

# Identifying Comparative Structures in Biomedical Text

Samir Gupta<sup>1</sup>, A. S. M. Ashique Mahmood<sup>1</sup>, Karen E. Ross<sup>2</sup>, Cathy H. Wu<sup>1,3</sup>, K. Vijay-Shanker<sup>1</sup>

<sup>1</sup>Department of Computer & Information Sciences

<sup>3</sup>Center for Bioinformatics and Computational Biology

University of Delaware, Newark, DE

{sgupta, ashique, wuc, vijay}@udel.edu

<sup>2</sup>Department of Biochemistry and Molecular & Cellular Biology

Georgetown University Medical Center, Washington, DC

ker25@georgetown.edu

## Abstract

Comparison sentences are very commonly used by authors in biomedical literature to report results of experiments. In such comparisons, authors typically make observations under two different scenarios. In this paper, we present a system to automatically identify such comparative sentences and their components i.e. the compared entities, the scale of the comparison and the aspect on which the entities are being compared. Our methodology is based on dependencies obtained by applying a parser to extract a wide range of comparison structures. We evaluated our system for its effectiveness in identifying comparisons and their components. The system achieved a F-score of 0.87 for comparison sentence identification and 0.77-0.81 for identifying its components.

## 1 Introduction

Biomedical researchers conduct experiments to validate their hypotheses and infer associations between biological concepts and entities, such as mutation and disease or therapy and outcome. It is often not enough to simply report the effects of an intervention; instead, the most common way to validate such observations is to perform comparisons. In such studies, researchers make observations under two different scenarios (e.g., disease sample vs. control sample). When the differences between the groups are statistically significant, association can be inferred.

Comparative studies are prevalent in nearly every field of biomedical/clinical research. For example, in the experimental approach known as “reverse genetics”, researchers draw inferences about gene function by comparing the pheno-

type of a gene knockdown sample to that of a sample expressing the gene at the normal level. In clinical trial studies, researchers study the effectiveness or side-effects of a drug compared to a placebo. A simple PubMed query “compared[TIAB] OR than[TIAB] OR versus[TIAB]” returned 3,149,702 citations, which provides a rough estimate of the pervasive nature of comparisons in the biomedical literature. Thus, development of automated techniques to identify such statements would be highly useful.

Comparative sentences typically contain two (or more) entities, which are being compared with respect to some common aspect. Consider sentence (1), which compares gene expression level in cancerous vs. non-cancerous tissues:

- (1) The expression of GPC5 gene was lower in lung cancer tissues compared with adjacent noncancerous tissues.

Typically, the entities, which we will refer as **compared entities**, are of the same type. In the example, the entities being compared are two tissues: “lung cancer tissues” and “adjacent noncancerous tissues”, which are separated by the phrase “compared with”. “Expression of GPC5 gene”, which we call the **compared aspect**, is the aspect on which comparison between the two entities is being made. The word “lower” indicates the scale of the comparison, thereby providing an ordering of the compared entities with respect to the compared aspect. These definitions are similar to those described in (Park and Blake, 2012).

In this paper, we describe a system to automatically identify comparative structures from text. We have developed patterns based on sentence syntactic dependency information to identify comparison sentences and also extract the various components (compared aspect, compared entities

and scale). The developed system identifies explicit comparative structures at the sentence level, where all the components of the comparison are present in the sentence. The main challenge is to capture patterns at a sufficiently high level given the sheer variety of comparative structures. In the rest of the paper we will define the task, describe our approach and comparison patterns and present the results of our evaluation. We achieved a F-score of 0.87 for identifying comparison sentences and 0.78, 0.81, 0.77 for extracting the compared aspect, scale indicator and compared entities, respectively. Thus the major contributions of this work are:

- Development of a general approach for identifying comparison sentences using syntactic dependencies.
- Development of methods to extract all of the components of the comparative structure.

## 2 Related Works

The sentence constructions used to make comparisons in English are complex and variable. [Bresnan \(1973\)](#) discussed the syntax of comparative clause construction in English and noted its syntactic complexity, ‘exhibiting a variety of grammatical processes’. [Friedman \(1989\)](#) reported a general treatment of comparative structures based on basic linguistic principles and noted that automatically identifying them is computationally difficult. They also noted that comparative structures resemble and can be transformed into other syntactic forms such as general coordinate conjunctions, relative clauses, and certain subordinate and adverbial clauses and thus ‘syntactically the comparative is extraordinarily diverse’. In ([Staab and Hahn, 1997](#)), the authors proposed a model of comparative interpretation that abstracts from textual variations using descriptive logic representation.

