Exploring adjectival modification in biomedical discourse across two genres

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
National Library of Medicine
Bethesda, Maryland, 20894 – USA

olivier@nlm.nih.gov

Serguei V. Pakhomov
Division of Medical Informatics Research
Department of Health Sciences Research
Mayo Clinic
Rochester, Minnesota, 55905 – USA

Pakhomov.Serguei@mayo.edu

Abstract

Objectives: To explore the phenomenon of adjectival modification in biomedical discourse across two genres: the biomedical literature and patient records. **Methods**: Adjectival modifiers are removed from phrases extracted from two corpora (three million noun phrases extracted from MEDLINE, on the one hand, and clinical notes from the Mayo Clinic, on the other). The original phrases, the adjectives extracted, and the resulting demodified phrases are compared across the two corpora after normalization. Quantitative comparisons (frequency of occurrence) are performed on the whole domain. Qualitative comparisons are performed on the two subdomains (disorders and procedures). **Results**: Although the average number of adjectives per phrase is equivalent in the two corpora (1.4), there are more adjective types in MAYO than in MEDLINE for disorders and procedures. For disorder phrases, the 38% of adjective types common to the two corpora account for 85% of the occurrences. The predominance of adjectives in one corpus is analyzed. **Discussion**: Potential applications of this approach are discussed, namely terminology acquisition, information retrieval, and genre characterization.

1 Introduction

In previous studies, we demonstrated the feasibility of using NLP techniques such as shallow parsing of adjectival modification for identifying hierarchical relations among biomedical terms (Bodenreider et al., 2001) and for extending an existing biomedical terminology (Bodenreider et al., 2002). In these studies, the corpus was biomedical terminology or phrases extracted from the biomedical literature.

Other authors have explored adjectival modification in a clinical corpus. Chute and Elkin (1997) note, based on empirical observation of clinical data, that many clinical terms are accompanied by modifiers, including adjectives. The authors make a distinction between clinical modifiers (such as *chronic*, *severe*, and *acute*) and operational or administrative qualifiers (such as *no evidence of*, *history of*, and *status post*). It appears that the class of clinical modifiers consists primarily of adjectives that provide specific information regarding condition and are distributed on a scale. They suggest that operational modifiers be kept separate from the terms themselves in order to avoid combinatorial explosion.

Taking this idea one step further, we believe that, besides operational modifiers, other adjectives encountered in clinical phrases could receive a special treatment in applications such as information retrieval. For example, adjectives expressing nuances useful only in the context of clinical care could be removed from the phrase when searching the biomedical literature. This is the case of adjectives expressing degree of certainty (e.g., *probable*). In other cases, adjectives specific to clinical phrases can be mapped to synonyms or closely related modifiers (e.g., *greenish sputum*, *green*

sputum). The ability to map stylistic variations of the same adjective becomes especially important to establishing links between clinical records and scientific literature, which actually has significant implications for improving patient care in clinical practice as well as health science research. Finally, adjectives absent from the biomedical literature or terminologies may denote recent phenomena, not yet integrated in terminologies.

Knowledge about these classes of adjectives may help map across genres. Conversely, studying adjectival modification across genres may help identify adjectives whose representation varies across genres, possibly denoting one of these phenomena.

In the present paper, we explore the phenomenon of adjectival modification across two genres: the biomedical literature and patient records. The expected outcome of this study is to obtain a better characterization of adjectival modification in biomedical phrases of various origins, in order to fully take advantage of this phenomenon in applications such as the automatic construction of terminology and ontology resources and the retrieval of clinical documents.

2   Background

Adjectival modification as well as lexical semantics of adjectives has been studied extensively in the linguistic and NLP literature. Most approaches have been directed at creating adjective taxonomies and other ways of classifying and representing adjectives according to their properties and function. Raskin and Niernburg (1995) provide a comprehensive overview of the various approaches that have been taken to description, classification and representation of adjectives.

