Interpreting Comparative Constructions in Biomedical Text

Marcelo Fiszman,[1] **Dina Demner-Fushman,**[2]
Francois M. Lang,[2] **Philip Goetz,**[2]
Thomas C. Rindflesch[2]
[1]University of Tennessee – GSM, Knoxville, TN 37920
mfiszman@utmck.edu
[2]Lister Hill National Center for Biomedical Communications
National Library of Medicine, Bethesda, MD 20894
{ddemner|goetzp|flang|trindflesch}@mail.nih.gov

Abstract

We propose a methodology using underspecified semantic interpretation to process comparative constructions in MEDLINE citations, concentrating on two structures that are prevalent in the research literature reporting on clinical trials for drug therapies. The method exploits an existing semantic processor, SemRep, which constructs predications based on the Unified Medical Language System. Results of a preliminary evaluation were recall of 70%, precision of 96%, and F-score of 81%. We discuss the generalization of the methodology to other entities such as therapeutic and diagnostic procedures. The available structures in computable format are potentially useful for interpreting outcome statements in MEDLINE citations.

1 Introduction

As natural language processing (NLP) is increasingly able to support advanced information management techniques for research in medicine and biology, it is being incrementally improved to provide extended coverage and more accurate results. In this paper, we discuss the extension of an existing semantic interpretation system to address comparative structures. These structures provide a way of explicating the characteristics of one entity in terms of a second, thereby enhancing the description of the first. This phenomenon is important in clinical research literature reporting the results of clinical trials.

In the abstracts of these reports, a treatment for some disease is typically discussed using two types of comparative structures. The first announces that the (primary) therapy focused on in the study will be compared to some other (secondary) therapy. A typical example is (1).

```
(1)   Lansoprazole compared with
ranitidine for the treatment of
nonerosive gastroesophageal reflux
disease.
```

An outcome statement (2) often appears near the end of the abstract, asserting results in terms of the relative merits of the primary therapy compared to the secondary.

```
(2)   Lansoprazole is more
effective than ranitidine in
patients with endoscopically
confirmed non-erosive reflux
esophagitis.
```

The processing of comparative expressions such as (1) and (2) was incorporated into an existing system, SemRep [Rindflesch and Fiszman, 2003; Rindflesch et al., 2005], which constructs semantic predications by mapping assertions in biomedical text to the Unified Medical Language System® (UMLS)® [Humphreys et al., 1998].

2 Background

2.1 Comparative structures in English

The range of comparative expressions in English is extensive and complex. Several linguistic studies have investigated their characteristics, with differing assumptions about syntax and semantics (for example [Ryan, 1981; Rayner and Banks, 1990; Staab and Hahn, 1997; Huddleston and Pullum, 2002]). Our study concentrates on

BioNLP 2007: Biological, translational, and clinical language processing, pages 137–144,
Prague, June 2007.

structures in which two drugs are compared with respect to a shared attribute (e.g. how well they treat some disease). An assessment of their relative merit in this regard is indicated by their positions on a scale. The compared terms are expressed as noun phrases, which can be considered to be conjoined. The shared characteristic focused on is expressed as a predicate outside the comparative structure. An adjective or noun is used to denote the scale, and words such as *than, as, with,* and *to* serve as cues to identify the compared terms, the scale, and the relative position of the terms on the scale.

The first type of structure we address (called comp1 and illustrated in (3)) merely asserts that the primary and secondary terms (in bold) are being compared. A possible cue for identifying these structures is a form of *compare*. A further characteristic is that the compared terms are separated by a conjunction, or a preposition, as in (3).

(3) To compare **misoprostol** with **dinoprostone** for cervical ripening and labor induction.

As shown in (4), a scale may be mentioned (*efficacy*); however, in this study, we only identify the compared terms in structures of this type.

(4) To compare the *efficacy* of **misoprostol** with **dinoprostone** for cervical ripening and labor induction.