The above studies provide an analysis of comparative sentences from a linguistic point of view. Computational systems for identifying comparisons have also been developed. [Jindal and Liu \(2006a\)](#) proposed a machine learning approach to identify comparative sentences from text. The system first categorizes comparative sentences into different types, and then presents a pattern discovery and supervised learning approach to classify each sentence into two classes: comparative and non-comparative. Class sequential rules

based on words and part-of-speech tags automatically generated while learning the model were used as features in this work. The authors evaluated their classifier on product review sentences containing comparison between products and reported a precision of 79% and a recall of 81%. The authors extended their work ([Jindal and Liu, 2006b](#)) to extract comparative relations i.e. the compared entities and their features, and comparison keywords from the identified comparison sentences. In ([Xu et al., 2011](#)), the authors described a machine learning approach to extract and visualize comparative relations between products from Amazon customer reviews. They describe a comparative relation as a 4-tuple, containing the two compared products, the compared aspect and a comparison direction (better, worse, same). They reported a F-score of 38.81% using multi-class SVM and 56.68% using Conditional Random Fields (CRF). ([Jindal and Liu, 2006b](#); [Xu et al., 2011](#)) are the only works that extract the different components of the comparison. In ([Ganapathibhotla and Liu, 2008](#)), the authors focused on mining opinions from comparative sentences from product review sentences and extracting the preferred product. [Yang and Ko \(2009\)](#) proposed a machine learning approach to identify comparative sentences from Korean web-based text but did not address the extraction of the comparison arguments. They first constructed a set of comparative keywords manually and extracted candidate comparative sentences and then used Maximum Entropy Model (MEM) and Naive Bayes (NB) to eliminate non-comparative sentences from the candidates.

Relatively few works on identifying comparative sentences and/or its components from biomedical text have been developed. [Park and Blake \(2012\)](#) reported a machine learning approach to identify comparative claims automatically from full-text scientific articles. They introduced a set of semantic and syntactic features for classifications using three different classifiers: Naive Bayes (NB), a Support Vector Machine (SVM) and a Bayesian network (BN). They evaluated their approach on full-text toxicology articles and achieved F1 score of 0.76, 0.65, and 0.74 on a validation set for the NB, SVM and BN, respectively. The focus of this work was on identifying comparison sentences and the extraction of its components was not addressed.

Fiszman et al. (2007) described a technique to identify comparative constructions in MEDLINE citations using under-specified semantic interpretation. The authors used textual patterns combined with semantic predications extracted from the semantic processor SemRep (Rindflesch and Fiszman, 2003; Rindflesch et al., 2005). The predications extracted by SemRep are based on the Unified Medical Language System (UMLS) (Humphreys et al., 1998). Their system extracts the compared entities (limited to drugs) and the scale of the comparison. They reported an average F-score of 0.78 for identifying the compared drug names, scale and scale position. To the best of our knowledge, (Fiszman et al., 2007) is the only reported work that goes beyond identification of comparison sentences to identify the different components of the comparison in biomedical text. But unlike our work, theirs is limited to comparison between drugs, does not extract the comparison aspect and appears to be limited in their coverage of comparison structures.

### 3 Method

#### 3.1 Task Definition

Basic comparison sentences contain two or more *compared entities (CE)* and a *comparison aspect (CA)* on which compared entities are being compared. Additionally, there are two parts in such sentences indicating the comparison. The first is the presence of a word that indicates the scale of the comparison and the other separates the two compared entities. The former is often comparative adjectives or adverbs (such as “higher”, “lower”, “better”, etc.), while the latter can be expressed with phrases or words (such as “than”, “compared with”, “versus” etc.). We will refer to the former comparative word indicating the scale as the *Scale Indicator (SI)* and the latter, separating the entities, as the *Entity Separator (ES)*. In example (2) below the key parts of such a comparison structure are highlighted.

- (2) [Arteriolar sclerosis]<sub>CA</sub> was significantly higher<sub>SI</sub> in addicts<sub>CE</sub> than<sub>ES</sub> controls<sub>CE</sub>.