From the NLP standpoint, Fellbaum (1993) partitions adjectives in WordNet® [1] into two large classes: descriptive and relational. Descriptive adjectives "ascribe a value of an attribute to a noun" (p.27) (i.e., *big child*) while relational adjectives are usually derived from and are somehow associated with a noun (i.e., *musical child*). Another prominent distinction has to do with whether an adjective can express continuous (scalar) or discrete (non-scalar) values. Raskin and Niernburg (1996) point out that for text meaning representa-

tion for computational semantics, the most important distinction to make is between scalar and non-scalar. They also present a method for incorporating the semantics of the modifier adjective into the semantics of the modified noun by representing nouns as frames with elements such as ATTRIBUTE_SIZE than can be filled in by the semantic content of the modifying adjectives.

The major contribution of this study is to explore adjectival modification across two genres in the biomedical domain. Our approach is essentially practical and oriented towards applied perspectives.

3   Resources

The two genres compared in this study are the biomedical literature and patient records. More precisely, we use MEDLINE as our bibliographic corpus and clinical notes recorded at the Mayo Clinic as our clinical corpus.

MEDLINE® [2], the U.S. National Library of Medicine's (NLM) premier bibliographic database, contains over twelve million references to articles from more than 4,600 worldwide journals in life sciences with a concentration on biomedicine. Srinivasan et al. (2002) performed a shallow syntactic analysis on the entire MEDLINE collection, using only titles and abstracts in English. From the 175 million noun phrase types identified in their study, we selected the subset of "simple" phrases, i.e., noun phrases excluding prepositional modification or any other complex feature. In this study, a randomly selected subset of three million of these simple noun phrases constitutes our bibliographic corpus.

The Mayo Clinic is a group medical practice in the United States and spans all recognized medical care settings and specialties. Currently over 50,000 patient visits occur each week that generate 40,000 medical documentation entries in Mayo electronic record that principally consists of text narratives. The current size of the collection is approaching fifteen million notes and each note has on average 200 to 250 words of text. For this study we considered only the most current sample of the clinical notes collection – 1,783,377 documents recorded in 2002. Only simple noun phrases of the same type extracted from MEDLINE were extracted

---

from this corpus, resulting in a set of 9,665,942 phrases. A randomly selected subset of three million of these simple noun phrases constitutes our clinical corpus.

In both cases, the noun phrases were first normalized for case, so that the two subsets studied represent three million noun phrase types each.

Another resource used in this study is the Unified Medical Language System® [3] (UMLS®) Metathesaurus®. The Metathesaurus, also developed by NLM, is organized by concept or meaning. A concept is defined as a cluster of terms representing the same meaning (synonyms, lexical variants, acronyms, translations). The 14th edition (2003AA) of the UMLS Metathesaurus contains over 1.75 million unique English terms drawn from more than sixty families of medical vocabularies, and organized in some 875,000 concepts.

In the UMLS, each concept is categorized by semantic types from the Semantic Network. McCray et al. (2001) designed groupings of semantic types that provide a partition the Metathesaurus and, therefore, can be used to extract consistent sets of concepts corresponding to a subdomain, such as disorders or procedures.

4 Methods

In order to compare the linguistic phenomenon of adjectival modification across two corpora of noun phrases, we first extracted the adjectives after submitting the phrases to a shallow syntactic analysis and normalizing the head noun of the phrase for inflectional variation. Then, we compared across corpora the adjectives on the one hand and the "demodified" noun phrases [4] (i.e., noun phrases from which the adjectives have been removed) on the other. In order to address the size of these corpora, we limited the focus of our study to a significant subdomain of clinical medicine: disorders and procedures.

4.1 Extracting adjectives

Figure 1 illustrates the sequence of methods used for extracting adjectives from the original noun phrases. It also presents the number of phrases present before and after each of the four steps detailed below.

Step 1. Syntactic analysis

The phrases in our bibliographic and clinical samples were then submitted to an underspecified syntactic analysis described by Rindflesch et al. (2000) that draws on a stochastic tagger (see (Cutting et al., 1992) for details) as well as the SPECIALIST Lexicon [5], a large syntactic lexicon of both general and medical English that is distributed with the UMLS. Although not perfect, this combination of resources effectively addresses the phenomenon of part-of-speech ambiguity in English.

