AN EXPLORATORY STUDY USING THE PREDICATE-ARGUMENT STRUCTURE TO DEVELOP METHODOLOGY FOR MEASURING SEMANTIC SIMILARITY OF RADIOLOGY SENTENCES

Eric Tyner Newsom

Submitted to the faculty of the School of Informatics
in partial fulfillment of the requirements
for the degree of
Master of Science in Healthcare Informatics
Indiana University

May 2013

Accepted by the Faculty of Indiana University,
in partial fulfillment of the requirements for the degree of Master of Science
in Healthcare Informatics

Master's Thesis Committee

Dr. Josette Jones, RN, Ph.D., Chair

Dr. Roland E. Gamache, Jr., Ph.D.

Dr. Malika Mahoui, Ph.D.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Eric Tyner Newsom

AN EXPLORATORY STUDY USING THE PREDICATE-ARGUMENT STRUCTURE TO DEVELOP METHODOLOGY FOR MEASURING SEMANTIC SIMILARITY OF RADIOLOGY SENTENCES

The amount of information produced in the form of electronic free text in healthcare is increasing to levels incapable of being processed by humans for advancement of his/her professional practice. Information extraction (IE) is a sub-field of natural language processing with the goal of data reduction of unstructured free text. Pertinent to IE is an annotated corpus that frames how IE methods should create a logical expression necessary for processing meaning of text. Most annotation approaches seek to maximize meaning and knowledge by chunking sentences into phrases and mapping these phrases to a knowledge source to create a logical expression. However, these studies consistently have problems addressing semantics and none have addressed the issue of semantic similarity (or synonymy) to achieve data reduction. To achieve data reduction, a successful methodology for data reduction is dependent on a framework that can represent currently popular phrasal methods of IE but also fully represent the sentence. This study explores and reports on the benefits, problems, and requirements to using the predicate-argument statement (PAS) as the framework. A convenient sample from a prior

study with ten synsets of 100 unique sentences from radiology reports deemed by domain experts to mean the same thing will be the text from which PAS structures are formed.

CHAPTER ONE: INTRODUCTION & BACKGROUND

Today's knowledge worker has far too much published information to review for advancement of his/her professional practice (Zweigenbaum & Demner-Fushman, 2009). In healthcare, the majority of knowledge is in the form of free text such as professional journals and patient clinical notes (Demner-Fushman, Chapman, & McDonald, 2009). These forms of free text have been linked to profits (Friedman & Hripcsak, 1999) and are the basis for evidence-based practice. Analyzing these unique forms of knowledge by a machine-driven mechanism in order to decrease the knowledge worker's burden of reading and maximizing patient outcomes is a goal of natural language processing (NLP). According to Friedman & Hripscsak, NLP has three essential components: a) understanding the syntax of free text, b) translating the syntax into semantics, and c) inferring the meaning of the text to a domain's knowledge. If those essential components are addressed, NLP has achieved data reduction.

Due to an increasing amount and complexity of free text forms in healthcare, NLP needs more sophisticated methods (Zweigenbaum & Demner-Fushman, 2009). Information extraction (IE), a branch of NLP, uses such methods as question & answering (Q & A) and text summarization to achieve data reduction of specialized and sophisticated forms of free-text of which radiology reports are the most analyzed in healthcare (Demner-Fushman, Chapman, & McDonald, 2009). Zeigenbaum & Demner-Fushman contend that in order for IE methods to be successful an essential component is an annotated corpus. A corpus is a collection of free-text documents that represent a unique domain of knowledge, such as radiology reports, and allows for identification of lexical and syntactical traits (Millar & Budgell, 2008). A corpus is annotated to help train IE methods to establish the sense or meaning of the text (Miltsakaki et al, 2010; Kim, Ohta, & Tsujii, 2008) and to create a recognizable symbol that can be translated into a logical expression necessary for artificial intelligence software processing. While studies experiment with size and style of annotated corpora (Xu, Tsujii & Change, 2012; Miller, Guinnes, & Zamanian, 2004), studies do show that a lack of annotated corpora impede the effectiveness of IE (Chapman et al., 2011; Jones, Newsom, & Delaney, 2009).

The most used type of NLP annotated based processing is phrasal (Goulart, Strube de Lima & Xavier, 2011). Nadkarni, Ohno-Machado & Chapman (2011) describe in detail the typical low-level NLP processing that takes a detected bound sentence and chunks it into indentifying phrases for higher-level processing. Higher-level tasks, greatly dependent on the annotated corpus, categorize the phrases and map them to a concept in a vocabulary where a logical expression can be assigned. A logical expression is typically a record ID code in some reference source by which the computer can process. Table 1 gives an example of how two sentences are parsed into phrases and how those phrases may be assigned a logical expression.

Complete Sentence:	This lesion is suspicious for a neoplasm such as a brainstem glioma or astrocytoma.	The CT image shows a lesion in the brainstem possibly indicating a glioma.	
Phrase	**Logical Expression (UMLS Code)**	**Phrase**	**Logical Expression (UMLS Code)**
This		The	
lesion	C0221198	CT image	C0040405
is		shows	
possibly	C2362652	a	
a		tumor	C3273930
neoplasm	C0027651	in the	
such as		brainstem	C1306665
a		possibly	C2362652
brainstem	C1306665	indicating	
glioma	C0017638	secondary	C2939419
or		metastases	
astrocytoma	C0004114		
Sentential Logical Expression:	C0221198 & C2362652 &C1306665 & C0017638 & C0004114	C0040405 & C3273930 & C2362652 & C1306665 &C2939419	

Table 1--Phrasal Annotation & Associated Logical Expression

A complete, complex thought is then the combination of multiple concepts resulting in multiple logical expressions (bottom row of Table 1). Data reduction occurs by synonymy relationships in the knowledge source. In the two sentences in Table 1, the knowledge source would need to know that neoplasm (C0027651) and tumor (C3273930) are synonyms and that glioma (C0017638), astrocytoma (C0004114) and secondary metastases (C2939419) are synonyms for something cancerous. Through such semantic relationships, NLP methods achieve data reduction that both sentences represent the concept for potential cancerous growth in the brain stem. The size of annotated documents for phrasal NLP has had relative success in processing free text meaning from minimal corpus sizes as long as the corpus is well representative of the domain (Juckett, 2012).

Another approach to annotation is using the sentence as the unit of representation (Friedlin, Mahoui, Jones & Jamieson, 2011). In sentential annotation, the sentence could represent a single, simple concept or a complex, compound concept. Either way, the sentence is assigned a single logical expression within a knowledgebase of sentential propositions. A product of this annotation method is the creation of sentential synsets, or syntactically different sentences, that mean the same thing. Table 2 shows how two sentences, both with same meaning, form a synset represented by a single logical expression.

<table>
<tr><td rowspan="4">Sentential Synset</td><td>Proposition:</td><td colspan="2">There is potential cancerous growth in the brainstem</td></tr>
<tr><td colspan="2">Sentence</td><td>Logical Expression</td></tr>
<tr><td colspan="2">This lesion is suspicious for a neoplasm such as a brainstem glioma or astrocytoma.</td><td>S1001</td></tr>
<tr><td colspan="2">The CT image shows a tumor in the brainstem possibly indicating secondary metastases.</td><td>S1001</td></tr>
</table>

Table 2--Sentential Annotation & Synset for a Discrete Concept

Phrasal annotation needs multiple relationships among terms established in order to achieve data reduction on complex concepts, whereas for sentential annotation, one relationship is all that is needed. Because a sentence is not chunked into phrases, a large number of annotated sentences are necessary to represent the knowledge of a domain. Sentential annotation has to have a proposition for all the concepts possibly expressed in combination of various modifiers. There is no or little published research available to-date on the performance of sentential annotation.

1.1. Problem Statement

To date, most annotation methods have focused on a minimal amount of human expert time (Juckett, 2012; McCrae & Collier, 2008). This renders the question of how much annotated data is needed to maximize IE (Cohen & Hunter, 2006) and how to evaluate the contribution of larger annotated corpora such as used in sentential annotation (Demner-Fushman, Chapman, & McDonald, 2009). Because of the cost involved with annotation, a method to evaluate large scale annotation would help determine appropriate levels of resource investment.

Although most widely used, the phrasal annotation approach is consistently plagued with performance problems, among which semantics is the most common (Ware, Mullett, Jagannathan & El-Rawas, 2012; Hope, 2012; Cohen et al, 2005). Specifically, the term synonymy is used to represent the difficulties of an NLP engine to assign a common logical expression to two different terms or symbols (Nadkarni, Ohno-Machado & Chapman, 2011). To the contrary, sentential annotation creates a synset which serves as an index of semantically equivalent pieces of text (Friedlin, Mahoui, Jones & Jamieson, 2011). It is not known if sentential annotation can process the issue of synonymy better than phrasal annotation. However, if a method can be developed that accurately measures how the two methods process synonymy, then through synonymy, it could be possible to ascertain if large scale sentential annotation has a legitimate place in NLP for data reduction. To accomplish this evaluation, a phrasal annotation method has to be used that can retain the semantics of the entire sentence as is accomplished in sentential annotation. This is extremely important because methods based on phrasal annotation can tend to lose parts of a sentence because of uncertainty of how to relate the phrase to other phrases in the sentence (Sevenster, van Ommering & Qian, 2012).

1.2. Studies that Have Addressed the Problem

IE methods have to address the phenomenon of synonymy in order to achieve data reduction (Liu, Hogan, & Crowley, 2010; McCrae & Collier, 2008). This means that two different text elements are compared to determine what information is common between them. An integral aspect to consider when developing an evaluation method of annotated corpora is the annotation structure of a corpus that lends to analysis of synonymy (Miyao et al, 2006). A few in-depth studies show that an annotation structure with the sentence as the minimal text element of analysis will best achieve data reduction (Sevenster, van Ommering & Qian, 2012; Madnani & Dorr, 2010).

One method uses a form of semantic role labeling (SRL) called predicate-argument structures (PAS) for IE of free text and appears to be the most commonly researched and used method for analysis of sentence structures (Albright et al, 2013; Tan, Kaliyaperumal & Benis, 2012; Athenikos & Han, 2010; Chou et al, 2006; Cohen & Hunter, 2006; Godbert & Royaute, 2010; Kilicoglu et al, 2010; Kogan et al, 2005; Miltsakaki et al, 2010; Pyysalo et al, 2007; Surdeanu et al, 2003; Tsai et al, 2007; Wattarujeekrit, Shah, & Collier, 2004; Zeigenbaum & Demner-Fushman, 2009). PAS is a form of annotation that centers on the predicate of a sentence. How the predicate is used in language determines what semantic roles, or arguments, specific phrases of the sentence's text are assigned. In addition to arguments, modifiers such as location, temporal, and extent serve to clarify arguments but their role-labels are not dependent on verb. Table 3 shows how the description of argument roles differ among predicates. Table 4 shows how a sentence is annotated using PAS.

