OntoNotes: A Large Training Corpus for Enhanced Processing

Ralph Weischedel weischedel @bbn.com

Eduard Hovy hovy @isi.edu

Mitchell Marcus mitch @cis.upenn.edu

Martha Palmer martha.palmer @colorado.edu

Robert Belvin rsbelvin @mac.com

Sameer Pradhan pradhan @bbn.com

Lance Ramshaw lance.ramshaw @bbn.com

Nianwen Xue xuen @cs.brandeis.edu

Abstract[1]

This paper describes a large multilingual richly annotated corpus which is being made available to the community. There is an emphasis on quality and consistency with interannotator agreement rates targeted at 90%. The data covers multiple genres in English, Chinese, and Arabic, including a significant amount of parallel data. The annotation, intended to capture a skeletal representation of literal meaning, includes parse trees, predicate argument structures , word senses localized in an ontology, coreference, and name types. The resource is delivered as an integrated database, supporting combined queries that access multiple annotation layers. Annual incremental releases are distributed via LDC.

1 Motivation, Goals, and Rationale

Our goal is to provide data in multiple languages and multiple genres (newswire, broadcast news, broadcast conversation, and web text), richly annotated by a skeletal representation of the literal meaning of sentences, so that a new generation of language understanding would deliver new functional capability. Our inspiration has been the impact on research and on applications of two seminal annotation products: the UPenn Treebank for syntax (Marcus, *et al*., 1993) and PropBank for semantic role labeling (Palmer *et al*., 2005).

As shown in Figure 1, to the baseline structure of parse trees and propositions, OntoNotes adds

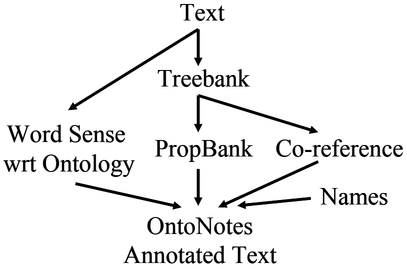


Figure 1. Annotation levels in OntoNotes

- Referring expressions and the textual phrases they refer to
- Terms disambiguated by word sense and localized in an ontology
- Named Entities

Based on our interpretation of the criteria for successfully applying learning algorithms, our guiding principle has been to find a "sweet spot" in the space of

- Inter-tagger agreement, so that human agreement as a ceiling on algorithm performance is as high as possible.
- Productivity, so that the amount of training data is maximized, given a budget,
- Depth of representation, so that the added semantic features are as deep as possible.

The methodology described here was tested prior to entering production mode, where pilot rounds of annotation were conducted to find the sweet spot above. In particular, only those classes of co-reference satisfying the methodology above during the pilot study are annotated. The methodology has been applied for each word; the sense inventory for a word is selected according to the criteria above.

Another dimension of the OntoNotes product is the integration of all of the annotations in a data-

[1] This work was supported under the GALE program of the Defense Advanced Research Projects Agency, Contract No. HR0011-06-C-0022.

base (Pradhan *et al.,* 2007a), which has at least two benefits:

- Consistency checks on entering each annotation element flag many inconsistencies across annotations for manual correction.
- The data may be searched for phenomena of interest.

This paper illustrates annotation primarily of English, though OntoNotes covers Arabic and Chinese as well. In the next sections, we describe each of the component annotations: treebanking, proposition banking, word sense, ontology creation, coreference, and names. The paper concludes with a summary of related work.

2 Treebanking

The Treebank style in OntoNotes for English is a modification of the Treebank II style for the Penn Treebank (Marcus *et al*., 1993; Marcus *et al., 1994*). For Chinese, the style follows the Chinese Treebank (Xue *et al.*, 2005). (Arabic Treebank annotation is being performed at the Linguistic Data Consortium.) These are annotated with information to make predicate-argument structure easy to decode, including function tags and markers of "empty" categories that represent displaced constituents.

To facilitate merging of the syntactically annotated material with the PropBanked material, both the Treebank style and PropBank style were modified to correct for some small mismatches between the annotations (Babko-Malaya *et el.,* 2006). The major changes to the Treebank stylebook involved modifying the list of verbs considered to take so-called "small clauses" to conform to the argument structures assigned by PropBank, and changing the structures of resultatives to match the PropBank analysis.