In the more complex comparative expression we accommodate (called comp2), the relative ranking of two compared terms is indicated on a scale denoted by an adjective (e.g. *effective* in (5)). The relative position of the compared terms in scalar comparative structures of this type expresses either equality or inequality. Inequality is further divided into superiority, where the primary compared term is higher on the scale than the secondary, and inferiority, where the opposite is true. Cues associated with the adjective designating the scale signal these phenomena (e.g. *as* ADJ *as* in (5) for equality, ADJer *than* in (6) for superiority, and *less* ADJ *than* in (7) for inferiority).

(5) **Azithromycin** is as effective as **erythromycin estolate** for the treatment of pertussis in children.

(6) **Naproxen** is safer than **aspirin** in the treatment of the arthritis of rheumatic fever.

(7) **Sodium valproate** was significantly less effective than **prochlorperazine** in reducing pain or nausea.

In examples (3) through (7), the characteristic the compared drugs have in common is treatment of some disorder, for example *treatment of pertussis in children* in (5).

Few studies describe an implemented automatic analysis of comparatives; however, Friedman [Friedman, 1989] is a notable exception. Jindal and Liu [Jindal and Liu, 2006] use machine learning to identify some comparative structures, but do not provide a semantic interpretation. We exploit SemRep machinery to interpret the aspects of comparative structures just described.

2.2 SemRep

SemRep [Rindflesch and Fiszman, 2003; Rindflesch et al., 2005] recovers underspecified semantic propositions in biomedical text based on a partial syntactic analysis and structured domain knowledge from the UMLS. Several systems that extract entities and relations are under development in both the clinical and molecular biology domains. Examples of systems for clinical text are described in [Friedman et al., 1994], [Johnson et al., 1993], [Hahn et al., 2002], and [Christensen et al., 2002]. In molecular biology, examples include [Yen et al., 2006], [Chun et al., 2006], [Blaschke et al., 1999], [Leroy et al., 2003], [Rindflesch et al., 2005], [Friedman et al., 2001], and [Lussier et al., 2006].

During SemRep processing, a partial syntactic parse is produced that depends on lexical look-up in the SPECIALIST lexicon [McCray et al., 1994] and a part-of-speech tagger [Smith et al., 2004]. MetaMap [Aronson, 2001] then matches noun phrases to concepts in the Metathesaurus® and determines the semantic type for each concept. For example, the structure in (9), produced for (8), allows both syntactic and semantic information to be used in further SemRep processing that interprets semantic predications.

(8) Lansoprazole for the treatment of gastroesophageal reflux disease

```
(9)  [[head(noun(Lansoprazole),metaconc('lansoprazole':[phsu]))],[prep(for),det(the),head(noun(treatment))],[prep(of),mod(adj(gastroesophageal)),mod(noun(reflux)),head(noun(disease),metaconc('Gastroesophageal reflux disease':[dsyn]))]]
```

Predicates are derived from indicator rules that map syntactic phenomena (such as verbs and nominalizations) to relationships in the UMLS Semantic Network. Argument identification is guided by dependency grammar rules as well as constraints imposed by the Semantic Network. In processing (8), for example, an indicator rule links the nominalization *treatment* with the Semantic Network relation "Pharmacologic Substance TREATS Disease or Syndrome." Since the semantic types of the syntactic arguments identified for *treatment* in this sentence ('Pharmacologic Substance' for "lansoprazole" and 'Disease or Syndrome' for "Gastroesophageal reflux disease") match the corresponding semantic types in the relation from the Semantic Network, the predication in (10) is constructed, where subject and object are Metathesaurus concepts.

```
(10)  lansoprazole TREATS
Gastroesophageal reflux disease
```

3 Methods

3.1 Linguistic patterns

We extracted sentences for developing comparative processing from a set of some 10,000 MEDLINE citations reporting on the results of clinical trials, a rich source of comparative structures. In this sample, the most frequent patterns for comp1 (only announces that two terms are compared) and comp2 (includes a scale and positions on that scale) are given in (11) and (12). In the patterns, Term1 and Term2 refer to the primary and secondary compared terms, respectively. "{BE}" means that some form of *be* is optional, and slash indicates disjunction. These patterns served as guides for enhancing SemRep argument identification machinery but were not implemented as such. That is, they indicate necessary components but do not preclude intervening modifiers and qualifiers.