Jindal and Liu (2006b) categorized comparative structures into four classes: (1) Non-Equal Gradable, (2) Equative, (3) Superlative and (4) Non-Gradable. *Non-Equal Gradable* comparison indicate relations of the type greater or less than, providing an ordering of the compared entities. *Equa-*

*tive* structures indicate equal relation between the two entities with respect to the aspect. Comparisons where one entity is “better” than all other entities are termed as *Superlative*. Sentences in which the compared entities are not explicitly graded are called *Non-Gradable*.

Based on our previous discussion, we will be addressing only the first two types: **Non-Equal Gradable** and **Equative** comparison. First, we consider processing at the sentence-level only. While there are cases of comparisons, where the context provided by a larger body of text might provide the information about all the components, they are not considered in this work. Thus most of the superlative cases will not be considered because all the compared entities are rarely mentioned within a single sentence. It also rules out cases such as in Example (3a), where the second compared entity must be inferred from previous sentences. Second, we consider only those sentences where the authors mention the result or conclusion of an experiment/study. Thus, we will not consider sentences such as in Example (3b), since it only mentions the intention to perform a comparison but does not indicate the result of the experiment. While such sentences can still be captured with minor changes to our existing patterns, our goal here is to only consider sentences that indicate the results of experiment by means of comparison. The patterns developed in this work identify explicit comparative structures at the sentence level and extract all components of the comparison relations, i.e., the compared aspect, entities and the scale indicator.

- (3) a. Mean procedure time was significantly shorter for the percutaneous procedure.  
 b. We compared lesion growth between placebo and tissue plasminogen activator-treated patients.

#### 3.2 Approach

The different steps of our system are depicted in Figure 1. Given an input text, typically a Medline abstract, we first tokenize and split the text into sentences using the Stanford CoreNLP toolkit (Manning et al., 2014). We then use the Charniak-Johnson parser (Charniak, 2000; Charniak and Johnson, 2005) with David McClosky’s adaptation to the biomedical domain (McClosky, 2010) to obtain constituency parse trees for each sentence. Next we use the Stanford conversion

tool (Manning et al., 2014; De Marneffe et al., 2014) to convert the parse tree into the syntactic dependency graph (SDG). We use the “CCProcessed” representation, which collapses and propagates dependencies allowing for an appropriate treatment of sentences that involve conjunctions. Note that “CCProcessed” is helpful as dependencies involving preposition, conjuncts, as well as referent of relative clauses are “collapsed” to get direct dependencies between context words. Thus, as seen in Figure 2, which shows the “CCProcessed” SDG, there is a direct edge from “lower” to the cells in the Noun Phrase (NP) “Hep3B cells” rather than a path with two edges where the first reaches the preposition “in” and the second from “in” word to the word “cells”. This simplifies the pattern development in relation extraction.

Based on this syntactic dependencies representation, we have developed patterns to identify the different arguments of the comparison relation. Next we use Semgrep, which is a part of the Stanford NLP Toolkit, to specify the translated patterns as regular expressions based on lemmas, part-of-speech tags, and dependency labels, which will automatically match with the sentence dependency parse structure. We have developed a total of 35 and 8 patterns to identify Non-Equal Gradable and Equative comparisons respectively. The developed Semgrep rules as well as the evaluation test set can be found at the link below<sup>1</sup>. Each Semgrep rule/pattern identifies all components of the comparison, specifically the head of the comparison aspect, entities and scale. Since the components are typically Noun Phrases (NPs), we look at the outgoing edges from the head nouns to obtain the NPs corresponding to the comparison components. In the next subsection, we will discuss the development of different comparison patterns.

### 3.3 Comparative Patterns

As discussed earlier in subsection 3.1, the two key parts in a basic comparison sentence are a Scale Indicator (SI), indicating the scale of the comparison and a Entity Separator (ES), separating the compared entities. We will use dependencies from these SI and ES words to extract the compared aspect and the compared entities. We have developed rules based on syntactic dependencies for various combinations of the two keys parts. We broadly categorize our comparison patterns based

<sup>1</sup><http://biotm.cis.udel.edu/biotm/projects/comparison/>

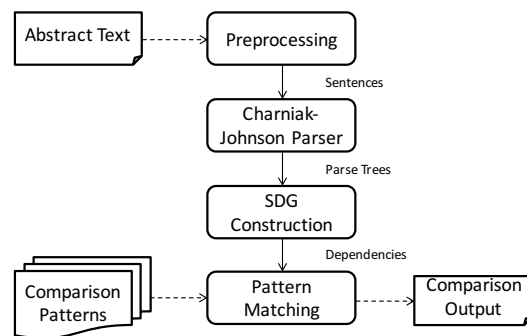


Figure 1: Comparison Pipeline.