Verb	To be *(Arg* Role Definition)*	Pay *(Arg* Role Definition)*	Show *(Arg* Role Definition)*
Arg0		payer or buyer	shower
Arg1	topic	money or attention	thing seen/shown
Arg2	comment	person being paid, destination of attention	seer
Arg3		commodity, paid for what	
Arg4			
	*Arg = Argument		

Table 3--Argument Role Differences Among Predicates

Sentence	The student paid $300 per credit hour at IUPUI last year for tuition	
Verb	paid	
Possible Arguments	**Argument Description (Role Lable)**	**Source Text**
Arg0	*payer or buyer*	The student
Arg1	*money or attention*	$300 per credit hour
Arg2	*person being paid, destination of attention*	
Arg3	*commodity, paid for what*	for tuition
Argument Modifiers	**Modifier Description**	**Source Text**
ArgM-loc	*where action took place*	at IUPUI
ArgM-tmp	*when action took place*	last year

Table 4--Annotation of a Sentence Using PAS Based on SRL

While most studies involved with sentential studies use PAS to translate the syntax of a sentence into semantics, a few studies have tested semantic relatedness between two PAS frames (Tan, Kaliyaperumal & Benis, 2012; Miyao et al, 2006; Wang & Zhang, 2009). A methodology for evaluating annotated corpora intended for data reduction would seem to adhere to standard practice by using PAS.

1.3. Deficiencies in Past Literature

While the PAS is a suitable method to deconstruct the sentence for machine processing and to test semantic relatedness, testing for semantic similarity between two PAS frames is a test yet to be determined. The difference between semantic relatedness and similarity is an issue of specificity (Pedersen et al, 2007). Semantic relatedness relates to a general broad category that two different texts share something in common, such as bird and feather. In the largest study that compared two PAS frames, Miyao et al. (2006) used as PAS query to find matching pre-processed PAS frames out of 70 million. However, the PAS query frames were generalized in that the queries asked to find <dystrophin cause 'disease'> or <'something' enhances p53 (negative)>. One side of the query seeks a match that is broad and general. In the study by Wang & Zhang (2009), two PAS frames are compared for semantic relatedness as a pre-processing step and the study does not experiment with any further issues of semantics.

Semantic similarity is degree of shared meaning between two different texts or symbols and is concept used to explain how humans represent relationships within a language's vocabulary expressions to formulate meaning of experiences (Samsonovic & Ascoli, 2010). Specifically, semantic similarity tests how close or far the conceptual distance is between the true senses of words. (McCrae & Collier, 2008; Miltsakaki et al, 2010). Neither Miyao et al. (2006) nor Wang & Zhang (2009) measure the degree of how similar the two PAS frames are to each other. In order to understand the true meaning shared between two representative structures of a sentence, methods for data reduction should test for semantic similarity (Madnani & Dorr, 2010).

To test semantic similarity, concepts need to be represented at the atomic level in order to understand how broad or narrow one text element is to another (Caviedes & Cimino, 2004). Successful studies that have tested for semantic similarity have focused on terms or lexical elements (Chaves-Gonzales & Martinez-Gil, 2013; Builtelaar & Sacaleanu, 2001; Elhadad & Sutaria, 2007; Mihalcea, Corley, & Strapparava, 2006). These studies attest to the idea that less complexity of an expressed thought the easier it is to measure semantic similarity. In another study that comes closest to evaluating sentential semantic similarity through a method called Named Entity Recognition (NER), the method in Savova et al. (2010) fails with issues of complex synonymy, overlapping text span, and structured interpretations. Savova et al. attempt to look at the whole meaning of a sentence, but the study underlines the difficulty of comparing the equivalency between two sentences when dissected and chunked into smaller parts. It remains to be evaluated if PAS frames can rise to a level of complexity to analyze sentential semantic similarity.

1.4 The Significance of the Study

Although there is support in the literature that larger corpora can improve semantic processing in IE (Fan & Freidman, 2007), a large annotated corpus does not guarantee usage (Cohen et al., 2005). If the PAS is an accepted annotation method to deconstruct but maintain the meaning of complete sentences, it may bode well to determine if it is appropriate to measure sentential semantic similarity. If the PAS is able to successfully measure sentential semantic similarity and prove that phrasal annotation is just as good as or better than sentential annotation, it may be that there would not be a need for large annotated corpora (Wilbur, Rzhetsky, & Shatkay, 2006). Measurements used by the PAS process could be applied across various texts to determine shortcomings of various NLP methods because PAS is a domain independent method (Surdeanu et al, 2003). The reduction of human involvement with annotation would allow for a focus of time and resources on domain knowledge sources (such as domain vocabularies, terminologies, and ontologies) that will always serve as an essential component in any form of IE (Roberts et al, 2009; Shi & Mihalcea, 2005; Tsai et al, 2007). As such, the PAS could become an accepted method for data reduction and to deduce EBP for the fact that it lends well to automated annotation (Chou et al, 2006; Wattarujeekrit, Shah & Collier, 2004; Delisle & Szpakowicz, 1997). However, if PAS successfully measures sentential semantic similarity but shows that sentential annotation is better, then sentential annotation could very well be a justified expense as method to augment NLP for data reduction.

1.5 Purpose of the Study

The purpose of this study is to describe an annotation methodology using the predicate-argument structure as a pre-processing step for measurement of semantic similarity.

CHAPTER TWO: LITERATURE REVIEW

Since this study's focus is on semantic similarity, it is important to understand this concept and how it has been studied. According to Resnik (1999), semantic similarity tests derive their worthiness on the basis that the notion of similarity is a representation of human perception and intuition. He expands on the importance of similarity measures in that understanding a degree of how similar two concepts are versus just that they are related assists in predicting human performance. Since words represent concepts, such measures become a test of how similar two words are to each other. Even more generalized, semantic similarity measures the relationship between word senses (Elhadad & Sutaria, 2007). What is an appropriate use of a word's sense in one situation may not hold in another. This means that semantic similarity not only tests for shared meaning between two words but also tests for what is not shared and how two words are different (Lin, 1998).

There seems to be two methods of how to measure semantic similarity. The first method uses a taxonomy or an ontology to determine the shortest path between two words by counting the nodes between them (Rada et al, 1989). This method determines a path length or distance between two concepts. Caviedes (2004) expands on Rada's method by looking at multiple terminologies in the UMLS. Although this is a simple method whereby strong similarity is determined by a smaller number of nodes, this method assumes that there is an equal interval of distance between nodes (Resnik, 1999). If a taxonomy were designed with all elements of a domain represented at the atomic level to assure true granularity, this could be an effective method. This cannot always be assured. The irony of a disproportioned interval is that a path length of two nodes between two concepts could be larger in the real world because the taxonomy or ontology has failed to represent even more granular concepts. What this method does do is attempt to mimic the amount of information shared between two different words based on a knowledge representation of the real world and represent that measurement with a quantifiable number.

Resnik (1995) takes the path length method and develops a probabilistic measure based on the amount of shared information content between two concepts. Shared content is determined by the level of a node that is the most informative that subsumes both concepts. Resnik uses the example that ‘nickel’ and ‘dime’ will have a higher probability of sharing the same content because both are directly subsumed by the concept ‘coin’. To the contrary, ‘nickel’ and ‘credit card’ will have a lower probability of being similar because the level most common to them is further up the hierarchy--‘medium of exchange’. Resnik’s method avoids the granularity problem of Rada’s and addresses similarity based more on what is shared rather than distance.

Semantic similarity invokes another dimension other than shared meaning. The issue of directionality plays an important role in how humans form relationships with words (Wee & Hassan, 2008). ‘President Obama’ may trigger an association with ‘president,’ but ‘president’ may trigger an association with ‘boss’ as well as ‘President Obama’. The dimension of direction introduces the concept that one word or phrase may be more broad or narrow than another. Neither Resnik’s nor Rada’s method addresses the issue of directionality, but studies that have considered this dimension of semantic similarity have developed a scoring mechanism to indicate directionality of synonymy for human judgment to develop a ‘gold standard’ that serves to test aspects of NLP (Cambell et al, 1997; Chute et al, 1996; Humphreys, McCray, & Chey, 1997; Zielstorff et al, 1998). This study will use a scoring mechanism for semantic similarity to see if directionality can determine benefits to sentential synsets.

If the goal of IE is to evaluate semantic similarity, then annotation serves as the technical bridge. Annotation serves to train IE methods to mimic human decisions (Chou et al., 2006). One goal of annotation is to reach a balance between preserving the main idea and representing granularity in order to form a model of completeness of texts comprised of component words (Mihalcea et al, 2006). Because in this study phrasal annotation begins with the sentence, the first step of annotation begins with identifying granular syntactical structures forming the more sophisticated knowledge structure represented in the sentence’s idea (Pyysalo et al., 2007; Rost et al., 2008). These basic syntactical annotations will serve to test IE’s ability to automate semantic relationships from the basic to the more sophisticated.

In order to focus the direction of IE, annotation will have to address the uniqueness of the corpus. A decision has to be made as to what types of data to collect to develop framework to properly form relationships among the basic syntactical annotations (Cohen & Hunter, 2006). The types of data to collect are guided by the domain specific features of the corpus (Cohen et al, 2005; Tsai et al, 2007). In a study by Wilbur, Rzhetsky & Shatkay (2006), annotation focuses on identifying sentences that represent dimensions of science. These dimensions are focus (is a sentence expressive of scientific, general knowledge, or experimentation methods), polarity (does a sentence express positive or negative statement), certainty (does sentence express degree of certainty about assertions), evidence (does sentence express evidence to support focus or certainty), and trend (does sentence express a high/low or increase/decrease about finding). In a study by Kim, Ohta, & Tsujii (2008), data collection focuses around events pertaining to biomedicine (cell differentiation, protein translation, and etc.). For a radiology corpus, data collection may be centered on anatomical findings (Demner-Fushman, Chapman & McDonald, 2009). Finally, the style of corpus's written language has to be considered (Surdeanu et al, 2003). Rost et al. (2008) present the case that medical reports tend to be ungrammatical and pose complications for IE if the part of speech routines do not have an algorithm to address this characteristic. The annotation process selected for a radiology corpus will have to prove that its framework can represent anatomical findings and concepts that may often be expressed in ungrammatical structures.

An annotation framework functions as the technical method to guide annotators. Framework is the annotation's style (Roberts et al, 2009), and has to have a schema (way to address goal of IE) and a method. If the goal of annotation is to help IE achieve statistical results, then the schema will guide annotators to mark linguistic and lexical traits of the corpus. If goal of IE is to make inference, then annotators will determine relevant domain knowledge specific to corpus. Of course, all annotation should be anchored in text (Cohen et al, 2005). For IE designed to make inference, two methods are NER (Savova et al, 2010) and SRL (Cohen & Hunter, 2006). NER is a phrasal method (noun, verb, or prepositional phrase) that identifies text boundaries relevant to a knowledge representation source and relies on the context to resolve ambiguity in meaning (Demner-Fushman, Chapman & McDonald, 2009). NER also seeks to annotate at the atomic level (Wattarujeekrit, Shah & Collier, 2004) to address granularity. Section 1.3 of this proposal addresses the deficiencies of NER to address complex synonymy for sentential similarity. On the other hand, SRL, according to Cohen & Hunter, is a method that seeks to balance completeness of a sentences expression with granularity. SRL is a shallow parser in that it serves nothing more than to identify sentential structures that revolve around the sentence's predicate. Section 1.2 of this proposal addresses the direction of this study to use PAS as a unique form of SRL for annotation to see if it can address the issue of complex sentential semantic similarity.