The resulting syntactic structure identifies the head and modifiers for the noun phrase analyzed. Each modifier is also labeled as being adjectival, adverbial, or nominal. Although all types of modification in the simple English noun phrase were labeled, only adjectives and nouns were selected for further analysis in this study. For example, the phrase *abnormal esophageal motility study* was analyzed as:

```
[[mod([abnormal,adj]),
  mod([esophageal,adj]),
  mod([motility,noun]),
  head([study,noun])]]
```

The result of the syntactic analysis was used to select the noun phrases suitable for studying the adjectival modification phenomenon, i.e., phrases having the following structure: `(adj+, noun*, head)`. The phrase is required to start with an adjectival modifier, possibly followed by other adjectives and end with a head noun, possibly preceded by other nouns. This specification excludes both simple phrases (e.g., one isolated noun) and complex phrases, not suitable for our analysis.

Step 2. Normalizing the head noun

In order to compare phrases across corpora, we normalized the head noun for inflectional variation in each noun phrase. As a result, the two noun phrases *cerebrovascular accident* (in MAYO) and *cerebrovascular accidents* (in MEDLINE) are considered equivalent. When both the singular and the plural form of a phrase appear in the same corpus, only the singular form is considered for further processing. In practice, to normalize head nouns, we used the program *lvg* [6], developed at NLM and distributed with the UMLS.
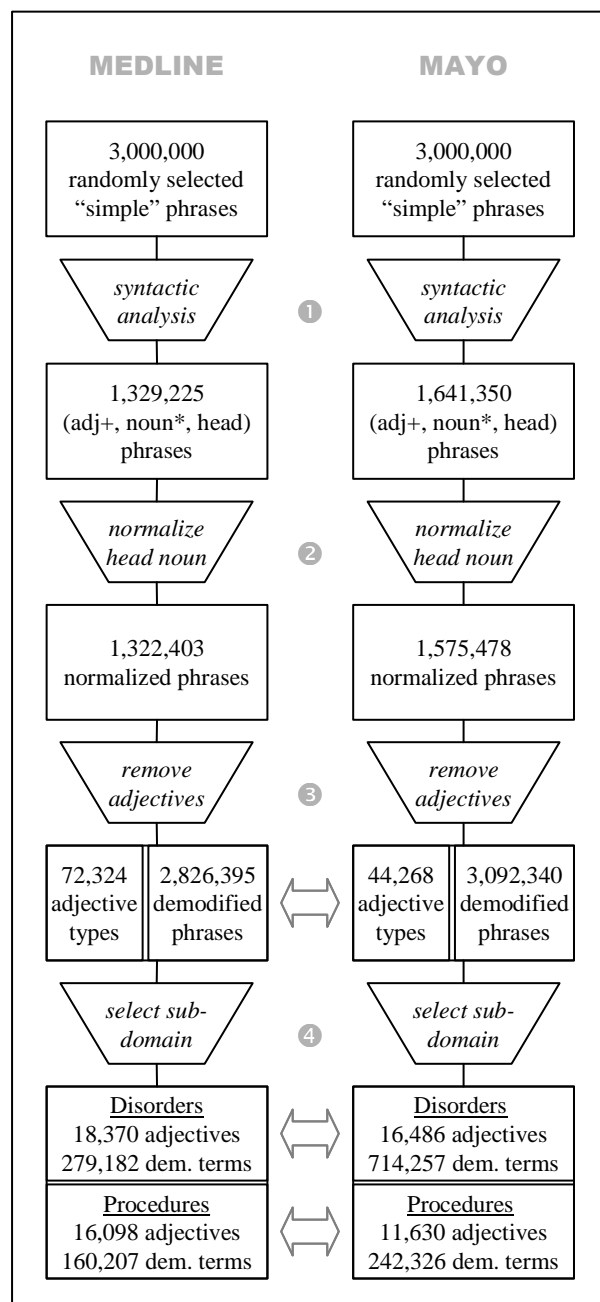
Figure 1. Summary of the methods.

Step 3. Creating demodified phrases

When adjectives are identified in a phrase O, a set of demodified phrases {T_1, T_2,…,T_n} is created by removing from phrase O any combinations of adjectival modifiers found in it. While the structure of the demodified phrases remains syntactically correct, the semantics of some phrases may be anomalous, especially when adjectives other than the leftmost are removed. Since most of them are semantically valid, we found it convenient to keep all demodified phrases for further analysis. Demodified phrases with incorrect semantics will be filtered out later in the experiment, since they will appear with a lower frequency.