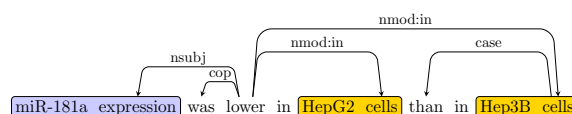


Figure 2: Example SDG

on the Scale Indicator word indicating either Non-Equal Gradable or Equative Comparison.

#### 3.3.1 Non-Equal Gradable

Non-Equal Gradable comparison indicates a difference between the compared entities. Based on three part-of-speech tags (POS) of the Scale Indicator, different syntactic structures are possible, as described below. Note that in all the figures depicting the dependency graph the compared aspect is highlighted in blue and the compared entities in yellow.

**Comparative Adjective:** Starting with the most frequent case for Scale Indicator, which is a *comparative adjective* (JJR) such as “better”, “higher”, “lower” etc., there are two broad categories of syntactic structures which we consider. The **first** category involves copular structures, where the JJR serves as the predicate of the comparison relation. The compared aspect is typically the subject of the JJR as shown in Figure 3a. Thus we follow the *nsubj* edge from the JJR to get the head of compared aspect. We use the *nmod:than* from JJR to extract one of the compared entities. The second entity will also have an edge from the JJR, which can be prepositional edge (*nmod:in* as in Figure 3a). Thus we use *nmod* edges from the predicate JJR to determine the second compared entity. Note all prepositional edges such as “with”, “for”, “during” etc. are considered. Additionally, the second compared entity will be separated by an Entity Separator (“than” in this case) from the first com-

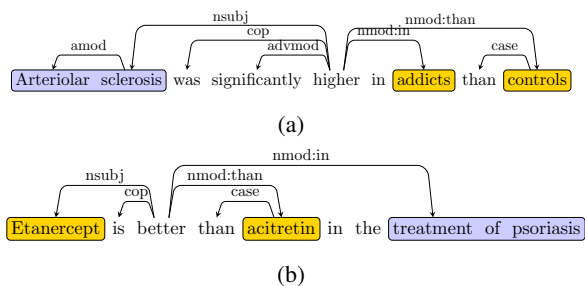


Figure 3: Comparative Adjective copular forms.

pared entity. Thus we further verify that the extracted compared entities are separated by an ES.

The position of the entity separator “than” is critical for determining the second compared entity as well as the first compared entity. As shown in Figure 3b, despite the similar copular structure to the sentence in Figure 3a, the subject of the JJR (“better” in this case) is the compared entity rather than the aspect. This is due to the fact that the JJR is followed by the ES “than”. Thus ordering of the words is an important clue when differentiating between these cases.

The **second** category involves sentences, where the comparative adjective modifies a head noun and this modified noun provides the compared aspect, as shown in Figures 4 and 5. Since the compared aspect is modified by the JJR, we used the *amod* edge to detect the aspect. In these cases, the noun phrase containing the Scale Indicator will be connected to a verb and typically serves as the predicate of the comparison relation. The entity separator in the sentence in Figure 4 is “compared to” and we can extract one of the compared entities (“intravenous morphine”) by following the *advcl:compared\_to* edge from the predicate verb (“offers”).

Note that in the first example (Figure 4), the Verb Group (“offer”) is in the active form and in the second example (Figure 5), it is in the passive form (“was observed in”). Due to the active/passive form difference, the aspect is in the object position and one of the compared entities in the subject position in the first example, while the reverse is true for the second example. In the dependency representation, the *nsubj* edge and the *nmod:in* edge provide the subjects in active and passive cases and *dobj* and *nsubjpass* provide the possible objects. Note that in certain cases, the author might use an adjective (JJ) instead of the comparative form (“high” instead of “higher”). We

treat such cases in the same way we treat the comparative adjective (JJR) form.

Note that the Semgrep patterns only identifies the head words of the various components, which are typically NPs. We follow outgoing dependency edges from these head words to extract phrases corresponding to each comparison component. For example, in Figure 3a “sclerosis” is identified as the aspect head and we follow the edge *amod* to extract the aspect phrase “Arteriolar sclerosis”. In Figure 5, we extract “TP expression” as the aspect phrase and not “Higher TP expression” as “higher” is the trigger of the comparison and identified as the scale.

