

# Semantic Annotation and Lexico-Syntactic Paraphrase

Bonnie J. Dorr<sup>†</sup>, Rebecca Green<sup>†</sup>, Lori Levin<sup>‡</sup>, Owen Rambow<sup>\*</sup>,  
David Farwell<sup>§</sup>, Nizar Habash<sup>†\*</sup>, Stephen Helmreich<sup>§</sup>, Eduard Hovy<sup>¶</sup>,  
Keith J. Miller<sup>^</sup>, Teruko Mitamura<sup>‡</sup>, Florence Reeder<sup>^</sup>, Advait Siddharthan<sup>\*</sup>

<sup>†</sup>Institute for Advanced Computer Studies  
University of Maryland  
{bonnie, rgreen, habash}@umiacs.umd.edu

<sup>‡</sup>Language Technologies Institute  
Carnegie Mellon University  
{lsl, teruko}@cs.cmu.edu

<sup>\*</sup>Department of Computer Science  
Columbia University {rambow, as372}@cs.columbia.edu

<sup>§</sup>Computing Research Laboratory  
New Mexico State University  
{david, shelmrei}@.crl.nmsu.edu

<sup>¶</sup>Information Sciences Institute  
University of Southern California hovy@isi.edu

<sup>^</sup>MITRE Corporation  
{freeder, keith}@mitre.org

## Abstract

The IAMTC project (Interlingual Annotation of Multilingual Translation Corpora) is developing an interlingual representation framework for annotation of parallel corpora (English paired with Arabic, French, Hindi, Japanese, Korean, and Spanish) with deep-semantic representations. In particular, we are investigating meaning equivalent paraphrases involving conversives and non-literal language use, as well as extended paraphrases involving syntax, lexicon, and grammatical features; meaning equivalent paraphrases will receive the same deep-semantic representation. Current evaluation efforts measure intercoder agreement on semantic concepts and thematic roles; ultimately, the ability to generate text from the interlingual representation comparable to the source texts will also be assessed. Examples of representations at incrementally deeper levels are given.

## 1. Introduction

An important issue for computational linguists and lexicographers is the question of meaning-equivalent paraphrases, including lexical synonymy, conversives (*buy/sell*), idioms (*kick the bucket/die*), and more extended paraphrases, such as *Initially the estimate given by X for Y w as Z* and *X had stated that Y would be Z*, which involve lexicon, syntax, and grammatical tense.

Annotation projects such as PropBank (Kingsbury and Palmer, 2002) and FrameNet (Baker et al., 1998) do not cover extended paraphrases. Automatic extraction of extended paraphrases (Barzilay and McKeown, 2001) relies on unsupervised learning and therefore requires large amounts of data (i.e., rarer paraphrases are not detected—non-perfect recall) and an error rate that may be fairly high (i.e., non-perfect precision).

Our approach in the IAMTC project (Interlingual Annotation of Multilingual Translation Corpora) is complementary to those of projects that focus on improvements to unsupervised learning techniques: We are producing a framework for hand annotation of parallel corpora with deep-semantic representations.

In contrast to other annotation projects, we are annotating texts in seven languages (Arabic, English, French, Hindi, Japanese, Korean, and Spanish). We are investigating meaning-equivalent paraphrases by annotating multiple versions of the same text, usually one non-English source language and two independently produced English translations.

## 2. Project Overview

The objectives of the IAMTC project are:

- To develop an interlingual representation framework based on a careful study of text corpora in six languages. The framework will include a formal definition of the representation, coding manuals, resources, etc.
- To annotate bilingual corpora using the interlingual representation. This representation allows us to relate potentially divergent pairs such as the following noun-modifier case: Spanish *anuncios sísmicos* (earthquake announcements) and English *predicted earthquakes*.

- To extend our current set of annotation tools (a tree editor, annotation interface, etc.).
- To design and test a variety of new evaluation metrics for assessing the interlingual representations and choosing an appropriate granularity of meaning representation.