The number of demodified phrases derived from a given phrase is $2^m – 1$, m being the number of adjectives in the phrase. For example, the phrase *acute respiratory infection syndrome* starts with the two adjectival modifiers *acute* and *respiratory*, so that the following three demodified phrases are generated *respiratory infection syndrome*, *acute infection syndrome*, and *infection syndrome*.

Step 4. Restricting to disorders and procedures

Because of the large size of the two corpora, we only performed a quantitative analysis of adjectival modification for the whole biomedical domain. We restricted the qualitative study to disorders and procedures. These represent a significant subdomain of clinical medicine, yet are small enough to be able to perform at least a somewhat detailed analysis.

All phrases, original and demodified, were mapped to the UMLS Metathesaurus by first attempting an exact match between phrases and Metathesaurus concepts. If an exact match failed, normalization was then attempted. This process makes the input and target terms potentially compatible by eliminating such inessential differences as inflection, case and hyphen variation, as well as word order variation. From the phrases mapping to some concept in the UMLS, we selected those for which the semantic category of the concept mapped to corresponded to the subdomains of interest. In practice, for a phrase to be considered a procedure, it had to map to a UMLS concept and the semantic type of this concept had to belong to the semantic group Procedures. The same principle was used for selecting disorders, using the semantic group Disorders. For example, the demodified phrase *arthroscopic surgery* (derived from *decompressive arthroscopic surgery*) is considered a procedure because it maps, as a synonym, to the concept *Surgical Procedures, Arthroscopic*, whose semantic group is Procedures. Exceptionally (32 UMLS concepts), a term may name both a disorder and a procedure. These terms are simply counted twice, once with Disorders and once with Procedures.

4.2 Comparing corpora

In order to investigate the characteristics of each corpus (noun phrases extracted from the biomedical literature and from patient records), we used two kinds of comparisons: quantitative and qualitative. The quantitative part consists of comparing frequencies of adjectives and demodified phrases across corpora, for the whole corpus as well as on specific subsets (Disorders and Procedures). In the qualitative part, we examined only phrases form the subdomains of Disorders and Procedures.

Quantitative comparisons

As mentioned earlier, the head noun of each phrase was normalized for inflectional variation (see Step 2 above). The purpose of normalizing the head noun is two-fold. First, it contributes to identifying phrase variants within each corpus, resulting in accurate counts of phrase types after duplicates had been removed. Second, it provides a simple means (string match) for identifying equivalent phrases across corpora.

We computed the number of original phrases, adjectives, and demodified phrases in each corpus, counting tokens and types in each category. Additionally, we explored similarities between the two genres by computing the number of phrases and adjectives common to the two corpora (intersection). Finally, we computed the number of phrase and adjective types for the two corpora taken together (union) in order to better characterize the whole domain. From these frequencies, we derived additional parameters such as the ratio of the number of adjectives to the number of original phrases.

Qualitative comparisons

We first extracted adjectives from the original phrases corresponding to Disorders and Procedures and computed their frequency of occurrence. Because phrases must map to a UMLS term in order to be identified as members of a subdomain, only the adjectives present in biomedical terms can be analyzed. For this reason, their rank will be studied rather than their frequency[7].

In order to better represent the whole spectrum of adjectives present in the two corpora, we then turned to the demodified phrases instead of the original phrases. In this second part, the condition

---

[7] rank n simply corresponds to the nth highest frequency

for a phrase to be considered a member of a subdomain was that the demodified phrase (not the entire phrase) map to a UMLS term. However, some adjectives may be overrepresented when several demodified phrases map to a UMLS term in the subdomains considered. For example, the phrase *abdominal vascular reconstructive surgery*, once demodified, maps to both *vascular surgery* (with modifiers *abdominal* and *reconstructive*) and *reconstructive surgery* (with modifiers *abdominal* and *vascular*). In this case, the adjective *abdominal* was counted twice.

For each adjective, we determined the corpus in which it was predominantly used. If more than half of the occurrences appear in one corpus, the adjective is considered predominant in this corpus. When more than half of the occurrences appear in both corpora, the adjective is considered common to the two corpora.

5 Results

5.1 Extracting adjectives

Out of the 3 million simple noun phrases randomly selected from MEDLINE, 1,322,403 phrase types were selected for further processing. Out of these, 72,324 adjective types (1,916,530 tokens) were extracted and 2,826,395 demodified phrases were generated. 1,575,478 phrase types were selected from the 3 million noun phrases in the MAYO corpus. Out of these, 44,268 adjective types (2,209,778 tokens) were extracted and 3,092,340 demodified phrases were generated. Details about the number of phrases selected at each step of the processing are given in Figure 1.