The internal consistency of the newly syntactically annotated material for English has been tested, and is quite good. The principal annotator for English reannotated sampled material a year after the original annotation. The F-measure of the newly annotated material against the initial annotation by the EVALB measure was 98.5.

A major editing pass of the OntoNotes Treebank materials is now underway to achieve full consistency with all materials treebanked under the GALE project. The first modification retrofits much of the OntoNotes treebanked materials to conform with the current LDC syntactic style for NPs, crucially adding branching structure whenever the default right branching structure of pre-head modifiers is violated. The second modification eliminates most token-internal hyphens, eliminating the anomaly created by the earlier tokenization of "the New York-based company,", which was based entirely on white space.

3 PropBanking

PropBanking focuses on annotating the argument structure of verbs, and provides a corpus annotated with semantic roles, including participants traditionally viewed as arguments and adjuncts. The style for English is that of the 1M word Penn Treebank II Wall Street Journal corpus (Palmer *et al.*, 2005). In addition to annotating verbs we are also applying Nombank style annotation to just those nouns with predicate-argument structures that can participate in event coreferences, such as nominalizations and eventive nouns. Links from the argument labels in the Frames Files to FrameNet frame elements and VerbNet thematic roles have been added. This style of annotation has also been successfully applied to other genres and languages. For Chinese, the style is that of (Xue & Palmer, 2009) The same style has also been applied to Arabic (Diab *et. al.*, 2007).

4 Word Sense

One of the daunting challenges was attaining 90% annotator agreement for word sense, since, for example, WordNet inter-annotator agreement averages in the low 70s. 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.,* 2007), 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 (Philpot *et al.*, 2005; see Section 5).

Figure 2 shows the empirical process for proposing meaningful sense distinctions and determining if they could be annotated at 90% accuracy. A 50-sentence sample of instances is annotated and immediately checked for inter-annotator agreement for all verbs and any noun with frequency over 100. ITA scores below 90% lead to a revision and clarification of the groupings by the linguist. It is

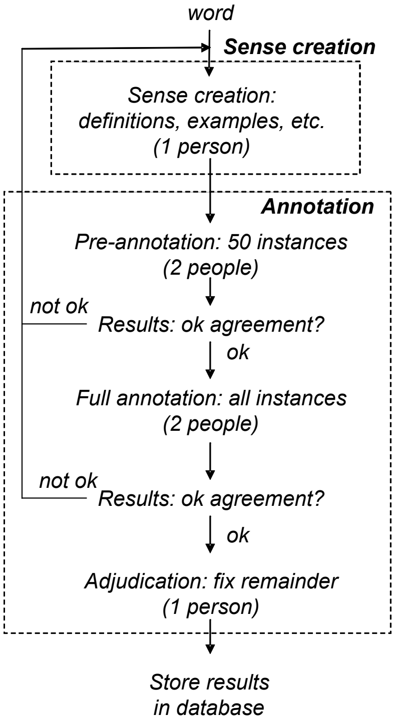


Figure 2. Annotation procedure

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.

The same methodology has been applied to English, Arabic, and Chinese; the only difference is the starting point for suggesting sense inventories. For English, WordNet has been our starting point of choice. For Chinese, diverse sources are reviewed before hypothesizing an inventory, including entries in web-accessible dictionaries, print dictionaries, samples from the corpora to be annotated, and general web searches. For Chinese verbs, the starting point has been the course-grained senses in the frame files created for the PropBank annotation, although print and electronic dictionaries are also consulted. Similar research including access to Arabic WordNet is carried out for each Arabic word prior to authoring its sense inventory file.

4.1 Verbs

The word sense inventories for English verbs come initially from grouping related WordNet senses (Palmer, et. al., 2007). Subcategorization frames and semantic classes of arguments play major roles in determining the groupings, as illustrated by the grouping for the 22 WN 2.1 senses for *drive* in Figure 3. The groupings are also linked to the ontology (see Section 5.) In addition to improved annotator productivity and accuracy, we have found a corresponding improvement in word sense disambiguation performance. Training on this new data, Dligach and Palmer (2008) report 83% accuracy for verbs using a Support Vector Machine and rich linguistic features, which is almost 20% higher than state-of-the art performance on ungrouped, fine-grained senses (Chen and Palmer, 2005). The sense inventories for Chinese and Arabic verbs are created by starting with the PropBank frame files and subdividing the verb entries into more fine-grained senses where deemed appropriate.