(11) **comp1:** Compared terms

C1: Term1 {BE} *compare with/to* Term2
C2: *compare* Term1 *with/to* Term2
C3: *compare* Term1 *and/versus* Term2
C4a: Term1 *comparison with/to* Term2
C4b: *comparison of* Term1 *with/to* Term2
C4c: *comparison of* Term1 *and/versus* Term2
C5 Term1 *versus* Term2

(12) **comp2:** Scalar patterns

S1: Term1 BE *as* ADJ *as* {BE} Term2
S2a: Term1 BE *more* ADJ *than* {BE} Term2
S2b: Term1 BE ADJer *than* {BE}Term2
S2c: Term1 BE *less* ADJ *than* {BE} Term2
S4: Term1 BE *superior to* Term2
S5: Term1 BE *inferior to* Term2

As with SemRep in general, the interpretation of comparative structures exploits underspecified syntactic structure enhanced with Metathesaurus concepts and semantic types. Semantic groups [McCray et al., 2001] from the Semantic Network are also available. For this project, we exploit the group Chemicals & Drugs, which contains such semantic types as 'Pharmacologic Substance', 'Antibiotic', and 'Immunologic Factor'. (The principles used here also apply to compared terms with semantic types from other semantic groups, such as 'Procedures'.) In the comp1 patterns, a form of *compare* acts as an indicator of a comparative predication. In comp2, the adjective serves that function. Other words appearing in the patterns cue the indicator word (in comp2) and help identify the compared terms (in both comp1 and comp2). The conjunction v*ersus* is special in that it cues the secondary compared term (Term2) in comp1, but may also indicate a comp1 structure in the absence of a form of *compare* (C5).

3.2 Interpreting comp1 patterns

When SemRep encounters a form of *compare*, it assumes a comp1 structure and looks to the right for the first noun phrase immediately preceded by *with, to, and*, or *versus*. If the head of this phrase is mapped to a concept having a semantic type in the group Chemicals & Drugs, it is marked as the secondary compared term. The algorithm then looks to the left of that term for a noun phrase having a semantic type also in the group Chemicals & Drugs, which becomes the primary compared term. When this processing is applied to (13), the semantic predication (14) is produced, in which the predicate is COMPARED_WITH; the first argument is the primary compared term and the other is the secondary. As noted earlier, although a scale is sometimes asserted in these structures (as in (13)), SemRep does not retrieve it. An assertion regarding position on the scale never appears in comp1 structures.

```
(13) To compare the efficacy and
tolerability of Hypericum
perforatum with imipramine in
patients with mild to moderate
depression.
```

```
(14) Hypericum perforatum
COMPARED_WITH Imipramine
```

SemRep considers noun phrases occurring immediately to the right and left of *versus* as being compared terms if their heads have been mapped to Metathesaurus concepts having semantic types belonging to the group Chemicals & Drugs. Such noun phrases are interpreted as part of a comp1 structure, even if a form of *compare* has not occurred. The predication (16) is derived from (15).

```
(15) Intravenous lorazepam versus
dimenhydrinate for treatment of
vertigo in the emergency
department: a randomized clinical
trial.
```

```
(16) Lorazepam COMPARED_WITH
Dimenhydrinate
```

SemRep treats compared terms as being coordinated. For example, this identification allows both "Lorazepam" and "Dimenhydrinate" to function as arguments of TREATS in (15). Consequently, in addition to (16), the predications in (17) are returned as the semantic interpretation of (15). Such processing is done for all comp1 and comp2 structures (although these results are not given for (13) and are not further discussed in this paper).

```
(17) Lorazepam TREATS Vertigo
     Dimenhydrinate TREATS
Vertigo
```

3.3 Interpreting comp2 patterns

In addition to identifying two compared terms when processing comp2 patterns, a scale must be named and the relative position of the terms on that scale indicated. The algorithm for finding compared terms in comp2 structures begins by locating one of the cues *as, than,* or *to* and then examines the next noun phrase to the right. If its head has been mapped to a concept with a semantic type in the group Chemicals & Drugs, it is marked as the secondary compared term. As in comp1, the algorithm then looks to the left for the first noun phrase having a head in the same semantic group, and that phrase is marked as the primary compared term.