Recognizing the complexity of interlinguas, we adopt an incrementally deepening approach, which allows us to produce relatively stable annotations while exploring alternatives at the next level down. We currently identify three levels of representation, referred to as IL0, IL1, and IL2. Each level of representation incorporates additional semantic features and removes existing syntactic ones.

IL0 is a deep syntactic dependency representation, constructed by hand-correcting the output of a dependency parser. Though this representation is purely syntactic, the presentation abstracts as much as possible from surface-syntactic phenomena. (Thus it is more abstract than the Praguian Analytical level, but more syntactic than the Tecogrammatical level.) By allowing annotators to see how textual units relate syntactically when making semantic judgments, IL0 is a useful starting point for semantic annotation at IL1.

IL1 is an intermediate semantic representation. It associates semantic concepts drawn from an ontology with nouns, adjectives, adverbs and verbs. It also replaces the syntactic relations in IL0 with thematic roles such as AGENT, THEME, and GOAL. However, IL1 is not an interlingua; it does not normalize over all linguistic realizations of the same semantics.

IL2 is intended to be an interlingua; it is still under development. As a representation of meaning that is (reasonably) independent of language, IL2 captures similarities in meaning across languages and across different lexical/syntactic realizations within a language. For example, IL2 is expected to normalize over:

- Conversives (e.g., *X bought a book from Y* vs. *Y sold a book to X*), as does FrameNet (Baker et al., 1998) at the more general level of Commercial.transaction.
- Non-literal language usage (e.g., *X started its business* vs. *X opened its doors to customers*).
- Extended paraphrases involving syntax, lexicon, and grammatical features (see example in introduction).

The range of paraphrase phenomena being addressed by the different representation levels is summarized in Table 1, which is based on examples from (Hirst, 2003), (Kozlowski et al., 2003), and (Rinaldi et al., 2003). The table indicates for which types we expect to produce normalized representations reflecting the similarity in meaning between paraphrases of that type and at which level the normalization will take place.

### 3. Work to Date

We have trained approximately 12 IL1 annotators, who have produced over 150 annotated English newspaper articles (six source language articles times two English trans-

lations of each article times 12 annotators, plus some practice annotations). Experts prepare the IL0 for each text using the tree editor TrEd from Prague (Pajas, 1998). Annotators use our interface, TIAMAT, and the Omega ontology (Hovy et al., 2003) to produce IL1 by adding concept names from WordNet (Fellbaum, 1998) and Mikrokosmos (O’Hara et al., 1998), and thematic role names from the Lexical Conceptual Structure verb database (Dorr et al., 2001).

For example, consider the following input sentence:

The study led them to ask the Czech government to recapitalize CSA at this level.

Below is a sample IL1 corresponding to this sentence. The TIAMAT interface displays the information more nicely for annotators, who can also consult the tree in TrEd. Concept names and thematic role names added by the annotators are in all caps; some nodes have two concepts.

```
[ 30, led, V, lead, Root, LEAD<GET, GUIDE ]
[ 20, study, N, study, AGENT, SURVEY<WORK, REPORT ]
[ 40, them, N, they, THEME, ---, --- ]
[ 60, ask, V, ask, PROPOSITION, ---, --- ]
[ 45, they, N, they, AGENT, ---, --- ]
[ 90, government, N, government, GOAL, AUTHORITIES, GOVERNMENTAL-ORGANIZATION ]
[ 80, Czech, Adj, Czech, MOD, CZECH~CZECHOSLOVAKIA, --- ]
[ 110, recapitalize, V, recapitalize, PROP, CAPITALIZE<SUPPLY, INVEST ]
[ 105, government, N, government, AGENT, AUTHORITIES, GOVERNMENTAL-ORGANIZATION ]
[ 108, Czech, Adj, Czech, MOD, CZECH~CZECHOSLOVAKIA, --- ]
[ 120, csa, N, csa, THEME, AIRLINE<LINE, --- ]
[ 160, at, P, value_at, GOAL, ---, --- ]
[ 150, level, N, level, ---, DEGREE, MEASURE ]
[ 140, this, Det, this, ---, ---, --- ]
```

Inter-coder agreement is being calculated (see below).