4.2 Nouns

Noun annotation follows a procedure similar to that for verbs. The noun senses are created starting with WordNet and other dictionaries. We aim to double-annotate, at the target agreement level, the 1100 most frequent polysemous English nouns in the entire corpus before the end of 2009, while maximizing overlap with the sentences containing annotated verbs. We have lower targets for the other two languages, which were started later.

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.," WN5: "She is driven by her passion," WN6: "drive back the invaders," WN7: "She finally drove him to change jobs," WN15: "drive the herd," WN22: "drive the game."
G3: to exert energy on behalf of something *NP drive NP/infinitival*	WN11: "What are you driving at?," 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"

Figure 3. Four Groups for "drive", Compared to the WordNet Senses

Certain nouns carry predicate structure. To ensure conformity with verbs, the structure of nominalizations (*destruction*) and eventive nouns (*party*) is created and assigned by the verb specialists at Colorado.

In order to speed up annotation, we investigated a form of active learning, in which nouns with high agreement in a subset of the whole corpus were used as training data by an automated annotation learner. Unfortunately, different sense distributions across corpora meant that we could not always use results from one year to automatically annotate another year's data. We investigated various strategies to bootstrap the learning, by mixing into the training data small amounts of annotated data from the new corpus. The results show that even 50 instances from the new distribution permit learning that is accurate enough for about 50% of the high-frequency nouns (Zhu and Hovy, 2007; Zhu *et al.*, 2008).

The OntoNotes (release 1.0) verb and noun word sense data was used in the Semeval-1 (Pradhan *et al.*, 2007b) Overall accuracy over 100 lemmas (65 verbs and 35 nouns) from WSJ corpus, for the best performing system was 86% — with average over verbs being 78% and over nouns being 89%.

4.3 Coverage Issues

There are far too many polysemous lexical items for any project to provide exhaustive coverage. Therefore the prioritization of items for annotation is of pressing concern.

Clearly high frequency items provide the most leverage, but they often have a predominant sense (as much as 90% of the data) which can overwhelm annotators with hundreds or even thousands of repetitive examples that will provide little if no system performance improvement. For example, 183 of the 186 instances of the word "bank" in the Ontonotes portion of the WSJ corpus are cases of the first of the 10 senses (a financial institution). In all corpora combined, 607 of the 640 instances of "investment" are the third sense (the activity of investing money for profit). With this type of data, double-blind manual annotation and adjudication is not really necessary.

Of course, prior to the manual annotation the entropy of a word's sense distribution is unknown. When we are partially through annotation of a given word and it is clear that the majority of its instances fall into one sense, we can then dispense with full double-annotation of all senses, and allow one of the "annotators" to be a trained classifier. This method is described in more detail in (Zhu and Hovy, 2007). This allows for somewhat quicker annotation progress to be carried out for the most common senses, so more time can be devoted to human annotation of rarer words and senses.

The desired aim is a balance between sufficient coverage of high frequency items and maximal coverage of low frequency ones. For these rare words and senses the greatest challenge is finding enough instances to provide adequate training material. We are also exploring techniques such as language modeling for preselecting instances of rare senses from a new corpus (Dligach and Palmer, submitted). In addition we have implemented a data selection plan which supplements our "whole document" based annotation approach with lexical samples for specific lexical items which require greater coverage.

5 Ontology

Standard dictionaries simply list the senses for each word. To support synonym access, inheritance of features and other properties such as predicate frames, links to instances, and so on, we group together the senses that share the same meaning, and then arrange them into a shallow taxonomy that we call the Omega Ontology (Philpot *et al.*, 2005) following the process of Figure 4.