**Comparative Adverb:** In these sentences, the comparison scale is indicated through comparative adverbs (RBR) such as “more”, “less” etc.. Typically, the RBR modifies an adjective (JJ) as shown in Figure 6, where the adjective is “effective”. This adjective serves as the predicate of the comparison and dependency edges from it are used to determine the aspect and entities. The syntactic structure and our rules are very similar to the first category of the Comparative Adjective case. Thus we use the *nsubj* and *advcl:compared\_to* edges from “effective” to determine the compared entities. Note that the compared aspect in this example is a clause headed by a VBG (“reducing MCP-1 levels”) and thus in addition to *nmod* edges, we need to consider the adverbial clause modifier (*advcl*) edge to determine the aspect.

**Verbs:** Certain verbs such as “increased”, “decreased” as well as “improved” indicate differences and can be used as a SI. This verb serves as the predicate of the comparison relation and outgoing dependencies can be used to determine the arguments of the comparison. We have observed two categories based on the voice (**passive** vs. **active**) of the Verb Group containing this verb. The passive case is depicted in Figure 7a (“was increased in”). In this case, we follow the *nsubjpass* edge to determine the compared aspect. In Figure 7b, since the scale indicator “improved” is in active voice, the direct object of the verb will instead provide the aspect. Extraction and verification of the compared entities is similar to the cases described previously (e.g. *nmod:in* in Figure 7a; *dobj* and *advcl:compared\_with* in Figure 7b).

Note that a verb in past participle tense (VBN) can be used as an adjective and modify a noun (e.g., *Increased* TP expression was found in ...).

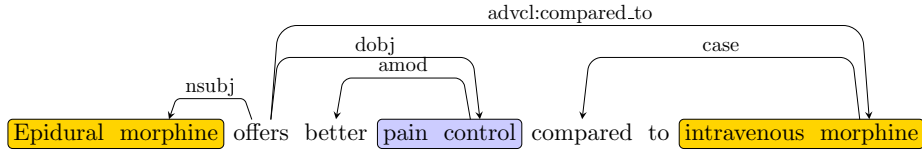


Figure 4: Comparative Adjective modifier form 1.

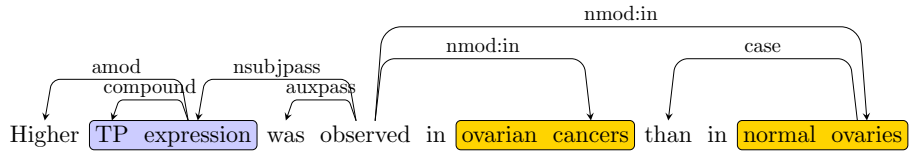


Figure 5: Comparative Adjective modifier form 2.

We treat cases when the scale indicator verb is used as a modifier of a NP like the second category of Comparative Adjectives.

### 3.3.2 Equative

A sentence with Equative comparison corresponds to cases, where the result of comparison indicates no difference between the compared entities (as in Figure 8). In these cases, it is very rare to find the usual Entity Separator (ES) and instead words such as conjunctions (“and”, “or”), “between” and “among” play the role of the ES. We have observed three frequently occurring types of such Equative comparative structures.

The **first** category involves the structure “**X as JJ as Y**”, where JJ is an adjective. In these cases, the adjective serves as the predicate of the comparison. Figure 8 depicts such a case, where the adjective is “effective”. Here one of the compared entity “botox” is the subject of the JJ “effective”. The second compared entity “oral medication” is preceded by the ES “as” and a *nmod:as* edge from the JJ to the entity is present. The compared aspect is typically attached to the second compared entity through a *nmod* edge (*nmod:for* in this case). Note that the ES “as” need not appear immediately after the JJ (e.g. “Botox is as effective for overactive bladder as oral medication”). Due to the “CCProcessed” representation of collapsing edges we can still consider the *nmod:as* from “effective” to determine the second compared entity. The only difference in this case is that the *nmod:for* edge used to determine the aspect is from the predicate “effective”.

The **second** case involves the Scale Indicator phrase “**similar to**” as shown in Figure 9. Here the subject of the adjective “similar” is the compared

aspect. The *nmod* edges (*nmod:in* in this example) from “similar” are used to determine the compared entities. The entities in these cases are separated through conjunctions. Note that the SI “similar” can also modify the compared aspect (e.g. “Similar CA was observed in CE1 and CE2”). This case closely resembles the second category of comparative adjectives and similar rules are used.