The contribution of PAS to IE is its ability to retain structure of sentences (Cohen & Hunter, 2006), contribute to inductive learning (Surdeanu et al, 2003), and facilitate mapping of arguments to ontological references (Miyao et al, 2006). These are three components essential for successful NLP (Friedman & Hripscsak, 1999). Typically, PAS development is guided by guidelines introduced through the Propbank project (Wattarujeekrit, Shah, & Collier, 2004) discussed in section 3.3.1. An example of PAS for a radiology sentence is presented in Table 5.

Sentence	The intervertebral discs are normal in height.	
Predicate	**Argument 1 *(Topic)***	**Argument 2 *(Comment)***
are	The intervertebral discs	normal in height

Table 5--Sentence Represented in a PAS Frame

For the sentence in table 5, all text is annotated to a role associated with the predicate and labeled. There is no loss of meaning.

Just as annotation should consider the uniqueness of a corpus, Surdeanu et al. (2003) and Miyao et al. (2006) contend that modification of a PAS frame is encouraged to correctly identify observations of a domain's structure necessary for semantic translation. For a corpus of radiology reports, Friedlin et al. (2011) frequently encountered missing predicates and implied concepts. These findings concerning a radiological corpus could mean that annotation will identify arguments and predicates not anchored in text. However, if domain experts make this kind of inference then the purpose of annotation serves to mimic the process of human judgment. For sentences lacking a predicate, this study will have to determine a method in the annotation process for measuring successfulness of constructing complete sentences from incomplete sentences and inferring implied/incomplete text.

If method of annotation is PAS, then focus is on the predicate, and to assure that predicates function to guide argument development, Godbert & Royaute (2010) argue that verbs should be classified based on syntactical structures present in corpus. For example, does the predicate accept a direct object or does it not. Such syntactical analysis of how a sentence is structured around specific verbs could help understand the arguments. Another issue that has to be addressed before annotating PAS frames is to understand how the corpus handles nominalization of verbs (Kilicoglu et al, 2010). Nominalized verbs take different forms, such as gerunds, and these forms may have arguments despite not being the predicate of the sentence. This study will have to make a conclusion if the Propbank annotation guidelines and knowledge base can serve as the formal schema to build a PAS for sentential semantic similarity centered on predicates in a radiology corpus.

The complexity of using PAS as a formal schema to measure sentential semantic similarity lies not in forming syntactical phrases but examining similarity of text phrases of one sentence to text phrases of another. This kind of measurement goes beyond the discreteness of word to word comparison and requires a representational scheme to assist with matching which phrase of one sentence should be compared to a phrase of another sentence. To adhere to principles of semantic similarity, this representational scheme not only has to address syntax and conceptual needs in a sentence, but it has to provide a foundation to capture the accumulative sense formed by concepts in the sentence (Miyao, et al, 2006). The goal, then, is to have a representational scheme whereby the output of syntactical phrases formed from one sentence are appropriately matched and compared to those of another sentence, and this representational scheme will serve as the basis for pattern discovery of synonymy (McCrae & Collier, 2008).

Creating this scheme could pose the biggest obstacle to using the PAS. Friedlin et al. (2011) state PAS may not scale; however, this statement has not yet been proven. The problem of scalability of PAS is presented when two semantically equivalent sentences have two different predicates. In such cases, it is possible that content of the arguments and modifiers of one sentence may not match to the other predicate's corresponding arguments and modifiers. In some cases, there may be no text to compare. Table 6 shows how three semantically equivalent sentences present the problem of scalability with three difference predicate senses.

Proposition	There is a left lower lobe pulmonary infiltrate(s).		
Sentence 1	The lungs demonstrate left basilar atelectasis or infiltrate .		
Sentence 2	There is interval clearing of the left lower lobe infiltrate .		
Sentence 3	The left lower lobe infiltrate is identified .		
	Sentence 1	Sentence 2	Sentence 3
Predicate	demonstrate	be	identified
Predicate Use	*show_off*	*existential*	*label, call*
Arg0	The lungs		
Arg1	atelectasis or infiltrate	interval clearing	infiltrate
Arg2			The left lower lob
Modifier-Locative	left basilar	of the left lower lobe	
Modifier-Cause		infiltrate	

Table 6--Scalability Problem of PAS

In each sentence, PAS annotates the complete syntactical structure of the sentence; therefore, the complete sense of each sentence is maintained. But when the goal is to compare the semantics across each role, problems arise from the fact that each predicate assumes a different structure for role assignments. It is evident that when trying to measure semantic similarity of varying predicates, not every role will contain like content. Arg1 of sentence 1 & 3 have semantic similarity but Arg1 of sentence 2 is not semantic similar despite the fact that the defined roles of the verb places the actual content that could be compared in the modifier ‘cause’. The question then is how is it possible to ‘cross’ compare arguments for a synset comprised of various predicates? Is it possible to compare differing semantic roles? This study will investigate patterns among synsets with varying predicates for feasibility of an algorithm allowing cross comparing of semantic roles.

2.1 Theoretical Perspectives

In order to extract semantic meaning from free text, deconstruction of free text must incorporate access to some reference based on a concept-oriented design that serves as a knowledge representation of the text's context (Campbell et al., 1998). C. K. Ogden's (1943) semiotic triangle provides the theoretical foundation for NLP to process appropriate meaning of a text's context. A central premise to this theory is that humans deduce meaning from symbols by comprehending the context in which the symbols are used. Machines cannot process relationships of text within a context like humans. To accomplish this, IE methods process semantics by accessing an ontology or reference that identifies relationships between textual terms that help determine semantics (Samsonovic & Ascoli, 2010). However, usage of textual terms changes as the context in which language changes. The term 'handcuffs' has a meaning easily understood in the context of police reports. To the contrary, 'handcuffs' is a term not easily understood in the context of radiology reports. A corpus, therefore, becomes essential to interpretation of a text's meaning by providing a context of reference.

Besides addressing the issue of context, Ogden's (1943) semiotic triangle also provides a theoretical foundation that syntax or text can indirectly represent the natural world and bypass the thought process (see Figure 1).

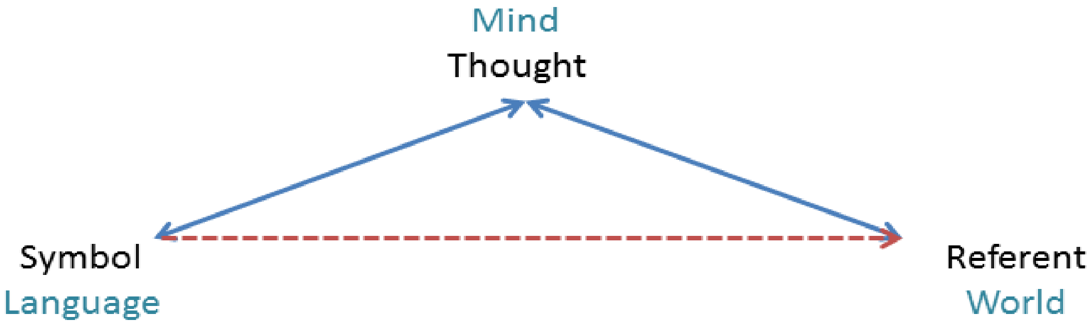


Figure 1- Ogden's Semiotic Triangle

This lends well to NLP because semantics or meaning is mirrored in the syntax of a sentence. This indirect relationship has seen the development of two methods of IE. One uses statistical semantics to infer through probability what free text means based on syntactical analysis (Caviedes & Cimino, 2004; Fan & Friedman, 2007; Pedersen et al., 2007). Another IE method uses propositional logical methods to deconstruct free text into syntactical phrases whereby the phrases become propositions and relationships are formed among these propositions in order to infer meaning (Roth & Yih, 2001). PAS has been used to create propositions (Pugeault, Saint-Dizier & Montreil, 1994).

The theoretical contribution of propositional logic to NLP plays an important role in how semantic are processed. English is unsuitable for computer computations to perform deductive reasoning (Gallier, 1986). Propositional logic becomes a way to express the syntax and semantics of English in order for a machine to deduce logic. With this method, text is assigned a unique variable ('Text Assignment' columns in Table 7) followed by assignment of that unique variable to a variable with meaning in some knowledge source ('Logical Form' columns in Table 7). As a formal language that uses sentential calculus to analyze well formed sentences of a language, symbols come to represent text and inference is made by examining the pattern of those symbols (Johnson-Laird, 2010). The output pattern of symbols becomes meaningful to humans when the representation of the symbols is translated back into text. This method of symbolic reasoning fits well for semantic relatedness because the output pattern of symbols can be compared as logical or not. However, it remains to be seen if propositional logic can be used to determine semantic similarity that seeks to test for directionality between senses of two different PAS frames.

<table>
<tr><td>Sentence1
(CharPAS1$)
(CharSentence1$)</td><td colspan="3">The gray white matter differentiation of the brain is normal.</td></tr>
<tr><td>Sentence2
(CharPAS21$)
(CharSentence2$)</td><td colspan="3">The gray/white matter interface is preserved.</td></tr>
<tr><td colspan="2">PAS</td><td colspan="2">Sentential</td></tr>
<tr><td>Text Assignment</td><td>Logical Form</td><td>Text Assignment</td><td>Logical Form</td></tr>
<tr><td>CharPAS1$(Rel) = A
CharPAS1$(Arg0) = B
CharPAS1$(Arg1-X) = C</td><td>A & B & C = **D**</td><td>CharSentence1$ = H</td><td>H = **D**</td></tr>
<tr><td>CharPAS2$(Rel) = E
CharPAS2$(Arg0) = F
CharPAS2$(Arg1-X) = G</td><td>E & F & G = **D**</td><td>CharSentence2$ = I</td><td>I = **D**</td></tr>
<tr><td>Therefore</td><td>A = E, B = F, C = G</td><td>Therefore</td><td>H = I</td></tr>
<tr><td colspan="4">**D** is typically referent to entry in some knowledge source referred to as the
PROPOSITION</td></tr>
</table>

Table 7--Propositional Logic Method for PAS and Sentential

2.2 Research Questions

Propositional logic lends to an assumption in this study: If a set of sentences are deemed to express the same meaning by humans, then the logical representation structure of those sentences should be the same. With that assumption, this study seeks to answer the following questions:

1) Can a PAS frame be annotated for each sentence that allows for a test of semantic similarity?

   Specifically, the following questions about annotation of a PAS frame will have to be answered:

   *Can PAS frame illuminate uniqueness of and adapt to nuances of a radiology corpus?

   *Can a consistent routine be applied to create complete sentences from incomplete sentences in a radiology corpus?

   *Is Propbank knowledge base sufficient to cover sense of verbs used in radiology corpus?

   *Can PAS frame show a schema to match and compare corresponding syntactical phrases from different sentences of synsets formed from radiology corpus?

2) Does the representative logical propositional sentence represent the synset of sentences through a semantic similarity measurement that considers directionality?

While the method for the second question will be explained, data collection and analysis in this paper will be limited to question one.