A manual procedure forms *sense pools* by selecting and grouping together individual noun and verb senses that convey the same meaning. (Sense pools correspond to WordNet's synsets, but are generally less fine-grained.) Each sense pool contains one or more definitions, examples, features, and pointers to the individual senses that comprise it, from which one can access their respective annotated sentences. It is thus possible to assemble for each meaning a set of sentences that contain different target words, each expressing that meaning, in order to train more-powerful sense disambiguation engines.

All sense pools are attached into Omega's Upper Model (Hovy *et al.*, 2009), a network of some 120 nodes that represent very abstract conceptualizations. Reference to VerbNet semantic classes has been helpful in creating nodes for the verb upper level ontology (Palmer *et. al.*, 2009). To date, about 5000 noun-derived and 3500 verb-derived pools (from English sources) have been created and attached, by multiple annotators who compare their decisions to ensure quality, using a specialized interface. Work is underway to create sense structures also for the other two languages, and either merge them into English-derived pools or attach them to the Upper Model separately. In addition, we have started creating sense structures for the 3000-odd monosemous English nouns occurring in the corpus, and merging or inserting them into Omega. Figure 5 shows the interface for aligning monosemous pools (left pane) into or nearby pools already taxonomized under the Upper Model (options listed in the right pane, some numbered as Pxxx and some named), and one option displayed in the center pane). In this example, a pool representing *Type* (subdivision, kind of something) is compared to a pool representing *Type/Font* (printed characters).

6 Coreference

The coreference annotation in OntoNotes connects coreferring instances of specific referring expressions, 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 noun 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”.

However, the sense of “be” is marked so that attributive information is annotated. 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.

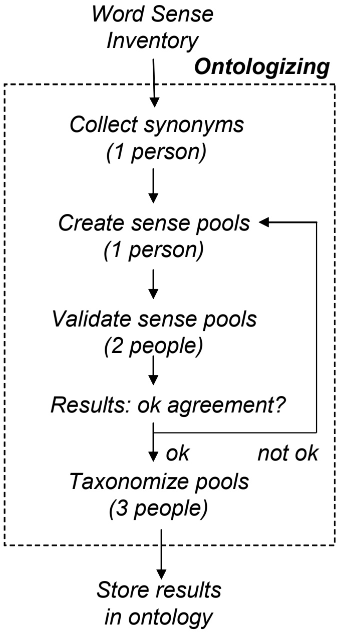

Word Sense Inventory
Ontologizing
Collect synonyms (1 person)
Create sense pools (1 person)
Validate sense pools (2 people)
Results: ok agreement?
ok
not ok
Taxonomize pools (3 people)
Store results in ontology

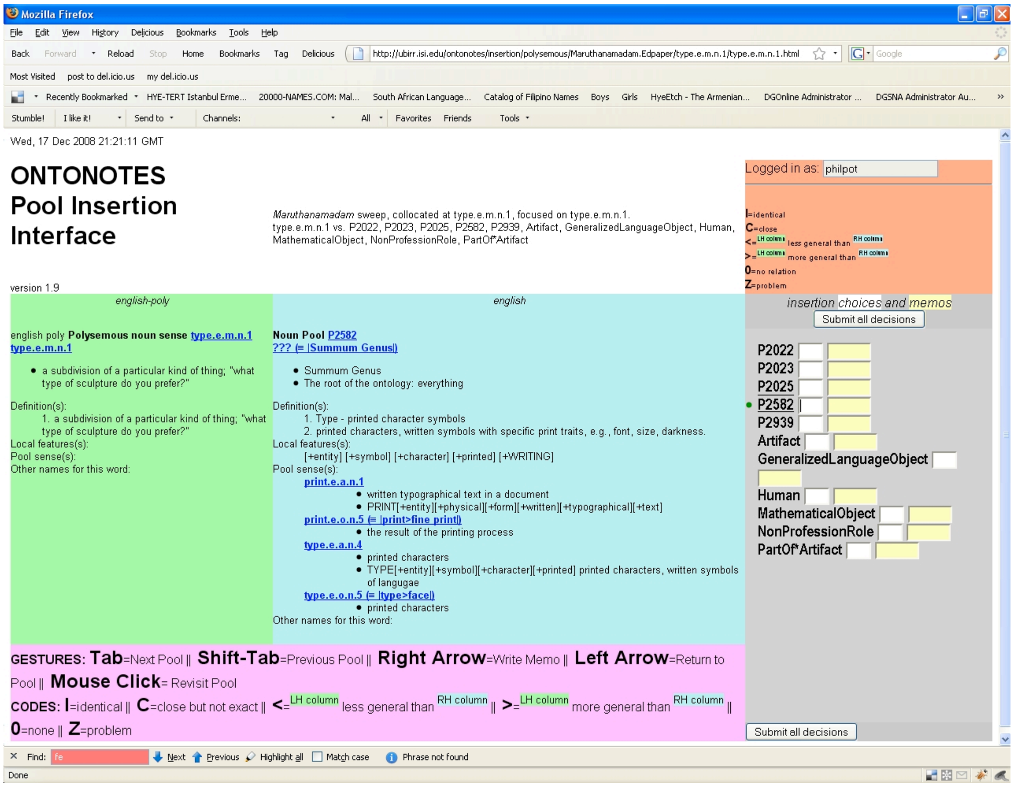
Figure 4: Process for adding to the Ontology