The **third** category involves Scale Indicator phrases such “**no differences**”, “**no changes**” etc. Similar to the case of the second category comparative adjectives, here the SI “difference” is part of a NP and hence is connected to a verb, which serves as the predicate. Typically these verbs can be “linking” verbs (“is”, “was” etc.) in active form or certain verbs indicating presence (“found in”, “noted in”, “observed in”) in the passive form. In active voice case, as shown in Figure 10, the SI typically follows an existential such as “there”. In these cases, the *nmod:between* from the predicate verb (“was” in this case) is used to determine the compared entities. Other *nmod* edges we consider are *nmod:among* and *nmod:in*. The compared aspect is attached to the second compared entity though *nmod* edges (*nmod:for* in this example). A large proportion of Equative structures do not mention the compared entities explicitly, and as per the definition of our task, we do not extract the comparison components in these cases.

## 4 Evaluation

We evaluated our system for its effectiveness in identifying comparative sentences and its components on a test set of 189 comparisons from 125 abstracts annotated by a co-author, who was not involved in the design and development of the sys-

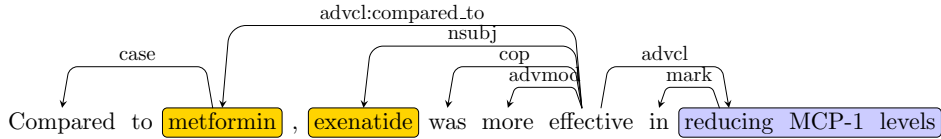


Figure 6: Comparative Adverb copular form.

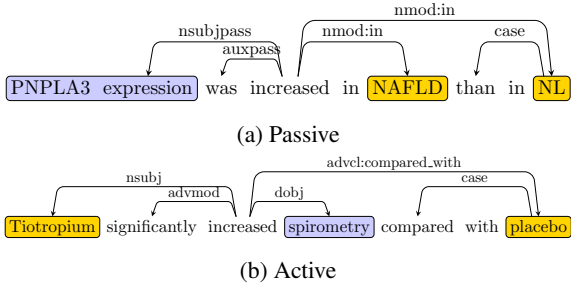


Figure 7: Comparative Verb forms.

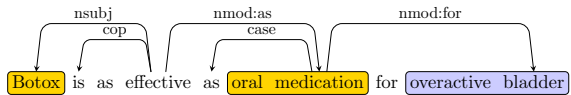


Figure 8: Equative Form 1.

tem. Note that the annotator also annotated an additional 50 abstracts, which was used in the development of the comparison patterns. Although the work by Fiszman et al. (2007) attempts to tackle the similar task of identifying comparison sentences and its components, we do not directly compare with their results. This is due to the fact that their implementation is limited to “direct comparisons of the pharmacological actions of two drugs”. We ran their system on our annotated test data and only 8 out of the 189 comparisons were identified by their system as their implementation only detects comparison if the two compared entities (CEs) are drugs. We also ran their system on some artificially created sentences obtained by replacing CEs with drugs and observed that their system seemed limited in the coverage of comparison structures. In the subsequent sections, we will describe the evaluation methodology, present the results and provide an analysis of errors.

#### 4.1 Experimental Setup

To evaluate our system’s performance, we have created a test set of 125 abstracts. We selected abstracts that usually draw conclusions by means of comparing between two contrasting situations. Randomized controlled trials (RCT), which compare the outcome between two randomly selected

groups, fit this definition very well. For this reason, we searched for RCTs in PubMed with the query “(Randomized Controlled Trial[Publication Type]). This query yielded 431,226 abstracts. However, we noticed that this set lacked abstracts concerning gene expression studies. Thus, we appended to our initial dataset with abstracts related to the effect of differential expression of genes on diseases. As we target to identify comparison sentences, we chose abstracts tagged as “comparative study” in PubMed because they tend to contain comparisons. We used the PubMed query: “(Comparative Study [Publication Type]) AND expression[TIAB] AND (cancer[TI] OR carcinoma[TI])”, restricting the comparative studies to gene expressions and cancer related studies. This query yielded 8,479 abstracts.

From this initial set of abstracts, we randomly selected 125 abstracts for annotation by a biomedical researcher expert who did not take part in the development of the system. 150 sentences from the 125 abstracts were annotated as comparison sentences and included 189 comparisons. Our guidelines required the annotation of the four components for each comparison: the compared aspect (CA), the two compared entities (CE1 and CE2) and a word or phrase that indicates the scale of comparison (SI). Additionally, they (the guidelines) required annotation at a sentence level for sentences which had an explicit conclusion i.e. indicated the scale of comparison and is not a mention of a planned investigation.