To find the scale name, SemRep examines the secondary compared term and then locates the first adjective to its left. The nominalization of that adjective (as found in the SPECIALIST Lexicon) is designated as the scale and serves as an argument of the predicate SCALE in the interpretation. For adjectives *superior* and *inferior* (patterns S4 and S5 in (12)) the scale name is "goodness."

In determining relative position on the scale, equality is contrasted with inequality. If the adjective of the construction is immediately preceded by *as* (pattern S1 in (12) above), the two compared terms have the same position on the scale (equality), and are construed as arguments of a predication with predicate SAME_AS. In all other comp2 constructions, the compared terms are in a relationship of inequality. The primary compared term is considered higher on the scale unless the adjective is *inferior* or is preceded by *less*, in which case the secondary term is higher. The predicates HIGHER_THAN and LOWER_THAN are used to construct predications with the compared terms to interpret position on the scale. The equality construction in (18) is expressed as the predications in (19).

```
(18) Candesartan is as effective
as lisinopril once daily in
reducing blood pressure.
```

```
(19) Candesartan COMPARED_WITH
lisinopril
     SCALE:Effectiveness
     Candesartan SAME_AS
lisinopril
```

The superiority construction in (20) is expressed as the predications in (21).

```
(20) Losartan was more effective
than atenolol in reducing
cardiovascular morbidity and
mortality in patients with
hypertension, diabetes, and LVH.
```

```
(21) Losartan COMPARED_WITH
Atenolol
```

```
      SCALE:Effectiveness
      Losartan HIGHER_THAN
Atenolol
```

The inferiority construction in (22) is expressed as the predications in (23).

(22) **Morphine-6-glucoronide** was significantly less potent than **morphine** in producing pupil constriction.

```
(23) morphine-6-glucoronide
COMPARED_WITH Morphine
      SCALE:Potency
      morphine-6-glucoronide
LOWER_THAN Morphine
```

3.4 Accommodating negation

Negation in comparative structures affects the position of the compared terms on the scale, and is accommodated differently for equality and for inequality. When a scalar comparison of equality (pattern S1, *as* ADJ *as*) is negated, the primary term is lower on the scale than the secondary (rather than being at least equal). For example, in interpreting the negated equality construction in (24), SemRep produces (25).

(24) **Amoxicillin-clavulanate** was *not* as effective as **ciprofloxacin** for treating uncomplicated bladder infection in women.

```
(25) Amoxicillin-clavulanate
COMPARED_WITH Ciprofloxaci
      SCALE:Effectiveness
      Amoxicillin-clavulanate
LOWER_THAN Ciprofloxacin
```

For patterns of inequality, SemRep negates the predication indicating position on the scale. For example, the predications in (27) represent the negated superiority comparison in (26). Negation of inferiority comparatives (e.g. "X is not less effective than Y") is extremely rare in our sample.

(26) These data show that **celecoxib** is *not* better than diclofenac (P = 0.414) in terms of ulcer complications.

```
(27) celecoxib COMPARED_WITH
diclofenac
      SCALE:Goodness
      celecoxib NEG_HIGHER_THAN
diclofenac
```

3.5 Evaluation

To evaluate the effectiveness of the developed methods we created a test set of 300 sentences containing comparative structures. These were extracted by the second author (who did not participate in the development of the methodology) from 3000 MEDLINE citations published later in date than the citations used to develop the methodology. The citations were retrieved with a PubMed query specifying randomized controlled studies and comparative studies on drug therapy.