All of the coreference annotation is being doubly annotated and adjudicated. Over the first two years, the overall average agreement between individual annotators and the adjudicated result for non-appositive coreference using the MUC coreference scorer was 86%.

Pradhan *et al.* (2007c) report baseline performance on the OntoNotes coreference data using a standard feature set. Coreference decoding contrasts with decoding other layers in that system performance on coreference still lags very much behind the ITA, in spite of the latter being very high. This is most likely due to the fact that richer semantic and word-knowledge components, in addition to annotation granularity and consistency, are important in identifying co-referring entities. Better learning strategies combined with the accompanying layers in OntoNotes would likely help bridge this gap in the future.

7 Names

Names are also annotated using an 18-type superset of the ACE name guidelines. This supplemental annotation is done in a single pass.


Mozilla Firefox
File Edit View History Delicious Bookmarks Tools Help
Back Forward Reload Stop Home Bookmarks Tag Delicious
http://ubirr.isi.edu/ontonotes/insertion/polysemous/Maruthanamadam.Edpaper/type.e.m.n.1/type.e.m.n.1.html
Google
Most Visited post to del.icio.us my del.icio.us
Recently Bookmarked HYE-TERT Istanbul Erme... 20000-NAMES.COM: Mal... South African Language... Catalog of Filipino Names Boys Girls HyeEtch - The Armenian... DGOnline Administrator ... DGSNA Administrator Au... »
Stumble! I like it! Send to Channels: All Favorites Friends Tools
Find: fe Next Previous Highlight all Match case Phrase not found
Done

Wed, 17 Dec 2008 21:21:11 GMT

ONTONOTES
Pool Insertion
Interface

Maruthanamadam sweep, collocated at type.e.m.n.1, focused on type.e.m.n.1.
type.e.m.n.1 vs. P2022, P2023, P2025, P2582, P2939, Artifact, GeneralizedLanguageObject, Human, MathematicalObject, NonProfessionRole, PartOf*Artifact

version 1.9

Logged in as: philpot

I=identical
C=close
<=LH column less general than RH column
>=LH column more general than RH column
0=no relation
Z=problem

english-poly

english poly **Polysemous noun sense** type.e.m.n.1
type.e.m.n.1

- a subdivision of a particular kind of thing; "what type of sculpture do you prefer?"

Definition(s):
1. a subdivision of a particular kind of thing; "what type of sculpture do you prefer?"

Local features(s):
Pool sense(s):
Other names for this word:

english

Noun Pool P2582
??? (≡ |Summum Genus|)

- Summum Genus
- The root of the ontology: everything

Definition(s):
1. Type - printed character symbols
2. printed characters, written symbols with specific print traits, e.g., font, size, darkness.