5.2 Comparing corpora

Quantitative results

The number of original phrases (Table 1), adjectives (Table 2), and demodified phrases (Table 3) are presented below in tabular format. Counts are broken down by corpus (MEDLINE and MAYO), on the one hand, and by subdomain (Disorders and Prodedures), on the other. Tables also include results obtained on the whole corpus (All), i.e., without subsetting, and on the union of the two corpora (Together). Except for original phrases (Table 1), which, by design, are phrase types, Table 2 and

Table 3 contain the numbers of types (upper left) and tokens (lower right).

The number of adjectives per phrase ranges from 1 to 16 in MEDLINE and from 1 to 7 for MAYO when the whole corpus is considered. The maximum number of adjectives per phrase is 6 or 7 for the various subsets. Phrases containing so many adjectives may look syntactically and semantically suspicious. While some of them denote extraction errors (often due to inappropriate part-of-speech tagging), most correspond to valid phrases and reflect the complexity of the biomedical domain (e.g., *diastolic systolic mean middle cerebral artery blood flow velocity* and *combined enteral parenteral synthetic hypercaloric nutrition*). The distribution of the number of adjectives per phrase is plotted in Figure 2.

Although the number of phrases processed is slightly more important for MAYO (1,575,476) than for MEDLINE (1,322,403), and although the ratio of the number of adjective tokens extracted to the number of original phrases is roughly similar in the two corpora (1.45 for MEDLINE and 1.40 for MAYO), there are significantly more adjective types in MEDLINE (72,324) than in MAYO (44,268). A difference in the opposite direction is observed in the Disorders and Procedures subsets, where the number of adjective types is higher in MAYO than in MEDLINE, while the average number of adjectives per phrase is still slightly higher in MEDLINE (1.27 vs. 1.21 for Disorders and 1.21 vs. 1.14 for Procedures). This finding requires further investigation.

Despite reducing the variation by normalizing head nouns for inflection, less than 3% of the original phrases are common to the two corpora. This proportion is significantly higher for the subset of disorder and procedure phrases where up to one third of MEDLINE phrases can be found in the MAYO corpus. Not surprisingly, the proportion of adjectives in common is higher. Overall, 44% of the adjectives in MAYO are also found in MEDLINE and up to 75% of the adjectives in MEDLINE are also found in MAYO (for disorders). Interestingly, the adjectives common to both corpora are also the most frequent. For example, as shown in Table 2, the 1,584 adjective types in common in the subset Disorders account for 38% of all adjectives for Disorders (4,148), but the corresponding 25,557 adjective tokens account for 85% of all tokens (30,046).

Table 1 – Number of original phrases (types), for Disorders (Di) and Procedures (Pr)

	MEDLINE	MAYO	Together	Common
Di	4,941	19,641	22,774	1,808
Pr	1,534	4,959	6,028	465
All	1,322,403	1,575,476	2,857,848	40,031

Table 2 – Number of adjectives (types [top] and tokens [bottom]), for Disorders (Di) and Procedures (Pr)

	MEDLINE	MAYO	Together	Common
Di	2,048	3,684	4,148	1,584
	6,299	23,747	30,046	25,557
Pr	902	1,499	1,790	611
	1,852	5,667	7,519	5,683
All	72,324	44,268	97,762	18,830
	1,916,530	2,209,778	4,126,308	3,885,852

Table 3 – Number of demodified phrases (types [top] and tokens [bottom]), for Disorders (Di) and Procedures (Pr)