Sentences containing direct comparisons of the pharmacological actions of two drugs expressed in the target structures (comp1 and comp2) were extracted starting from the latest retrieved citation and continuing until 300 sentences with comparative structures had been examined. These were annotated with the PubMed ID of the citation, names of two drugs (COMPARED_WITH predication), the scale on which they are compared (SCALE), and the relative position of the primary drug with respect to the secondary (SAME_AS, HIGHER_THAN, or LOWER_THAN).

The test sentences were processed using SemRep and evaluated against the annotated test set. We then computed recall and precision in several ways: overall for all comparative structures, for comp1 structures only, and for comp2 structures only. To understand how the overall identification of comparatives is influenced by the components of the construction, we also computed recall and precision separately for drug names, scale, and position on scale (SAME_AS, HIGHER_THAN and LOWER_THAN taken together). Recall measures the proportion of manually annotated categories that have been correctly identified automatically. Precision measures what proportion of the automatically annotated categories is correct.

In addition, the overall identification of comparative structures was evaluated using the F-measure [Rijsbergen, 1979], which combines recall and precision. The F-measure was computed using macro-averaging and micro-averaging. Macro-averaging was computed over each category first and then averaged over the three categories (drug names, scale, and position on scale). This approach gives equal weight to each category. In micro-averaging (which gives an equal weight to the performance on each sentence) recall and precision were obtained by summing over all individual sentences. Because it is impossible to enumerate all entities and relations which are not drugs, scale, or position we did not use the classification error rate and other metrics that require computing of true negative values.

4 Results

Upon inspection of the SemRep processing results we noticed that the test set contained nine duplicates. In addition, four sentences were not processed for various technical reasons. We report the results for the remaining 287 sentences, which contain 288 comparative structures occurring in 168 MEDLINE citations. Seventy four citations contain 85 comp2 structures. The remaining 203 structures are comp1.

Correct identification of comparative structures of both types depends on two factors: 1) recognition of both drugs being compared, and 2) recognition of the presence of a comparative structure itself. In addition, correct identification of the comp2 structures depends on recognition of the scale on which the drugs are compared and the relative position of the drugs on the scale. Table 1 presents recall, precision, and F-score reflecting these factors.

Table 1. SemRep performance

Task	Recall	Precision	F-score
Overall	0.70	0.96	0.81
Drug extraction	0.69	0.96	0.81
Comp1	0.74	0.98	0.84
Comp2	0.62	0.92	0.74
Scale	0.62	1.00	0.77
Position on scale	0.62	0.98	0.76

We considered drug identification to be correct only if both drugs participating in the relationship were identified correctly. The recall results indicate that approximately 30% of the drugs and comparative structures of comp1, as well as 40% of comp2 structures, remain unrecognized; however, all components are identified with high precision. Macro-averaging over compared drug names, scale, and position on scale categories we achieve an F-score = 0.78. The micro-average score for 287 comparative sentences is 0.5.

5 Discussion

In examining SemRep errors, we determined that more than 60% of the false negatives (for both comp1 and comp2) were due to “empty heads” [Chodorow et al., 1985; Guthrie et al., 1990], in which the syntactic head of a noun phrase does not reflect semantic thrust. Such heads prevent SemRep from accurately determining the semantic type and group of the noun phrase. In our sample, expressions interpreted as empty heads include those referring to drug dosage and formulations, such as *extended release* (the latter often abbreviated as *XR*). Examples of missed interpretations are in sentences (28) and (29), where the empty heads are in bold. Ahlers et al. [Ahlers et al., 2007] discuss enhancements to SemRep for accommodating empty heads. These mechanisms are being incorporated into the processing for comparative structures.

(28) Oxybutynin **15 mg** was more effective than propiverine **20 mg** in reducing symptomatic and asymptomatic IDCs in ambulatory patients.

(29) Intravesical atropine was as effective as oxybutynin **immediate release** for increasing bladder capacity and it was probably better with less antimuscarinic side effects

False positives were due exclusively to word sense ambiguity. For example, in (30) *bid* (twice a day) was mapped to the concept “BID protein”, which belongs to the semantic group Chemicals & Drugs. The most recent version of MetaMap, which will soon be called by comparative processing, exploits word sense disambiguation [Humphrey et al., 2006] and will likely resolve some of these errors.