Local features(s):
[+entity] [+symbol] [+character] [+printed] [+WRITING]

Pool sense(s):
print.e.a.n.1
- written typographical text in a document
- PRINT[+entity][+physical][+form][+written][+typographical][+text]

print.e.o.n.5 (≡ |print>fine print|)
- the result of the printing process

type.e.a.n.4
- printed characters
- TYPE[+entity][+symbol][+character][+printed] printed characters, written symbols of langugae

type.e.o.n.5 (≡ |type>face|)
- printed characters

Other names for this word:

insertion choices and memos
Submit all decisions

P2022
P2023
P2025
● P2582
P2939
Artifact
GeneralizedLanguageObject
Human
MathematicalObject
NonProfessionRole
PartOf*Artifact

Submit all decisions

GESTURES: **Tab**=Next Pool || **Shift-Tab**=Previous Pool || **Right Arrow**=Write Memo || **Left Arrow**=Return to Pool || **Mouse Click**= Revisit Pool
CODES: **I**=identical || **C**=close but not exact || **<**=LH column less general than RH column || **>**=LH column more general than RH column || **0**=none || **Z**=problem

Figure 5. Manually aligning pools to form Ontology contents

8 Database

Since we are delivering multiple levels of annotation (syntax, propositions, coreference, word sense and ontology), several questions arose:

1. How could we ensure that all the components are consistent with each other, avoiding engineering/formatting inconsistencies?
2. Should the annotations be delivered as independent pieces provided in an integrated representation?
3. What representation would best facilitate use of this information as training data for systems that will be incorporated into applications? Can this representation also support leveraging these additional knowledge sources during the training process?

We have created a corpus with diverse levels of semantic information integrated in one database (Pradhan *et al.*, 2007a). Figure 6 illustrates some of the interconnections captured.

This process identified several levels of inconsistencies that were resolved, ensuring a clean, consistent final product. The relationships between all the layers and within the layers themselves are efficiently captured in the database schema.

We have also provided an object layer on top of the database layer, written in Python, which can flexibly manipulate the data at the level of the database or as objects, to extract information across layers. It can also produce the individual layers by themselves as well as a human-readable representation.

This facilitates defining custom views of the data as well as extracting cross-layer features for use in predictive models, neither of which was easily possible before.

9 Related 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 annotation (eventuality ID's, coarse-grained sense tags, nominal coreference and selected discourse connectives) has been applied to a small (100K) parallel Chinese/English corpus (Babko-Malaya *et al.*,

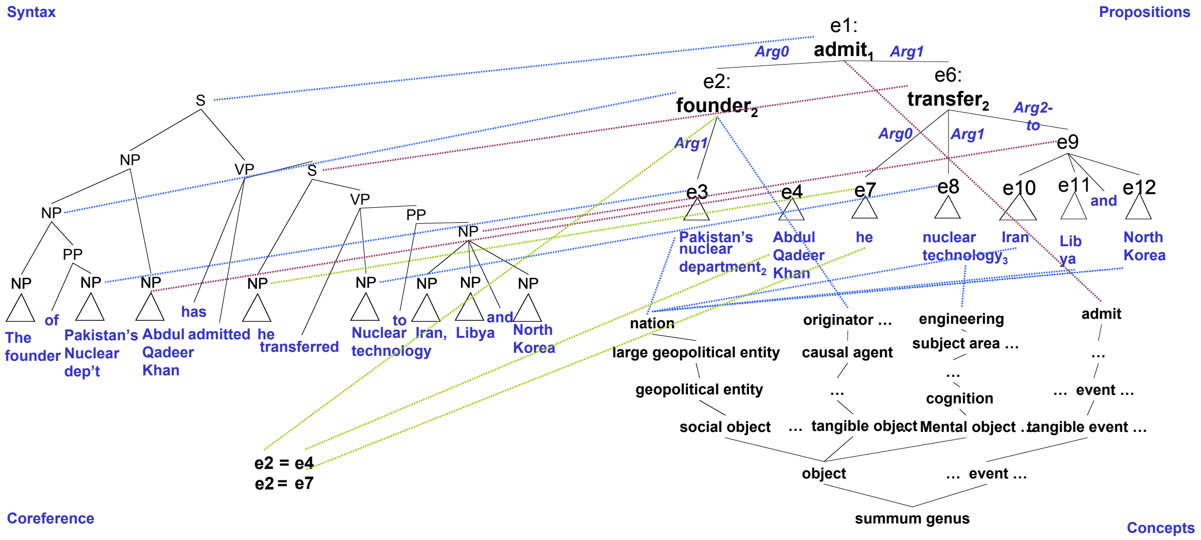


Figure 6. Simplified diagram of the interconnections between annotation layers.