2.3 Scope and Limitations

This study is exploratory in nature and seeks to develop a methodology for semantic measurement that may need further testing. The sample set is small and composed of large synsets and is not representative of an entire corpus. Therefore, a full evaluation of the corpus’s syntactical, semantic, and language characteristics cannot be assumed. An analysis of this study will determine if the annotation process warrants a larger scale study.

CHAPTER THREE: METHODOLOGY

This study will use a traditional annotation schema for clinical corpus (Xia & Yetisgen-Yildiz, 2012). Annotation of clinical text depends on the manual interpretation of trained experts (clinicians, NLP researchers) to mark and code unstructured text. The team for traditional clinical annotation is smaller due to the expense involved with recruiting physicians and expert clinicians often with no formal training in linguistics and NLP. Because this study has no budget, annotation of the PAS frame will be done by a graduate healthcare informatics student with 4 years of nursing critical care experience. Traditional annotation schema relies on procedural guidelines. Table shows the guidelines adopted from Xia & Yetisgen-Yildiz, used for this study and where in the study each procedure has been addressed:

	Procedure	Where Addressed
1.	Define annotation task based on clinical need	1.1
2.	Select data to be annotated	3.2
3.	Get IRB	Not Applicable
4.	Write annotation guidelines	3.3.1
5.	Create good annotation tools	3.3.1
6.	Annotate text	3.3.1
7.	Analyze annotation—Identify patterns	4.0 & 5.0
8.	Assess annotation method for next phase	6.0

Table 8--Guidelines for Annotation Schema

3.1 Type of Research Design

This study investigates using methods in NLP never used in measuring sentential semantic similarity; it does not have a dependent or independent variable; and it will assess if the Propbank knowledge source, annotation tools and processes can guide humans to successfully create a PAS frame in a radiology corpus. For these reasons, the design of this study meets the scientific definition for an exploratory, non-experimental, methodological study (Nieswiadomy, 2008). However, this study will also look for patterns across synsets from a radiology corpus to understand if there is a possible algorithm for scalability within the PAS frames so that appropriate content is compared. Pattern recognition of meaning in text is a content analysis process and is argued for NLP to be a qualitative perspective (Yu & Jannasch-Pennell, 2011). Due to only one annotator, this study will not have a inter-rater score to determine validity and reliability. Although inter-rater score is recommended before moving onto larger annotation (Roberts et al., 2008), analysis and findings from the one annotator of the ten synsets should still produce valuable findings from which to assess feasibility of larger annotation.

3.2 Dataset

A convenience sample from a prior study (Friedlin et al, 2011) is used as the data set. The data set represents ten synsets. Each synset has 100 randomly selected sentences (n=1000) from synsets with more than 1,000 sentences of an ongoing annotation project. A propositional sentence serves as the representing the semantics for the synset (see Table 9).

	Propositional Sentence of Synset
1.	The endotracheal tube is above the carina.
2.	There is no pneumothorax.
3.	There is a left lower lobe pulmonary infiltrate(s).
4.	The pulmonary vessels are prominent.
5.	A posterior anterior (PA) chest x-ray was performed.
6.	The gray white matter differentiation of the brain is normal.
7.	The intervertebral disc heights are normal.
8.	There are pelvic phlebolith(s).
9.	There is small vessel ischemic disease of the brain.
10.	The lungs are diffusely hazy bilaterally.

Table 9--Proposition Sentence of 10 Synsets

3.3.1 Data Collection—Annotation

Reliable annotation depends on tools and guidelines (Roberts et al, 2009). Tools comprise the manuals, knowledge references, and notation aids from which the project guidelines form. NLP researchers review the resources, modify and extract from these resources elements necessary to understand and complete annotation task, and create notation aids to record/document annotation extractions from text. This project reviewed the manual for Propbank annotation (Babko-Malaya, 2005) as the primary source to develop guidelines and used the Unified Verb Index (2012) to facilitate construction of PAS frames. From these resources, a notation tool was developed as a worksheet in Excel to structure data collection (See Appendix A). An aim of these resources was to establish guidelines annotation rules for passive sentences, compound sentences, and understanding of argument modifiers. It was important for the worksheet to have a structure that would allow for analysis of secondary research questions (i.e., implied predicates and no matching Roleset ID). It also reduces the complexity of linguistic knowledge required allowing for experts of the corpus content to focus on the text. The successfulness of using the worksheet depends on three assumptions: 1) no predicate can have more than five arguments, 2) all text of sentence is annotated, and 3) no modifier is used more than once.

Annotation resources used to create guidelines should strive for an annotation recipe (Roberts et al., 2009). An annotation recipe formalizes a framework for consistent data collection but refrains from being restrictive to allow for investigation of text. The annotation process adopted for this study is the following:

1. Read sentence in its entirety without making entries in worksheet
2. Read sentence 2nd time and enter predicate in worksheet
3. Look up predicate in Propbank and enter appropriate Roleset ID into worksheet
4. Copy corresponding argument descriptions from Propbank into corresponding entry in worksheet
5. Go through sentence and copy text phrase in sentence that corresponds to argument descriptions and paste on worksheet
6. Go through sentence and copy text phrase in sentence that corresponds to a modifier and paste on worksheet selecting modifier type from drop down
7. Review annotation that all text of sentence has been annotated into worksheet
8. Record any questions, uncertainties, ambiguities, problems, or issues in comment section of worksheet

The distribution of data should follow an iterative process whereby completed annotation is reviewed to address changes in guidelines based on discoveries from the corpus's text.

Appendix B presents the manual used in this study to train and familiarize annotators with the tools, references, and process. The manual originally assumed that two radiologists would fulfill the annotator responsibilities but registered nurse student with four years of critical care experience. After development of resources, annotation process presented in manual took 60 hours.

3.3.2 Data Analysis/Validity

Because this study had only one annotator, there are no agreement metrics to report. Holden (2010) argues that annotation studies do not necessarily require a multiple analyst approach to assess validity. A study warrants a single research approach if validity can be assessed through alternative methods. The alternative methods in this study that meat criteria for a single researcher approach are:

1) Use of Propbank (Babko-Malaya, 2005) and Unified Verb Index (2012) to interpret raw data
2) Reporting of coding changes resulting from challenges to subsequent data
3) Historical record of all annotation decisions available for external auditing

The annotator did keep a historical log of significant events that challenged annotation. These events perhaps could be considered comparable to analysis of annotator differences (Roberts et al, 2009). These events are described in section 4.0. Analysis will report relative frequencies for the following:

1. Predicates for each synset
2. Predicates overall
3. Implied predicates

4. No appropriate predicate in Unified Verb Index

The major data analysis approach will be content analysis from the perspective of text mining. This study will employ steps presented in Yu, Jannasch-Pennel & DiGangi (2011):

a) Researcher examines textual data and takes notes
b) Researcher performs data reduction in way to answer research question
c) Researcher organizes, arranges, and displays condensed data—identifying themes, patterns, connections and omissions
d) Researcher revisits data many time to verify and confirm themes and patterns identified

The content analysis approach will be used to answer the research question in this study if a pattern exists to allow for scalability of PAS. Section 4.2 describes in detail the guidelines used for content analysis.

CHAPTER FOUR: RESULTS

Findings will begin with a presentation of overall predicate usage in project corpus followed by a discussion on method developed for content analysis. Then, each synset will be presented that will include frequency proportions on predicates and discussion about linguistic, syntactic, or language particular to that synset.

4.1 Predicate Usage Across Synsets

Fig 2 shows a graphic interpretation of major predicates compared to aggregation of less frequently used predicates. Fig 3 presents a frequency count of all predicates used on project corpus. There were 33 unique predicates with 'be' being the most frequent.

