. Babko-Malaya, and H. T. Dang. 2004. Different Sense Granularities for Different Applications, *2nd Workshop on Scalable Natural Language Understanding Systems, at HLT/NAACL-04,*

M. Palmer, H. Dang and C. Fellbaum. 2006. Making Fine-grained and Coarse-grained Sense Distinctions, Both Manually and Automatically, *Journal of Natural Language Engineering*, to appear.

M. Palmer, D. Gildea, and P. Kingsbury. 2005. The Proposition Bank: A Corpus Annotated with Semantic Roles, *Computational Linguistics*, 31(1).

A. Philpot, E.. Hovy, and P. Pantel. 2005. The Omega Ontology. Proceedings of the ONTOLEX Workshop at IJCNLP

S. Pradhan, W. Ward, K. Hacioglu, J. Martin, D. Jurafsky. 2005. Semantic Role Labeling Using Different Syntactic Views. *Proceedings of the ACL.*

F. Reeder, B. Dorr, D. Farwell, N. Habash, S. Helmreich, E.H. Hovy, L. Levin, T. Mitamura, K. Miller, O. Rambow, A. Siddharthan. 2004. Interlingual Annotation for MT Development. *Proceedings of AMTA.*