### 4. Expected Benefit – Identification of Paraphrases

The ability to discern paraphrases is beneficial to virtually all linguistic applications, including information retrieval, information extraction, question-answering, text summarization, and machine translation. Two sentences with the same IL2 are considered paraphrases even if they have different IL1’s. The IL2 annotation on the corpus will allow us to easily trace the different surface realizations of a given meaning pattern. Paraphrases can arise, for example, from conversives, as shown here:

Relationship Type	Example	Where Normalized
Syntactic variation	The gangster killed at least 3 innocent bystanders. <i>vs.</i> At least 3 innocent bystanders were killed by the gangster.	IL0
Lexical synonymy	The toddler sobbed, and he attempted to console her. <i>vs.</i> The baby wailed, and he tried to comfort her.	IL1
Morphological derivation	I was surprised that he destroyed the old house. <i>vs.</i> I was surprised by his destruction of the old house.	IL2
Clause subordination <i>vs.</i> anaphorically linked sentences	This is Joe's new car, which he bought in New York. <i>vs.</i> This is Joe's new car. He bought it in New York.	IL2
Different argument realizations	Bob enjoys playing with his kids. <i>vs.</i> Playing with his kids pleases Bob.	IL2
Noun-noun phrases	She loves velvet dresses. <i>vs.</i> She loves dresses made of velvet.	IL2
Head switching	Mike Mussina excels at pitching. <i>vs.</i> Mike Mussina pitches well. <i>vs.</i> Mike Mussina is a good pitcher.	IL2
Overlapping meanings	Lindbergh flew across the Atlantic Ocean. <i>vs.</i> Lindbergh crossed the Atlantic Ocean by plane.	IL2
Comparatives <i>vs.</i> superlatives	He's smarter than everybody else. <i>vs.</i> He's the smartest one.	Not normalized
Different sentence types	Who composed the Brandenburg Concertos? <i>vs.</i> Tell me who composed the Brandenburg Concertos.	Not normalized
Inverse relationship	Only 20% of the participants arrived on time. <i>vs.</i> Most of the participants arrived late.	Not normalized
Inference	The tight end caught the ball in the end zone. <i>vs.</i> The tight end scored a touchdown.	Not normalized
Viewpoint variation	The U.S.-led invasion/liberation/occupation of Iraq . . . You're getting in the way. <i>vs.</i> I'm only trying to help.	Not normalized

Table 1: Relationship Types Underlying Paraphrase

IL2:	English IL1:
GO BOOK FROM JOHN TO MARY VIA PURCHASE	lend
IL1 candidate #1:	AGENT: network
buy	MOD: comprise
MARY[agent]	PART: eighteen independent organizations
JOHN[source]	THEME: ...
BOOK[theme]	
IL0 candidate #1:	French:
Mary bought the book from John	Le réseau regroupe dix-huit organisations indépendantes qui ont déboursé ...
IL1 candidate #2:	'the network comprises eighteen independent organizations which have disbursed'...
sell	French IL1:
JOHN[agent]	comprise
MARY[goal]	WHOLE: network
BOOK[theme]	PART: 18 organizations
IL0 candidate #2:	RECL-CL: disburse
John sold the book to Mary	AGENT: network
Paraphrases also arise from different sentence plans for expressing the same information, as shown here:	THEME: ....
English:	
Its network of eighteen independent organizations in Latin America has lent ...	

IL1-IL2 Mappings:

of/regroupe <-> COMPRISE  
lend/debourse <-> TRANSFER-MONEY

IL2:

COMPRISE:

WHOLE: network  
PART: eighteen independent  
organizations

TRANSFER-MONEY

AGENT: network  
THEME: ...