#### 4.2 Results and Discussion

Annotations of the test set of 125 abstracts yielded 189 comparisons, each containing a compared aspect, a scale indicator and two compared entities. We ran our system on the test set and evaluated its performance on correctly identifying the (1) comparison sentences, (2) compared aspect, (3) scale indicator and (4) compared entities. When computing true positives, we compared the head word of the annotated components with the head words extracted by our system. A mismatch resulted

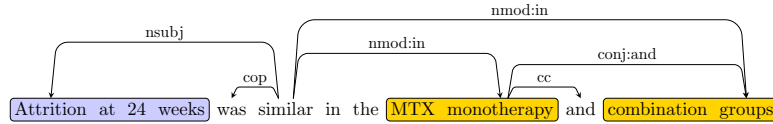


Figure 9: Equative Form 2.

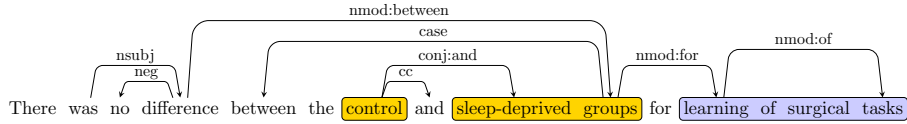


Figure 10: Equative Form 3.

Table 1: Evaluation Results.

Type	Precision	Recall	F-Score
<b>Sentence</b>	0.91	0.83	0.87
<b>Comparison Aspect</b>	0.85	0.72	0.78
<b>Scale Indicator</b>	0.87	0.75	0.81
<b>Compared Entities</b>	0.84	0.72	0.77

in both a false negative and false positive. We computed Precision (P), Recall (R), and F-score (F) measures for each evaluation type, results of which are shown in Table 1.

We analyzed the errors made by our system and majority of the errors (more than 80%) encountered were due to incorrect parsing of complicated sentences. For example, in sentence (4), the clause modifier edge *acl* to “compared” was from “feed” instead of the aspect “palatable”. If the clause “with significantly less consumption of treated feed” is removed, thereby simplifying the sentence, the parse is correct and we correctly extract the comparison.

- (4) Pro-Dynam was significantly less palatable, with significantly less consumption of treated feed compared with either Equipalazone Powder or Danilon Equidos

A second but rarer category of error involves cases, where we did not consider certain Scale Indicators (SI) such as “superior”, “non-inferior”, “extra” as in sentence (5). In such examples, the parser tagged the SI as adjective (JJ) and not a comparative adjective (JJR) even though these words indicate a comparison. Since our treatment of such patterns was limited to JJR scale indicators, we missed these cases. It is important to note that our system will identify such structures if we replace such JJ scale indicators by a JJR.

- (5) Moxifloxacin was non-inferior to ceftriaxone/metronidazole in terms of clinical response at test-of-cure in the PP population

The third category involved cases missed due to missing patterns such as seen in sentences (6). In sentence (6a), two set of patients are being compared with respect to improvement extent, while sentence (6b) compares the concentration of “plasma F2-isoprostane” before and after drug administration. These cases where a comparison sentence was not detected due to missing patterns were very few.

- (6) a. Both paroxetine and placebo-treated patients improved to a similar extent on self-rated pain measures  
 b. Maximal plasma F2-isoprostane concentrations after IS + C (iron sucrose + vitamin C) were significantly elevated from baseline

More than 90% of the false positive cases, where we detected a component of a comparison incorrectly was due to parsing error. For example, in sentence (7), the compared aspect is incorrectly identified as “Sixty minutes” as the parser detects it as the subject of “higher” rather than “FEV(1)% increase”. If the phrase “Sixty minutes after” is removed, the parse is correct and we correctly identify the aspect. We would like to emphasize that most of the errors, either FN or FP, were due to incorrect parsing of complicated sentences rather than the incompleteness of our developed patterns.

- (7) Sixty minutes after the bronchodilator inhalation, the FEV(1)% increase was higher in OXI groups than in the IB group.