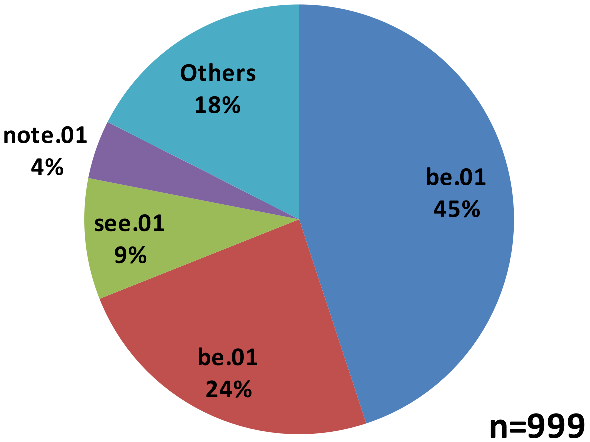


Figure 2--Aggregation of Infrequent Predicates (Others) to Frequent

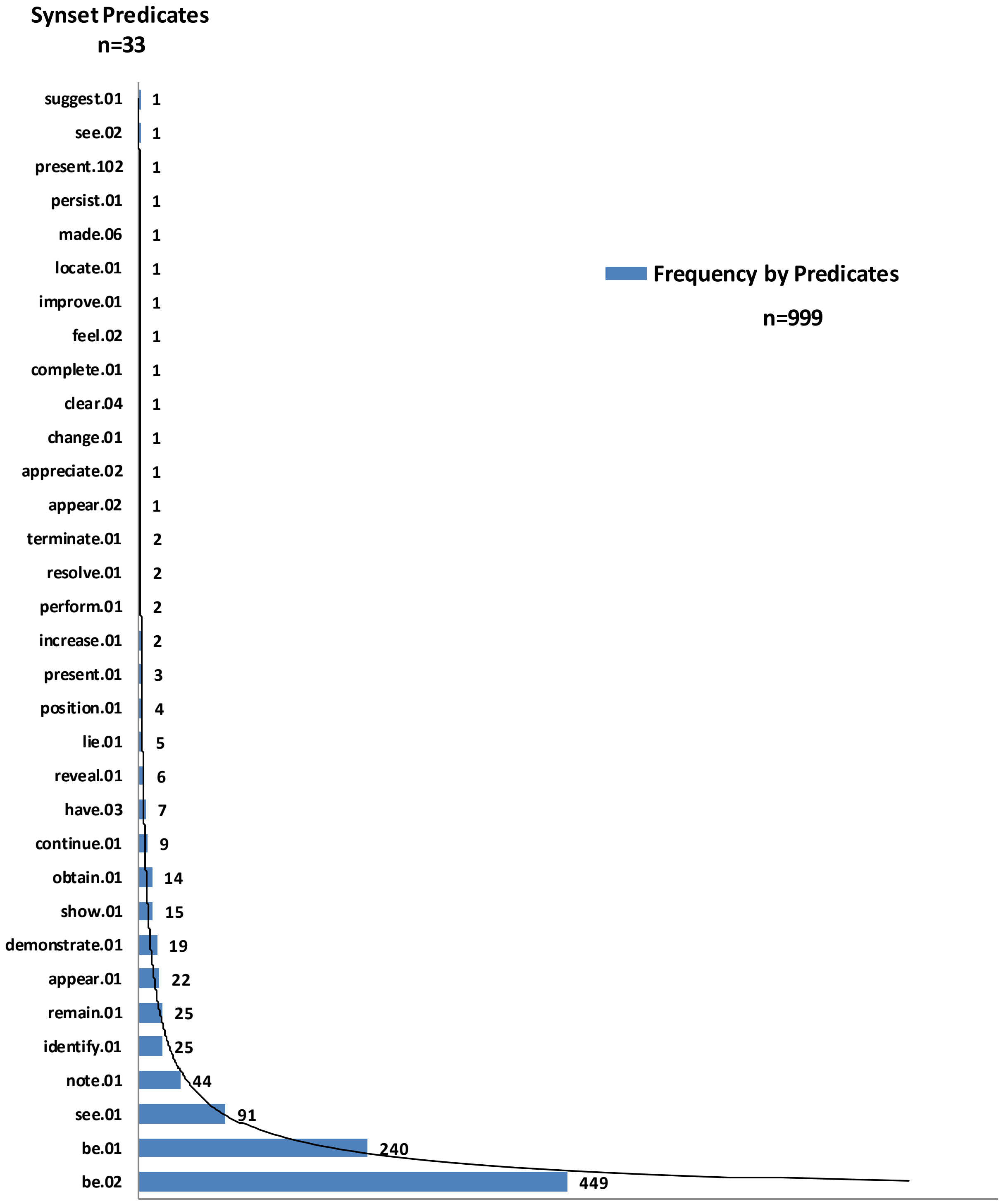


Figure 3--Frequency Distribution of Predicates Across 10 Synsets

Significance to these frequencies is the dominance of the two predicates, be.01 and be.02 which account for 2/3 of predicates expressed in project corpus. Be.02 represents the expletive form of “There are/is” and has one argument role. Its sense implies that <something> exists. This differs with be.01 which represents a copula statement. It does not imply that something exists. This makes their PAS frames different (see Table 10) and elucidates the general problem with using PAS frames as foundation to measure sentential semantic similarity of differing predicates (Discussed in 2.0). If these two predicates dominate a radiology corpus, a method that systematically and consistently identifies concepts needed to compare from different role labels will have to represent these two varying predicates usage in language. In the example in Table 10, the concept of prominence is expressed differently in two sentences—one expresses it as an adjective the other as a noun. For the concept of vessels, both use an identical form, ‘vasculature’, but annotation places ‘vasculature’ in Arg1 for be.01 and as a ‘cause’ modifier for be.02. To do a semantic comparison, the SRL of Arg1 for be.01 would have to be compared to ArgM-CAU for be.02 and the SRL of Arg2 for be.01(‘prominent’) would have to be compared to Arg1 SRL of be.02 (‘prominence’). As a whole, the pieces put together express semantic equivalence, but separated, a pattern has to be developed that could argue the cross comparing of SRLs. The investigation of this pattern for a consistent method is explained in the next section.

	The pulmonary vasculature is prominent			
	There is prominence of the pulmonary vasculature			
	be.01		be.02	
	Arg Desc	PAS Frame	Arg Desc	PAS Frame
Arg1	topic	vasculature	thing that is	prominence
Arg2	comment	prominent		
Argm-LOC		pulmonary		pulmonary
Argm-CAU				vasculature

Table 10--Scalability Problem of PAS

4.2 Content Analysis Method

When the PAS frames for two semantically equivalent sentences differ, it is intuitive for humans to forego the SRL and match content of one PAS argument to that of another. This is not so with NLP software. Rules have to be established that represent the patterns for how to compare appropriate content of differing PAS frames. An annotated corpus considered a "goldstandard" serves as the source from which to investigate these patterns for establishing rules. The "goldstandard" in this project is the synsets that have already been determined by humans to be semantically equivalent sentences expressive of a unique concept (see Table 9). This project will use a modified method described in the Miyao et al. (2006) study for building PAS frames to discover patterns within the dataset for cross comparing of differing PAS semantic roles. The method begins by identifying the noun phrase (NP) and verb phrase (VP) of a sentence (See Figure 4).

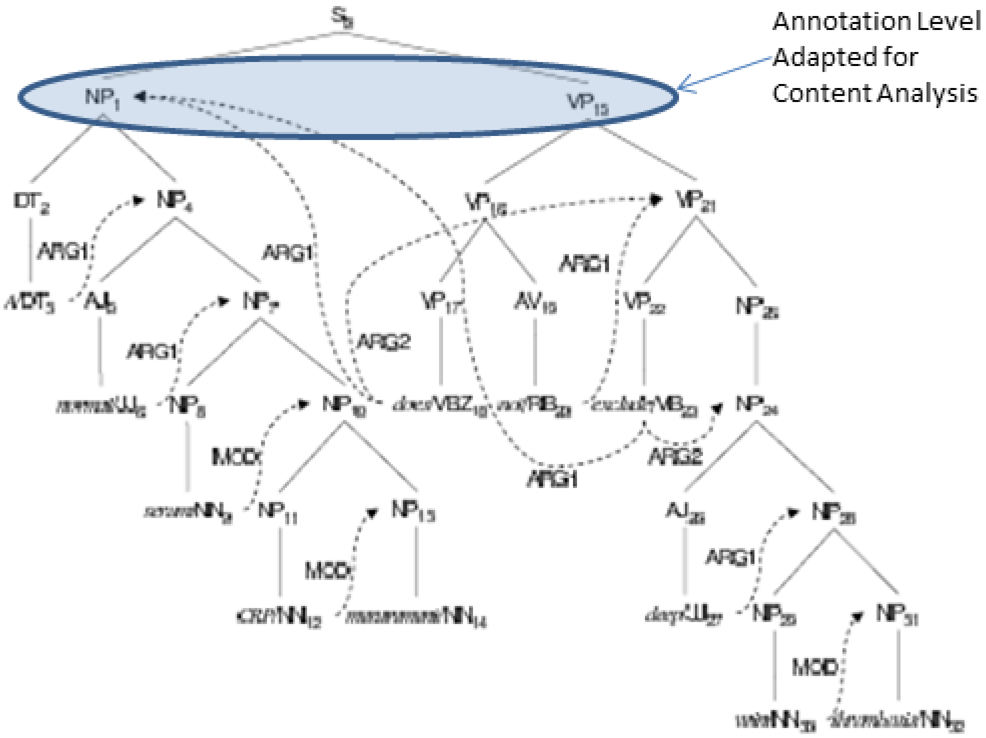


Figure 4--Annotation Process in Miyao et al.

A complete sentence should have a noun and verb phrase unless the sentence uses the predicate be.02 or is passive. If the sentence uses the predicate be.02, there is only a noun phrase. If the sentence is passive, it may only have a verb phrase. Each phrase may have modifiers of the phrase’s subject. In the Miyao et al. study, the text of these phrases is annotated further into adjective and adverb phrases. In this study, annotation of the phrases will simply identify the subject of the phrase and its modifier(s) (See Figure 5). With this method, it will be possible to see how content of the phrases of proposition sentence are assigned to arguments based on predicates of candidate sentences and if there’s a consistent patterns within the sysnet to cross compare differing SRLs. The analysis of this pattern will be presented in the form of a flowchart.

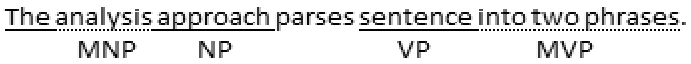


Figure 5--Identifying Noun/Verb Phrases and Modifiers of Proposition Sentence

4.3 Individual Synset Findings

For each synset, a propositional sentence indentifying noun and verb phrases (a solid underline) and their modifiers (dotted underline) are first presented. Then, there will be two frequency proportion summaries. One for predicates used in synset and one for predicates used to make an incomplete sentence complete. The last figure will be a flowchart showing how the content of the phrases and modifiers are related among the predicates. The section will conclude with a discussion about any linguistic, syntactical, and/or discoveries pertinent to research questions.