(30) Retapamulin ointment 1% (**bid**) for 5 days was as effective as oral cephalexin (**bid**) for 10 days in treatment of patients with SID, and was well tolerated.

Although, in this paper, we tested the method on structures in which the compared terms belong to the semantic group Chemicals & Drugs, we can straightforwardly generalize the method by adding other semantic groups to the algorithm. For example, if SemRep recognized the noun phrases in bold in (31) and (32) as belonging to the group Procedures, comparative processing could proceed as for Chemicals & Drugs.

(31) Comparison of **multi-slice spiral CT** and **magnetic resonance** imaging in evaluation of the un-resectability of blood vessels in pancreatic tumor.

(32) Dynamic **multi-slice spiral CT** is better than dynamic **magnetic resonance** to some extent in evaluating the un-resectability of peripancreatic blood vessels in pancreatic tumor.

The semantic predications returned by SemRep to represent comparative expressions can be considered a type of executable knowledge that supports reasoning. Since the arguments in these predications have been mapped to the UMLS, a structured knowledge source, they can be manipulated using that knowledge. It is also possible to compute the transitive closure of all SemRep output for a collection of texts to determine which drug was asserted in that collection to be the best with respect to some characteristic. This ability could be very useful in supporting question-answering applications.

As noted earlier, it is common in reporting on the results of randomized clinical trials and systematic reviews that a comp1 structure appears early in the discourse to announce the objectives of the study and that a comp2 structure often appears near the end to give the results. Another example of this phenomenon appears in (33) and (34) (from PMID 15943841).

(33) To compare the efficacy of famotidine and omeprazole in Japanese patients with non-erosive gastro-oesophageal reflux disease by a prospective randomized multicentre trial.

(34) Omeprazole is more effective than famotidine for the control of gastro-oesophageal reflux disease symptoms in H. pylori-negative patients.

We suggest one example of an application that can benefit from the information provided by the knowledge inherent in the semantic interpretation of comparative structures, and that is the interpretation of outcome statements in MEDLINE citations, as a method for supporting automatic access to the latest results from clinical trials research.

6 Conclusion

We expanded a symbolic semantic interpreter to identify comparative constructions in biomedical text. The method relies on underspecified syntactic analysis and domain knowledge from the UMLS. We identify two compared terms and scalar comparative structures in MEDLINE citations. Although we restricted the method to comparisons of drug therapies, the method can be easily generalized to other entities such as diagnostic and therapeutic procedures. The availability of this information in computable format can support the identification of outcome sentences in MEDLINE, which in turn supports translation of biomedical research into improvements in quality of patient care.

Acknowledgement This study was supported in part by the Intramural Research Programs of the National Institutes of Health, National Library of Medicine.

References

Ahlers C, Fiszman M, Demner-Fushman D, Lang F, Rindflesch TC. 2007. Extracting semantic predications from MEDLINE citations for pharmacogenomics. *Pacific Symposium on Biocomputing* 12:209-220.

Aronson AR. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. *Proc AMIA Symp*, 17-21.

Blaschke C, Andrade MA, Ouzounis C, and Valencia A. 1999. Automatic extraction of biological information from scientific text: protein-protein interactions. *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology*. Morgan Kaufman Publishers, San Francisco, CA.

Christensen L, Haug PJ, and Fiszman M. 2002. MPLUS: A probabilistic medical language understanding system. *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain, Association for Computational Linguistics,* 29-36.

Chodorow MS, Byrd RI, and Heidom GE. 1985. Extracting Semantic Hierarchies from a Large On-

Line Dictionary. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, 299-304.

Chun HW, Tsuruoka Y, Kim J-D, Shiba R, Nagata N, Hishiki T, and Tsujii J. 2006, Extraction of gene-disease relations from Medline using domain dictionaries and machine learning. *Pac Symp Biocomput*, 4-15.