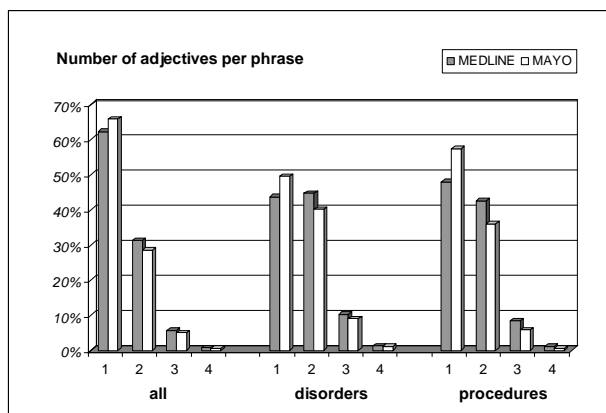|	MEDLINE	MAYO	Together	Common
Di	22,031	24,719	34,302	12,448
	174,548	463,097	637,645	571,041
Pr	9,850	8,595	13,691	4,754
	101,323	166,180	267,503	241,790
All	1,487,889	1,047,772	2,403,504	132,157
	2,826,395	3,092,340	5,918,735	2,709,100

Figure 2. Distribution of the number of adjectives per phrase

Qualitative results

The list of the most frequent adjectives found in the original phrases corresponding to Disorders and Procedures in the UMLS is given in Table 4, with their rank in each corpus. Interestingly, most high-ranking adjectives are found in both corpora.

Table 4 – Rank of the most frequent adjectives in MEDLINE (ME) and MAYO (Ma)

Disorders	ME	Ma	Procedures	ME	Ma
chronic	2	2	total	1	2
normal	3	1	surgical	2	3
acute	4	3	partial	5	1
congenital	1	8	serum	4	5
increased	6	5	patient	13	4
abnormal	8	4	percutaneous	3	15
neonatal	17	>100	renal	12	7
decreased	11	7	pulmonary	10	12
pulmonary	10	9	ultrasound	>100	22
benign	7	13	general	>100	23
renal	9	11	cardiac	16	8
recurrent	15	6	spinal	11	14
multiple	12	10	radical	14	13
increasing	14	12	evoked	29	>100
malignant	5	27	coronary	8	24
fetal	33	>100	femoral	>100	33
nasal	>100	33	studied	33	>100
joint	18	18	aortic	>100	34
intracranial	40	>100	fluid	7	27
positive	24	17	abdominal	24	11

Considering not the original phrases, but demodified phrases corresponding to disorders and procedures, most adjectives with a frequency greater than 10 are found in the two corpora (86% for disorder and 80% for procedures). However, their representation may differ largely across corpora. Examining the contexts of adjectives for Disorders (4978 adjectives with a frequency greater than 10), we found that 40% of the adjectives appear predominantly in MAYO (e.g., *mild*, *possible*, *recent*, *probable*, *questionable*, *greenish*), 20% predominantly in MEDLINE (e.g., *experimental*, *human*, *neonatal*, *canine*, *intracellular*), while 40% share most of their contexts across the two corpora (e.g., *acute*, *chronic*, *recurrent*). The repartition of the demodified phrases for Disorders (8263 phrases with a frequency greater than 10) is somewhat different. 65% of the demodified phrases appear predominantly in MAYO (e.g., *discomfort*, *tenderness*, *low back pain*, *chest pain*, *diarrhea*), 15% predominantly in MEDLINE (e.g., *resistance*, *strain*, *vesicle*, *hyperthermia*), while 20% share most of their contexts across the two corpora (e.g., *disease*, *lesion*, *pain*, *symptom*, *abnormality*).

6   Applications

In this section, we briefly examine some of the applications that may benefit from a better knowledge of adjectival modification in biomedical discourse: genre characterization, terminology and ontology acquisition, and information retrieval.

Genre characterization

Knowledge about adjectives and demodified phrases predominantly associated with one corpus may be useful to characterize corpora, and in this experiment, genres. Although limited, this study suggests, for example, that a clinical corpus contains markers for uncertainty (e.g., *possible*, *probable*, *questionable*) and non-specific symptoms (e.g., *discomfort*, *low back pain*). On the other hand, in a broad bibliographic corpus, precisions about organism or age groups must be given (e.g., *human*, *canine*, *neonatal*). Interestingly, while the term *fever* is found with no predominance in either corpus, its more scientific synonyms *hyperthermia* and *pyrexia* are used predominantly in MEDLINE. If corroborated, this finding may suggest that, although both scientific publications and medical records are geared toward peers, the language used in scientific publications tends to be more specialized.