4.3.1 Synset 1

Synset Proposition: The endotracheal tube is above the carina.
NP VP

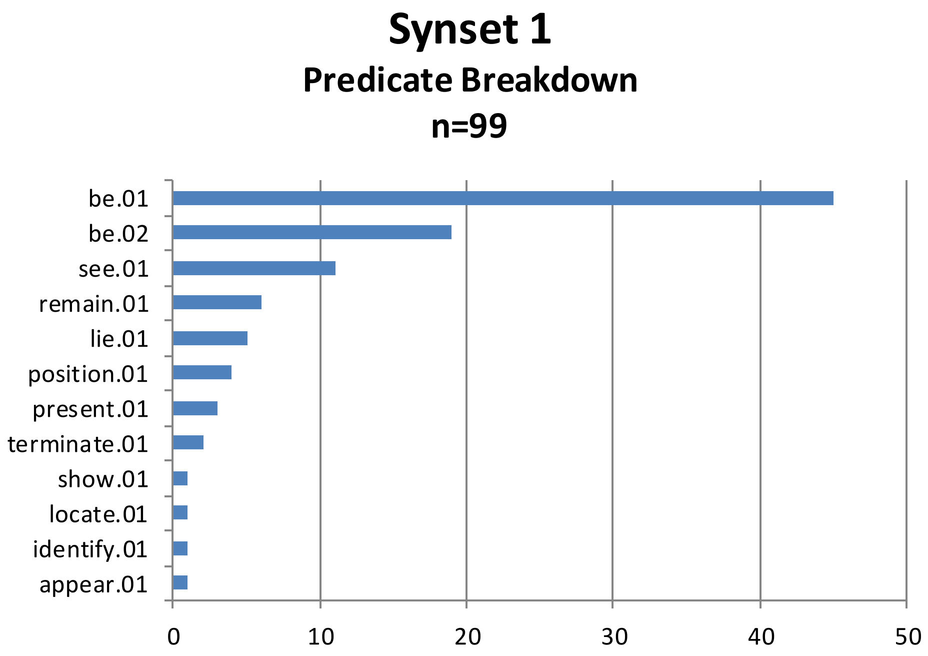


Figure 6—Synset 1 Predicates

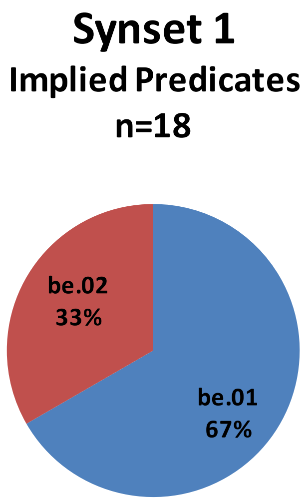


Figure 7—Synset 1 Implied Predicates

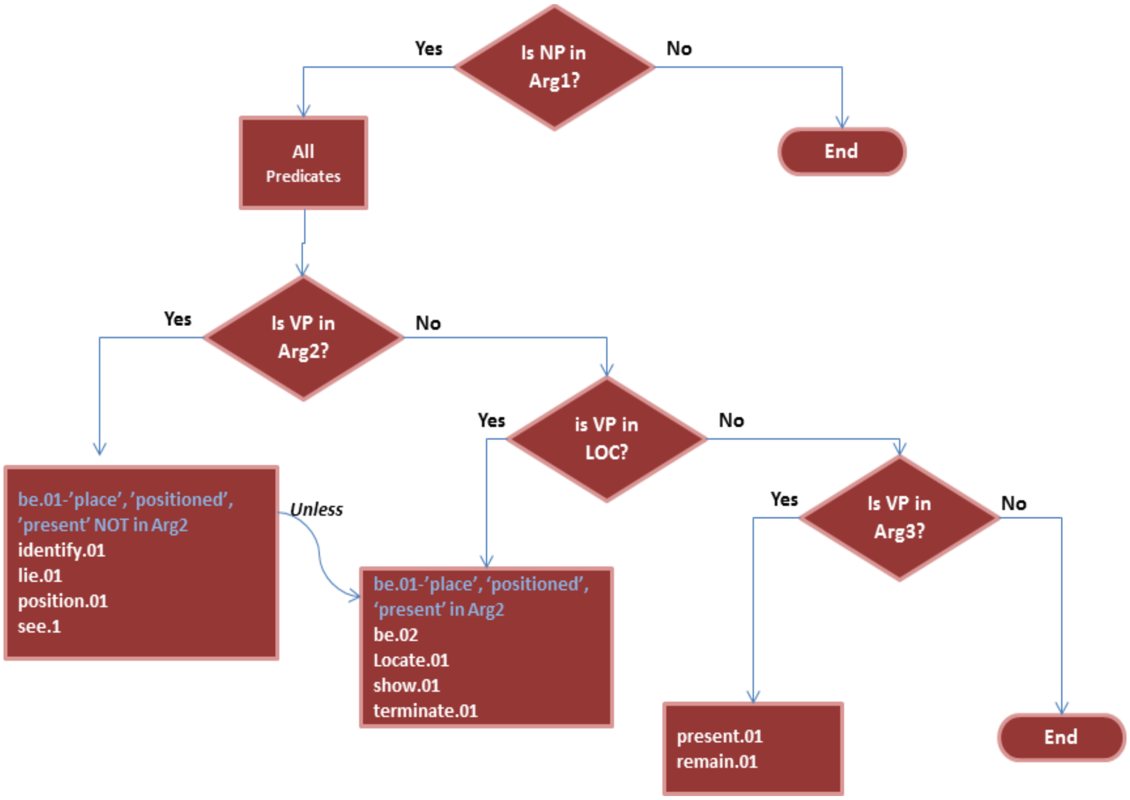


Figure 8—Synset 1 Argument Annotation Flowchart

Discussion:

No matter the predicate, the NP was found to be in Arg1. For the VP, predicates predict what argument contains the concept, except for be.01. Be.01 could be in 1 of 2 arguments depending on if the terms 'place', 'positioned', or 'present' are in Arg2. For semantics within the synset, granularity may be the only problem. Some sentences used 'tip' to be more specific about location of Endotracheal tube. For implied predicates, this synset is one of two where the predicate be.01 was needed more than be.02 to make a complete sentence.

4.3.2 Synset 2

Synset Proposition: There is <u>no pneumothorax.</u>
neg NP

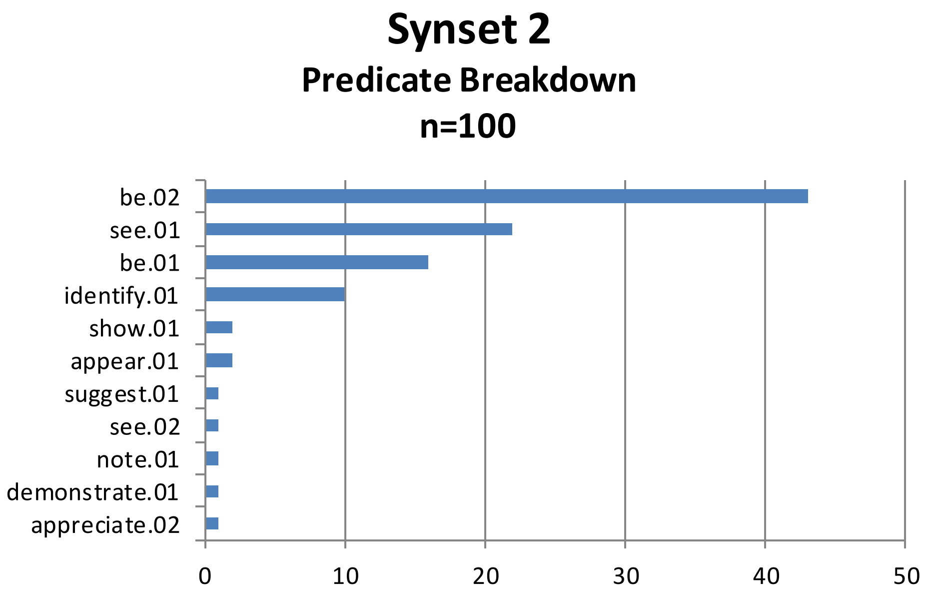


Figure 9—Synset 2 Predicates

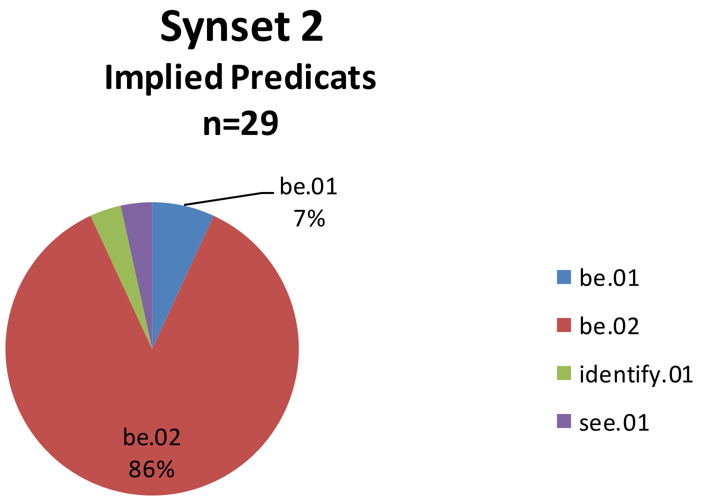


Figure 10— Synset 2 Implied Predicates

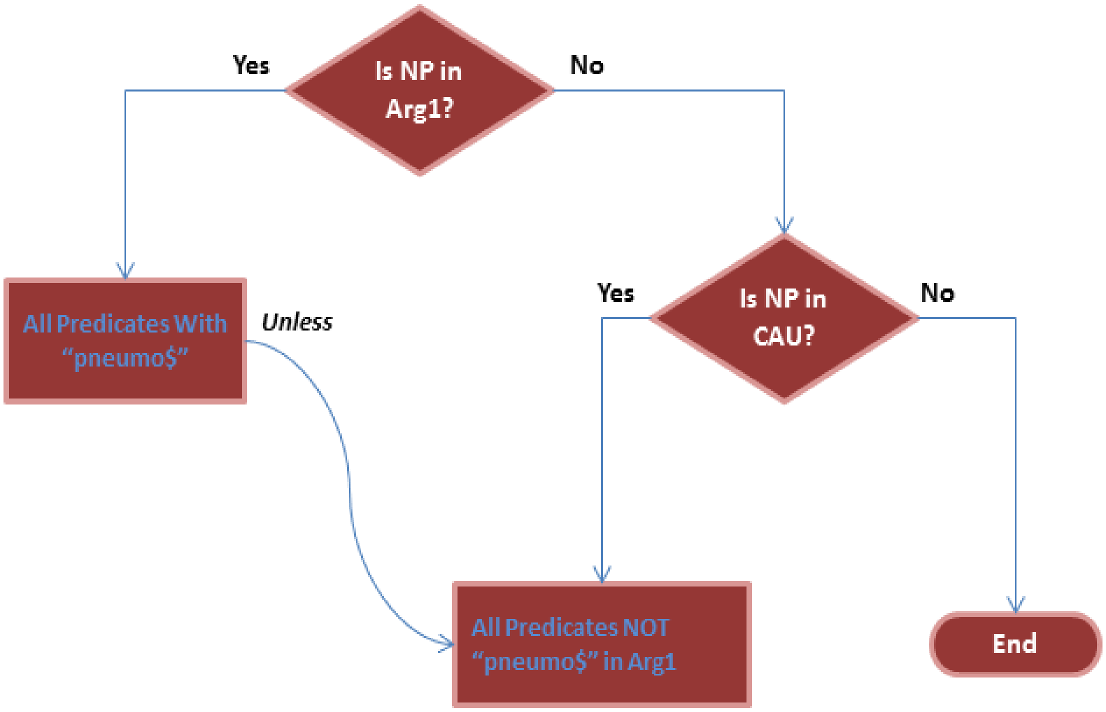


Figure 11—Synset 2 Argument Annotation Flowchart

Discussion:

This is the only sysnet with a negation, which is annotated to ArgM-NEG. The NP is found Arg1 for all predicates, unless Arg1 has a term beginning with the characters 'pneumo'. In which case, the NP is found in ArgM-CAU. There were two terms of negation, 'no', and 'without'. Other than singularity and plurality expression of pneumothorax, this synset had no semantic discrepancies.