## 5 Conclusion

We have presented a system to identify comparison sentences and extract their components from literature using syntactic dependencies. The significance of developing a system to identify comparisons arises from the prevalent nature of comparative structures in the biomedical literature. We have observed that in a sample of abstracts describing randomized controlled trials or comparative studies, almost every abstract contained at least one comparison. Moreover, other text-mining applications might rely on extracting the arguments of a comparison. For example, this approach could be applied to mining reports of differential expression experiments, which are inherently comparisons. In (Yang et al., 2010), the authors defined seven comparative classes of differential expression analyses relevant to the processes of neoplastic transformation and progression, including cancer vs. normal tissue, high grade vs. low grade samples, and metastasis vs. primary cancer. Because comparative statements are often used to summarize the results of a study, these sentences are often of high interest to the reader. To the best of our knowledge, ours is the only work that attempts to cover a wide range of comparisons, capture all comparison components, and does not impose any restrictions on the type of compared entities. Our system achieved F-scores of 0.87, 0.78, 0.81 and 0.77 for identifying comparison sentences, aspects, scale and entities respectively. We plan to extend this work to consider situations, where one of the entities is implied and needs to be extracted from context.

## Acknowledgments

Research reported in this manuscript is supported by the National Institutes of Health under Grants No. U01GM120953 and No. 1 U01 HG008390-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## References

- Joan W Bresnan. 1973. Syntax of the comparative clause construction in english. *Linguist. Inq.* 4(3):275–343.
- Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*. Association for Computational Linguistics, Stroudsburg, PA, USA, NAACL 2000, pages 132–139.
- Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '05, pages 173–180.
- Marie-Catherine De Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D Manning. 2014. Universal stanford dependencies: A cross-linguistic typology. In *LREC*. volume 14, pages 4585–4592.
- Marcelo Fiszman, Dina Demner-Fushman, Francois M Lang, Philip Goetz, and Thomas C Rindflesch. 2007. Interpreting comparative constructions in biomedical text. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*. Association for Computational Linguistics, Stroudsburg, PA, USA, BioNLP '07, pages 137–144.
- Carol Friedman. 1989. A general computational treatment of the comparative. In *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '89, pages 161–168.
- Murthy Ganapathibhotla and Bing Liu. 2008. Mining opinions in comparative sentences. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*. Association for Computational Linguistics, Stroudsburg, PA, USA, COLING '08, pages 241–248.
- Betsy L Humphreys, Donald A B Lindberg, Harold M Schoolman, and G Octo Barnett. 1998. The unified medical language system. *J. Am. Med. Inform. Assoc.* 5(1):1–11.
- Nitin Jindal and Bing Liu. 2006a. Identifying comparative sentences in text documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, SIGIR '06, pages 244–251.
- Nitin Jindal and Bing Liu. 2006b. Mining comparative sentences and relations. *AAAI*.
- Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*. pages 55–60.
- David McClosky. 2010. *Any Domain Parsing: Automatic Domain Adaptation for Natural Language Parsing*. Ph.D. thesis, Providence, RI, USA.

- Dae Hoon Park and Catherine Blake. 2012. Identifying comparative claim sentences in full-text scientific articles. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '12, pages 1–9.
- Thomas C Rindflesch and Marcelo Fiszman. 2003. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *J. Biomed. Inform.* 36(6):462–477.
- Thomas C Rindflesch, Marcelo Fiszman, and Bisharah Libbus. 2005. Semantic interpretation for the biomedical research literature. In Hsinchun Chen, Sherrilynne S Fuller, Carol Friedman, and William Hersh, editors, *Medical Informatics*, Springer US, Integrated Series in Information Systems, pages 399–422.
- Steffen Staab and Udo Hahn. 1997. Comparatives in context. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence*. AAAI Press, Providence, Rhode Island, AAAI'97/IAAI'97, pages 616–621.
- Kaiquan Xu, Stephen Shaoyi Liao, Jiexun Li, and Yuxia Song. 2011. Mining comparative opinions from customer reviews for competitive intelligence. *Decis. Support Syst.* 50(4):743–754.
- Seon Yang and Youngjoong Ko. 2009. Extracting comparative sentences from korean text documents using comparative lexical patterns and machine learning techniques. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACLShort '09, pages 153–156.
- Zhen Yang, Fei Ren, Changning Liu, Shunmin He, Gang Sun, Qian Gao, Lei Yao, Yangde Zhang, Ruoyu Miao, Ying Cao, Yi Zhao, Yang Zhong, and Haitao Zhao. 2010. dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics* 11 Suppl 4:S5.