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-score of 81.0 using an SVM model. 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. The Tsinghua Chinese Treebank TCT (Zhou, 2003) contains some 2 million Chinese characters, of which half has been treebanked, and manually annotated for syntactic and certain semantic relations, such as causality and conditionals. It covers various genres

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.

10 Summary

The plan for the full OntoNotes corpus is shown in Figure 7, covering three languages and four genres (NewsWire, Broadcast News, Broadcast Conversation, and Web text), and including significant amounts of parallel bilingual data. OntoNotes Version 2.0, released by the LDC in early 2008, covered NW and BN, and Version 3.0, to be released in April 2009, will add coverage of BC data.[2] It is our hope that this annotation will provide an enduring resource for the community.

[2] The Arabic data releases are on a separate genre schedule.

	English	Chinese	Arabic
NW	550 K	250 K	300 K
BN	200 K	300 K	200 K
BC	200 K	150 K	–
Web	300 K	150 K	–

Figure 7. Planned corpus (token counts)

References

O. Babko-Malaya, A. Bies, Ann Taylor, S. Yi, M. Palmer, M. Marcus, S. Kulick and L. Shen. 2006. Issues in Synchronizing the English Treebank and PropBank. *Workshop on Frontiers in Linguistically Annotated Corpora* 2006.

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 IJCNLP-2005*, pp. 933-944.

M. Diab; M. Alkhalifa; S. ElKateb; C. Fellbaum; A. Mansouri; M. Palmer, 2007. SemEval-2007 Task 18: Arabic Semantic Labeling, *Proceedings of SemEval-2007.*

D. Dligach and M. Palmer. 2008. Novel Semantic Features for Verb Sense Disambiguation. In *Proceedings of ACL-08.*

D. Dligach and M. Palmer. 2009. Using Language Modeling to Select Useful Annotation Data. Submitted to NAACL.

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

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

AJ. 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.

E.H. Hovy, A. Philpot, et al. 2009. The Omega Upper Model. Unpublished ms.

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

M. Marcus, G. Kim, M.A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn Treebank: annotating predicate argument structure. In *Proceedings of the HLT Workshop*.

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.*

M. Palmer, H. Dang and C. Fellbaum. 2007. Making Fine-grained and Coarse-grained Sense Distinctions, Both Manually and Automatically, *Journal of Natural Language Engineering*, 13:2, 137-163.

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

M. Palmer, J. Hwang, S. W. Brown, K. K. Schuler and A. Lanfranchi, 2009 Leveraging lexical resources for the detection of event relations. *AAAI 2009 Spring Symposium on Learning by Reading*.

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

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

S. Pradhan, E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel. 2007a. OntoNotes: A Unified Relational Representation. *International Journal of Semantic Computing.* 1:4 405-419

S. Pradhan, E. Loper, D. Dligach and M. Palmer. 2007b. SemEval-2007 Task 17: English Lexical Sample, SRL and All Words. *Proceedings of SemEval-2007.*

S. Pradhan, L. Ramshaw, R. Weischedel, J. MacBride, L. Miccuilla. 2007c. Unrestricted Coreference: Identifying Entities and Events in OntoNotes. *Proceedings of ICSC.*

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.*

N. Xue, F. Xia, C. Chiou and M. Palmer. 2005. The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. *Natural language Engineering*, 11(2): 207-238.

N. Xue and M. Palmer, Adding semantic roles to the Chinese Treebank. *Natural Language Engineering* 15(1): 143-172.

Q. Zhou. 2003. Build a Large-Scale Syntactically Annotated Chinese Corpus. *Springer Lecture Notes in Computer Science 2807*, pp 106–113.

J. Zhu, H. Wang, and E.H. Hovy. 2008. Multi-Criteria-based Strategy to Stop Active Learning for Data Annotation. *Proceedings of COLING*.

J. Zhu and E.H. Hovy. 2007. Active Learning for Word Sense Disambiguation with Methods for Addressing the Class Imbalance Problem. In *Proceedings of EMNLP.*