Terminology and ontology acquisition

The method described in this paper constitutes a useful technique for adapting existing terminologies and ontologies with empirically derived terms from a new subdomain. First, demodified phrases are more likely to be mapped to another corpus. And second, because adjectival modification often denotes a hyponymic relation between a phrase without modifier and a modified phrase, the modified phrase can be linked as a candidate hyponym to the phrase without modifier (Bodenreider et al., 2002).

This approach could be used, for example, for adapting biomedical terminologies to subtle clinical nuances. When used with exactly the same subdomain the existing terminology comes from, this technique could enable regular updates of the

terminology provided that current textual data is used for phrase extraction.

The approach is currently limited to simple adjectival modification; however, this is a self-imposed limitation. Theoretically, the same methodology can be adapted to work on nominal, prepositional phrase and other types of modification.

Information retrieval

Terminologies as well as ontologies are frequently used for information or document retrieval in the domains for which such terminologies or ontologies are available. Medicine is one such domain where there are numerous terminological resources. Integrated in a system such as the UMLS, these resources provide, for example, many synonyms for each concept, increasing the chances of retrieving documents from a given term. However, most terms in these resources are pre-coordinated and may not include all the variants needed in various contexts. Moreover, most terms are noun phrases and, while synonyms are often given for nouns, it may not be the case for their modifiers. For example, while the various synonyms for *fever* (e.g., *hyperthermia* and *pyrexia*) are present in the UMLS, there is no *greenish* variant for *green sputum*. Nor can there systematically be a variant denoting uncertainty. Therefore, identifying classes of adjectives that can be either ignored (e.g., uncertainty markers) or mapped to other adjectives (e.g., *greenish* to *green*) would increase the performance of information retrieval systems operating on clinical corpora. In light of these findings, existing terminologies and ontologies can provide a core of medical concepts common to most subdomains; whereas the methodology described here can be used to tailor the general-purpose terminological resources to accommodate subdomain-specific terminology services.

7 Conclusions

In conclusion, adjectival modification plays an important role in biomedical texts, and knowledge about this phenomenon can be exploited in applications such as the retrieval of biomedical documents and for developing terminology services in the biomedical domain.

In the future, we would like to identify patterns in biomedical terms and phrases based, in part, on classes of adjectival modifiers. Creating such a model for terms would constitute a generative approach to biomedical terminology, contrasting with the lists of precoordinated terms populating most terminology systems in the biomedical domain.

References

Bodenreider, O., Burgun, A., and Rindflesch, T. C. (2001). *Lexically-suggested hyponymic relations among medical terms and their representation in the UMLS*. Proceedings of TIA'2001 "Terminology and Artificial Intelligence", 11-21.

Bodenreider, O., Rindflesch, T. C., and Burgun, A. (2002). *Unsupervised, corpus-based method for extending a biomedical terminology*. Proceedings of the ACL'2002 Workshop "Natural Language Processing in the Biomedical Domain", 53-60.

Chute, C. G., and Elkin, P. L. (1997). *A clinically derived terminology: qualification to reduction*. Proc AMIA Annu Fall Symp, 570-574.

Cutting, D. R., Kupiec, J., Pedersen, J. O., and Sibun, P. (1992). *A practical part-of-speech tagger*. Proceedings of the Third Conference on Applied Natural Language Processing, 133-140.

Fellbaum, C. (1993). Five Papers on WordNet: Adjectives in Wordnet, D. Gross, ed.

McCray, A. T., Burgun, A., and Bodenreider, O. (2001). *Aggregating UMLS semantic types for reducing conceptual complexity*. Medinfo *10*, 216-220.

Raskin, V., and Niernburg, S. (1995). *Lexical Semantics of Adjectives: A Microtheory of Adjectival Meaning*. Memoranda In Cognitive and Computer Science *MCCS-95-288*.

Raskin, V., and Niernburg, S. (1996). *Adjectival Modification in Text Meaning Representation*. Proceedings of COLING '96, 842-847.

Rindflesch, T. C., Rajan, J. V., and Hunter, L. (2000). Extracting molecular binding relationships from biomedical text. In "Proceedings of the 6th Applied Natural Language Processing Conference" (San Francisco, Morgan Kaufmann Publishers), pp. 188-195.

Srinivasan, S., Rindflesch, T. C., Hole, W. T., Aronson, A. R., and Mork, J. G. (2002). *Finding UMLS Metathesaurus concepts in MEDLINE*. Proc AMIA Symp, 727-731.