## 5. Evaluation

One of our goals is to test for the effects of coding two texts that are semantically close. In our initial experiments, we have measured inter-annotator agreement of our semantic annotations. After collecting data about the annotations, inter-coder agreement (of semantic concepts and thematic roles) is measured in a profile that includes a Kappa measure (Carletta, 1996) and a “Wood Standard” similarity measure (Habash and Dorr, 2002).

Another goal is to measure the accuracy of the annotation by evaluating the use of the resulting representations. Since the ultimate goal is to generate a representation that is useful for MT (among other NLP tasks), we plan to measure the ability to generate accurate surface texts from the representation. We plan to use an available generator, Halogen (Langkilde and Knight, 1998). Sentences will be generated from interlinguas and then compared with the originals through a variety of standard MT metrics (ISLE, 2003). This will serve to determine whether the elements of the representation language are sufficiently well defined and whether they can serve as a basis for inferring interpretations from semantic representations or (target) semantic representations from interpretations.

## 6. References

- Baker, Collin F., Charles J. Fillmore, and John B. Lowe, 1998. The Berkeley FrameNet Project. In Christian Boitet and Pete Whitelock (eds.), *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*. San Francisco, California: Morgan Kaufmann Publishers.
- Barzilay, R. and K.R. McKeown, 2001. Extracting paraphrases from a parallel corpus. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. Toulouse, France.
- Carletta, J.C., 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.
- Dorr, Bonnie J., Mari Olsen, Nizar Habash, and Scott Thomas, 2001. LCS Verb Database. Technical Report Online Software Database, University of Maryland, College Park, MD. [http://www.umiacs.umd.edu/~bonnie/LCS\\_Database\\_Documentation.html](http://www.umiacs.umd.edu/~bonnie/LCS_Database_Documentation.html).
- Fellbaum, Christiane, 1998. *WordNet: An Electronic Lexical Database*. MIT Press. <http://www.cogsci.princeton.edu/~wn> [2000, September 7].
- Habash, Nizar and Bonnie J. Dorr, 2002. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. In *Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*. Tiburon, CA.
- Hirst, Graeme, 2003. Paraphrasing paraphrased. In *Keynote address for The Second International Workshop on Paraphrasing: Paraphrase acquisition and Applications*. Association for Computational Linguistics.
- Hovy, Eduard, Andrew Philpot, Judith L. Klavans, Ulrich Germann, and Peter T. Davis, 2003. Extending metadata definitions by automatically extracting and organizing glossary definitions. In *Proceedings of the National Conference on Digital Government Research*. Boston, MA.
- ISLE, 2003. Towards Systematizing MT Evaluation. In *Proceedings of MT Summit IX Evaluation Workshop*. New Orleans, Louisiana.
- Kingsbury, P. and M. Palmer, 2002. From Treebank to PropBank. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*. Las Palmas, Spain.
- Kozłowski, Raymond, Kathleen F. McCoy, and K. Vijay-Shanker, 2003. Generation of single-sentence paraphrases from predicate/argument structure using lexicogrammatical resources. In *Proceedings of the Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP2003)*. ACL 2003.
- Langkilde, I. and K. Knight, 1998. Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics joint with the 17th International Conference on Computational Linguistics, ACL/COLING*. Montreal, Canada.
- O’Hara, Tom, Kavi Mahesh, and Sergei Nirenburg, 1998. Lexical acquisition with wordnet and the mikrokosmos ontology. In *Proceedings of the ACL Workshop on the Use of WordNet in NLP*. Montreal, Canada.
- Pajas, Petr, 1998. Tree editor manual. Technical report, CLSP Summer Sorkshop, Johns Hopkins University, Baltimore, MD.
- Rinaldi, Fabio, James Dowdall, Kaarel Kaljurand, Michael Hess, and Diego Moll, 2003. Exploiting paraphrases in a question-answering system. In *Proceedings of the Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IWP2003)*. ACL 2003.