4.3.3 Synset 3

Synset Proposition: There is a left lower lobe pulmonary infiltrate(s).
MNP NP

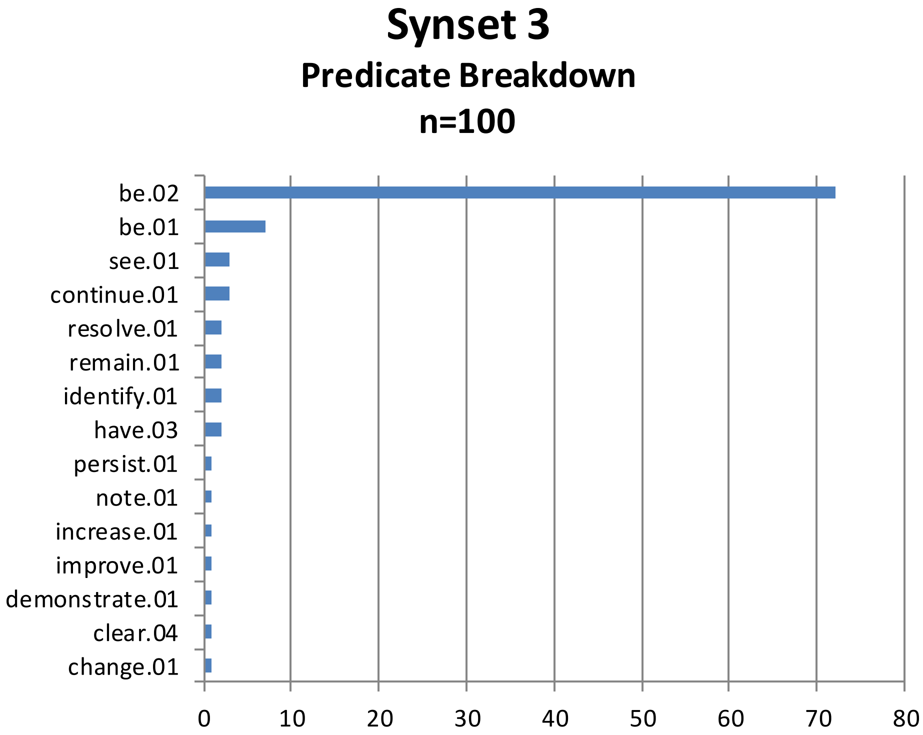


Figure 12— Synset 3 Predicates

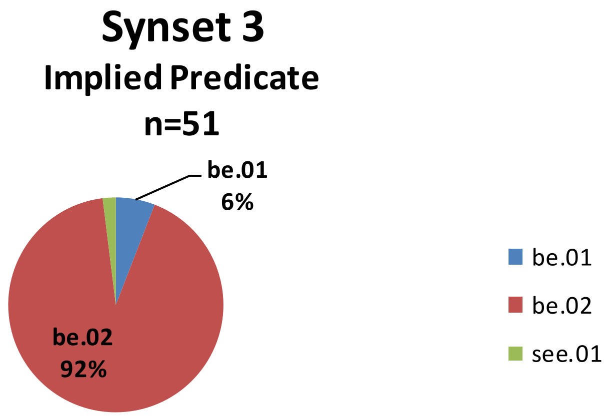


Figure 13— Synset 3 Implied Predicates

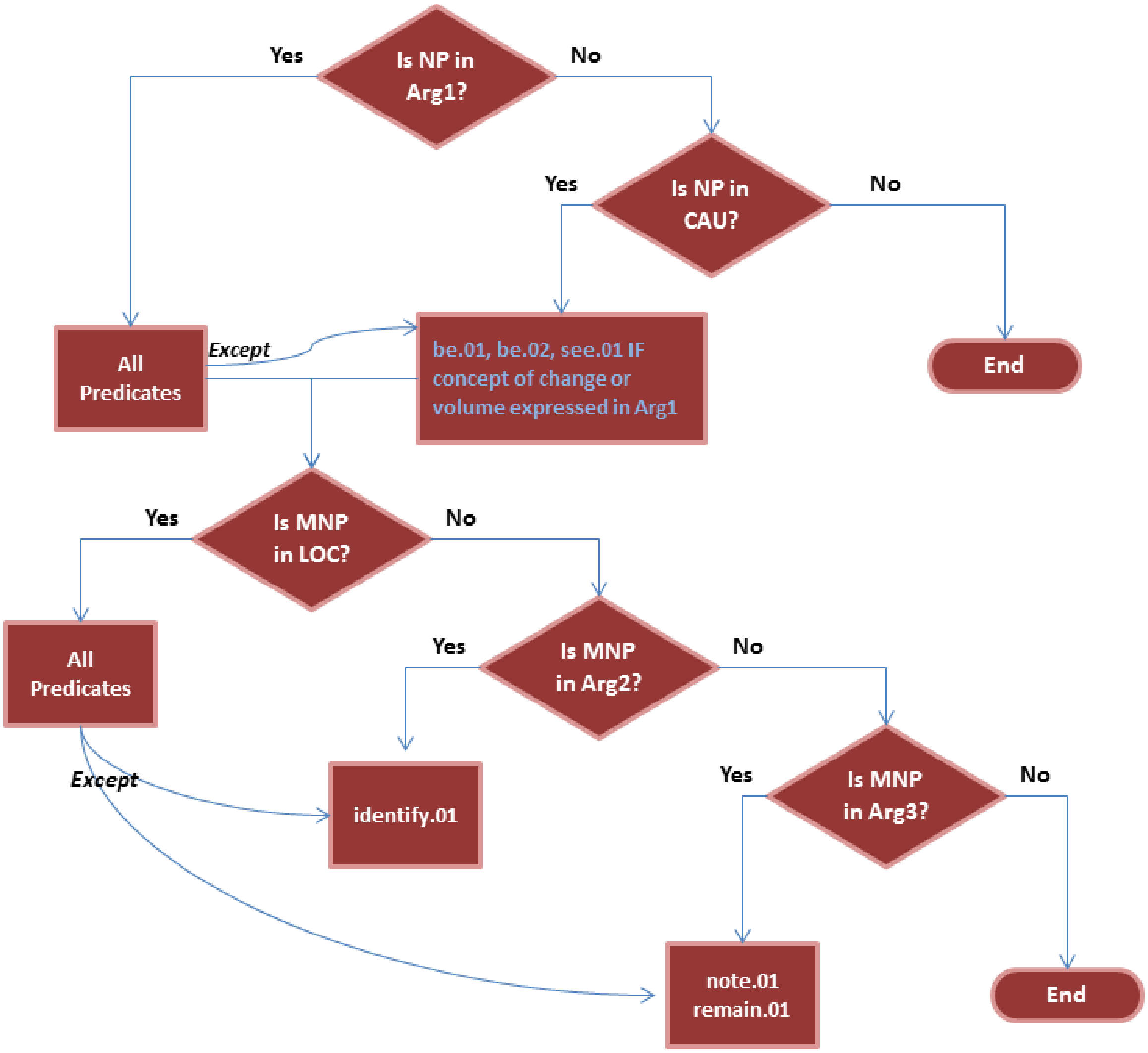


Figure 14— Synset 3 Argument Annotation Flowchart

Discussion:

Unless the concept of change or volume is expressed in Arg1 for predicates be.01, be.02, and see.01, the NP is found in Arg1, otherwise it is in ArgM-CAU. This synset will test the semantic similarity of two terms, 'infiltrate' & 'atelectasis'. If 'atelectasis' is used, 'infiltrates' is present in argument as well. Notice how this synset has the MNP (anatomical location of left lower lobe) annotated to three different arguments.

4.3.4 Synset 4

Synset Proposition: The pulmonary vessels are prominent.
MNP NP VP

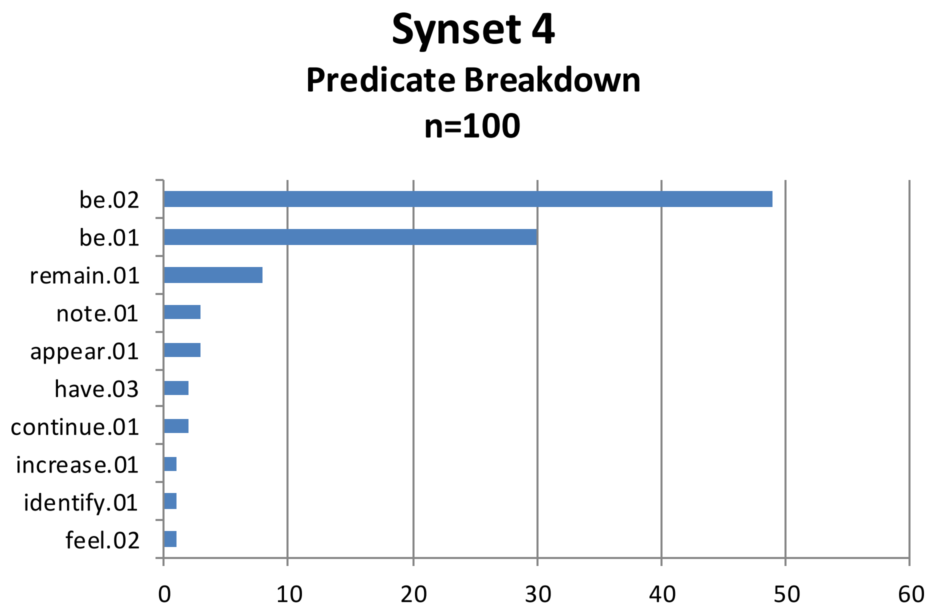


Figure 15— Synset 4 Predicates

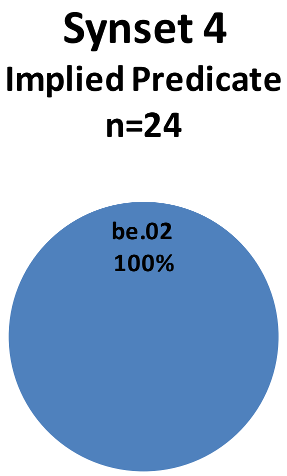


Figure 16— Synset 4 Implied Predicates

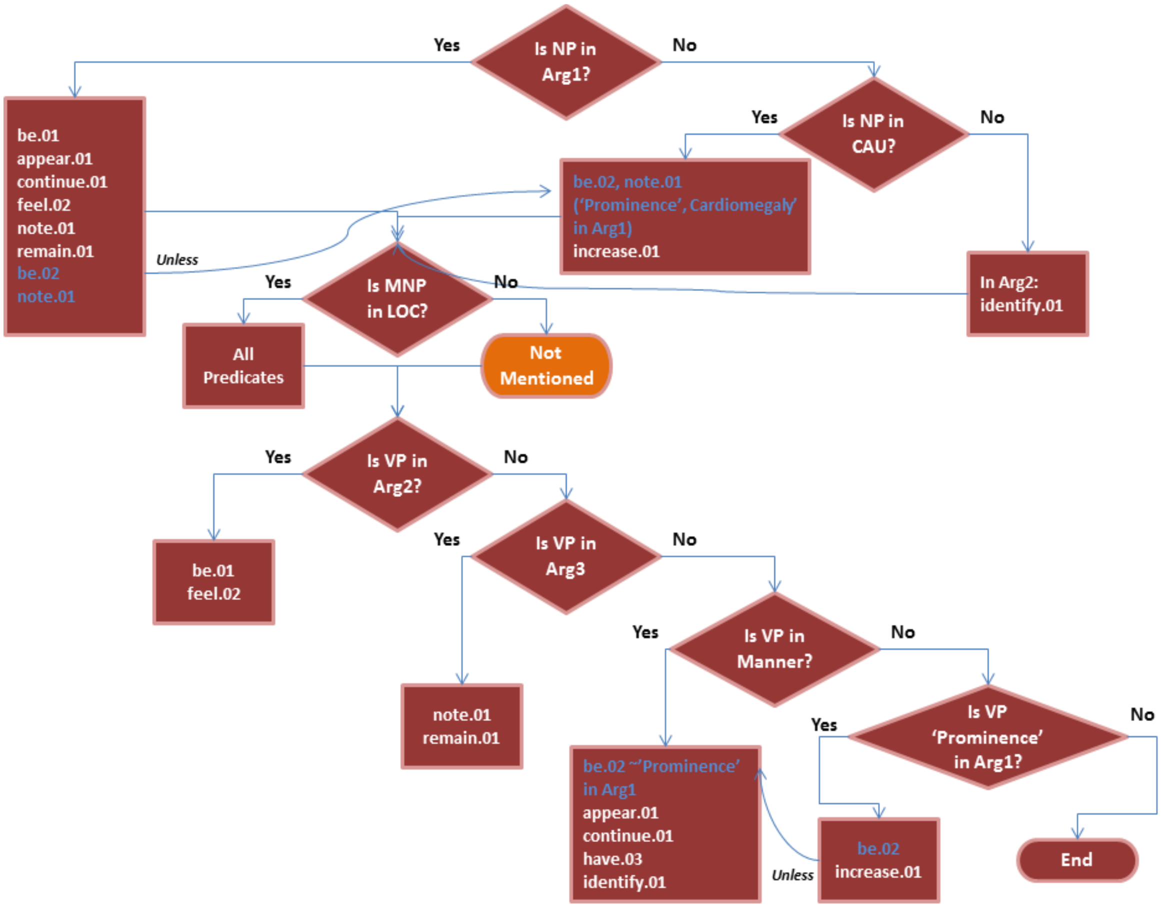


Figure 17— Synset 4 Argument Annotation Flowchart

Discussion:

In this synset, the annotation of NP or VP is dependent on language usage. Two different forms of a term express the concept ‘standing out in order to be seen’: ‘prominence’ and ‘prominent’. Depending on which form is used, the annotation of NP of VP for predicate be.02 is different. For the proposition, ‘pulmonary’ is treated as a modifier (anatomic location) because there could be two kinds of prominent vasculature, pulmonary and systemic. However, some sentences did not use ‘pulmonary’. This should be an interesting test of semantic similarity with the MNP. The NP will test semantic similarity of three terms, ‘vessels’, ‘vascularity’, ‘vasculature’.