Friedman C. 1989. A general computational treatment of the comparative. *Proc 27th Annual Meeting Assoc Comp Linguistics*, 161-168.

Friedman C, Alderson PO, Austin JH, Cimino JJ, and Johnson SB. 1994. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc*, 1(2):161-74.

Friedman C, Kra P, Yu H, Krauthammer M, and Rzhetsky A. 2001. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics*, 17 Suppl 1:S74-S82.

Guthrie L, Slater BM, Wilks Y, Bruce R. 1990. Is there content in empty heads? *Proceedings of the 13th Conference on Computational Linguistics*, v3:138 – 143.

Hahn U, Romacker M, and Schulz S. 2002. MEDSYNDIKATE--a natural language system for the extraction of medical information from findings reports. *Int J Med Inf*, 67(1-3):63-74.

Huddleston R, and Pullum GK. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, UK.

Humphrey SM, Rogers WJ, Kilicoglu H, Demner-Fushman D, Rindflesch TC. 2006. Word sense disambiguation by selecting the best semantic type based on Journal Descriptor Indexing: Preliminary experiment. *J Am Soc Inf SciTech* 57(1):96-113.

Humphreys BL, Lindberg DA, Schoolman HM, and Barnett OG. 1998. The Unified Medical Language System: An informatics research collaboration. *J Am Med Inform Assoc*, 5(1):1-11.

Jindal, Nitin and Bing Liu. 2006. Identifying comparative sentences in text documents. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research & Development on Information Retrieval*.

Johnson SB, Aguirre A, Peng P, and Cimino J. 1993. Interpreting natural language queries using the UMLS. Pro*c Annu Symp Comput Appl Med Care*, 294-8.

Leroy G, Chen H, and Martinez JD. 2003 A shallow parser based on closed-class words to capture relations in biomedical text. *J Biomed Inform,* 36(3):145-158.

Lussier YA, Borlawsky T, Rappaport D, Liu Y, and Friedman C. 2006 PhenoGO: assigning phenotypic context to Gene Ontology annotations with natural language processing. *Pac Symp Biocomput,* 64-75.

McCray AT, Srinivasan S, and Browne AC. 1994. Lexical methods for managing variation in biomedical terminologies. *Proc Annu Symp Comput Appl Med Care*, 235-9.

McCray AT, Burgun A, and Bodenreider O. 2001 Aggregating UMLS semantic types for reducing conceptual complexity. *Medinfo*, 10(Pt 1): 216-20.

Rayner M and Banks A. 1990. An implementable semantics for comparative constructions. *Computational Linguistics,* 16(2):86-112.

Rindflesch TC. 1995. Integrating natural language processing and biomedical domain knowledge for increased information retrieval effectiveness. *Proc 5th Annual Dual-use Technologies and Applications Conference,* 260-5.

Rindflesch TC and Fiszman M. 2003. The interaction of domain knowledge and linguistic structure in natural language processing: Interpreting hypernymic propositions in biomedical text. *J Biomed Inform,* 36(6):462-77.

Rindflesch TC, Marcelo Fiszman , and Bisharah Libbus. 2005. Semantic interpretation for the biomedical research literature. *Medical informatics: Knowledge management and data mining in biomedicine*. Springer, New York, NY.

Rijsbergen V. 1979. *Information Retrieval*, Butterworth-Heinemann, Newton, MA.

Ryan K. 1981. Corepresentational grammar and parsing English comparatives. *Proc 19th Annual Meeting Assoc Comp Linguistics,* 13-18.

Smith L, Rindflesch T, and Wilbur WJ. 2004. MedPost: a part-of-speech tagger for biomedical text. *Bioinformatics*, 20(14):2320-1.

Staab S and Hahn U. Comparatives in context. 1997. *Proc 14th National Conference on Artificial Intelligence and 9th Innovative Applications of Artificial Intelligence Conference*, 616-621.

Yen YT, Chen B, Chiu HW, Lee YC, Li YC, and Hsu CY. 2006. Developing an NLP and IR-based algorithm for analyzing gene-disease relationships.