4.3.5 Synset 5

Synset Proposition: A posterior anterior (PA) chest x-ray was performed.
MVP VP

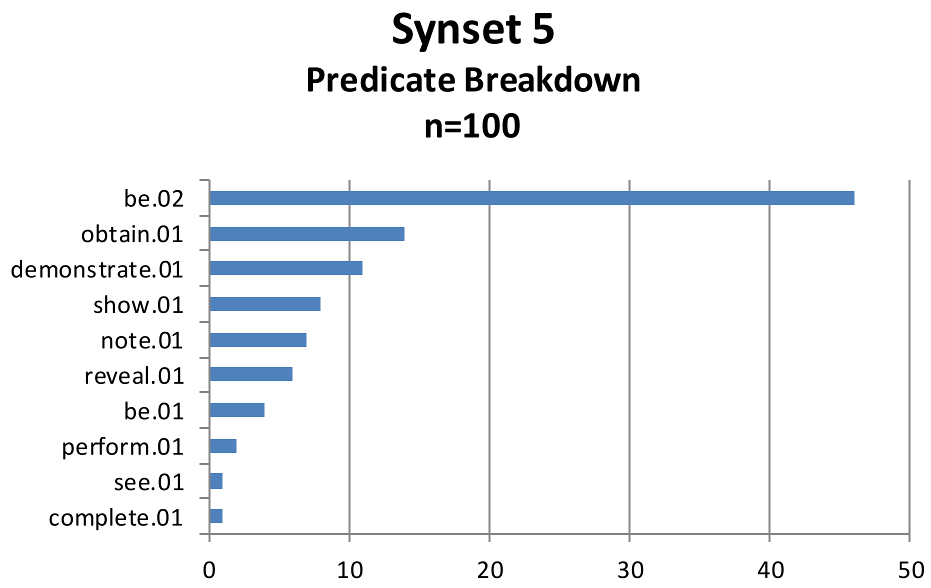


Figure 18— Synset 5 Predicates

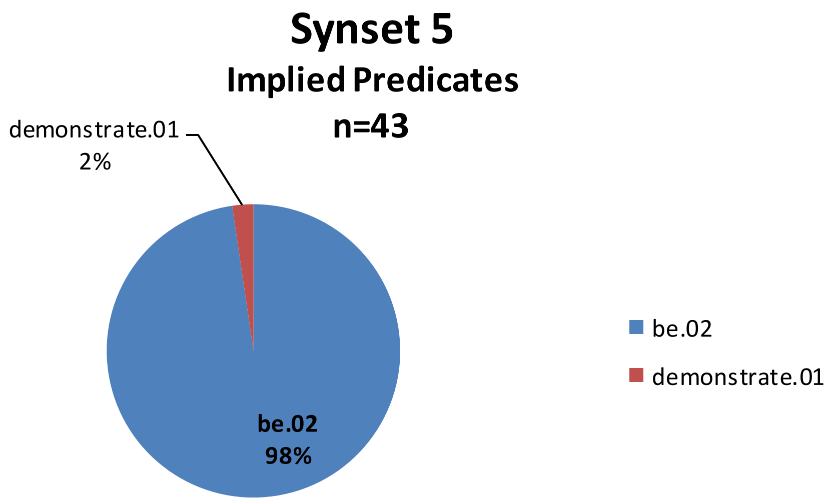


Figure 19— Synset 5 Implied Predicates

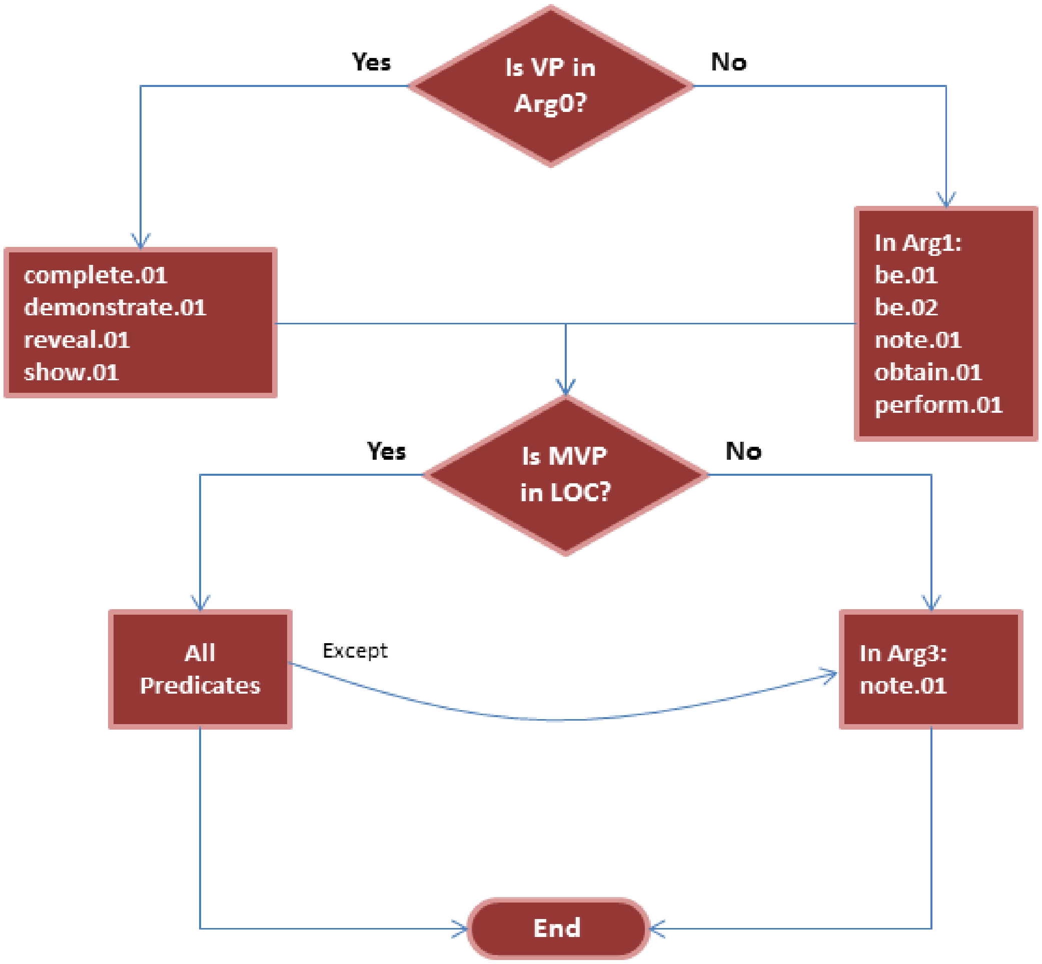


Figure 20— Synset 5 Argument Annotation Flowchart

Discussion:

The proposition sentence of this synset is passive which makes the sentence's subject the VP. There is no NP. This synset had 43% incomplete sentences presenting a syntactical challenge. This is likely due to what was construed as a sentence from the original tagging: 1) text between two '.', or 2) text between '.' & ":'—most likely a section header 'SINGLE PA CHEST VIEW'. The VP subject was annotated as 'chest x-ray' choosing not to make chest a modifier of x-ray. If that had been the case, the locative modifier (ArgM-LOC) could have two annotations, posterior anterior and chest. Very few sentences used ‘x-ray’ which would mean that there will have to be a great deal of implying of VP (a chest x-ray) through context of sentence, as in “PA CHEST : The heart and mediastinum are unremarkable .” An interesting finding will be to see if experts can infer a radiological procedure in this sentence when parsed into a PAS frame. This is the one synet that had text annotated to Arg0.

4.3.6 Synset 6

Synset Proposition: The gray white matter differentiation of the brain is normal.

NP MNP VP

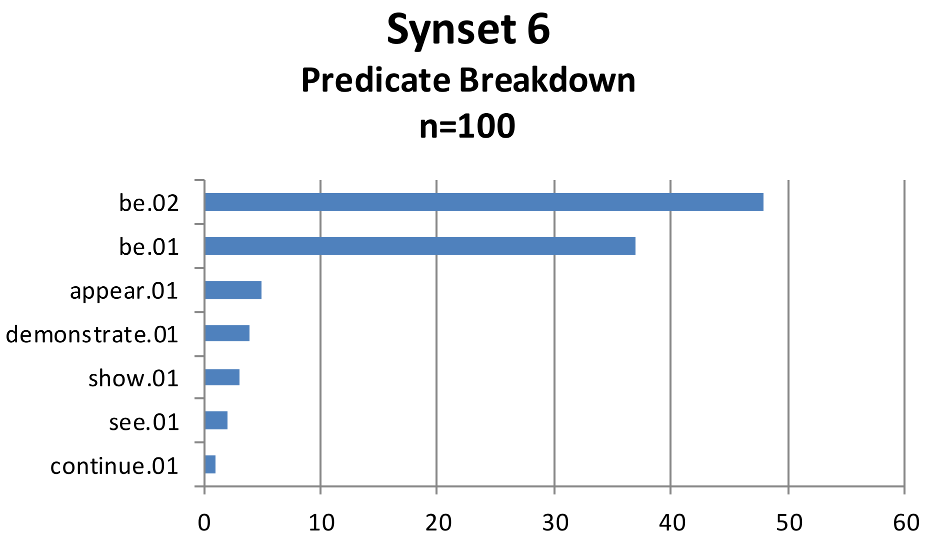


Figure 21— Synset 6 Predicates

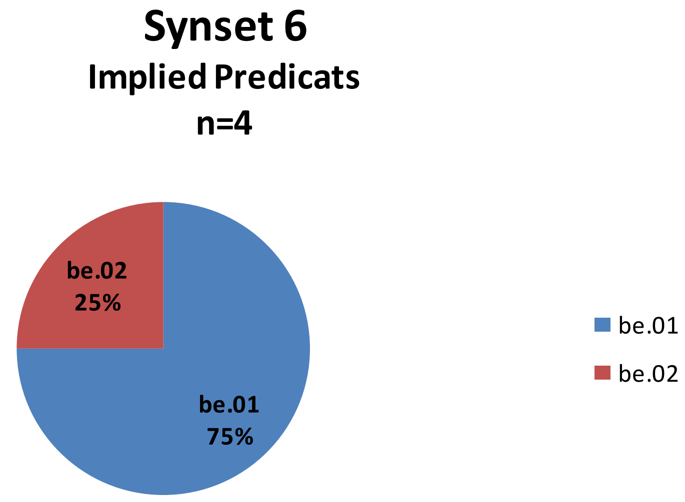


Figure 22— Synset 6 Implied Predicates

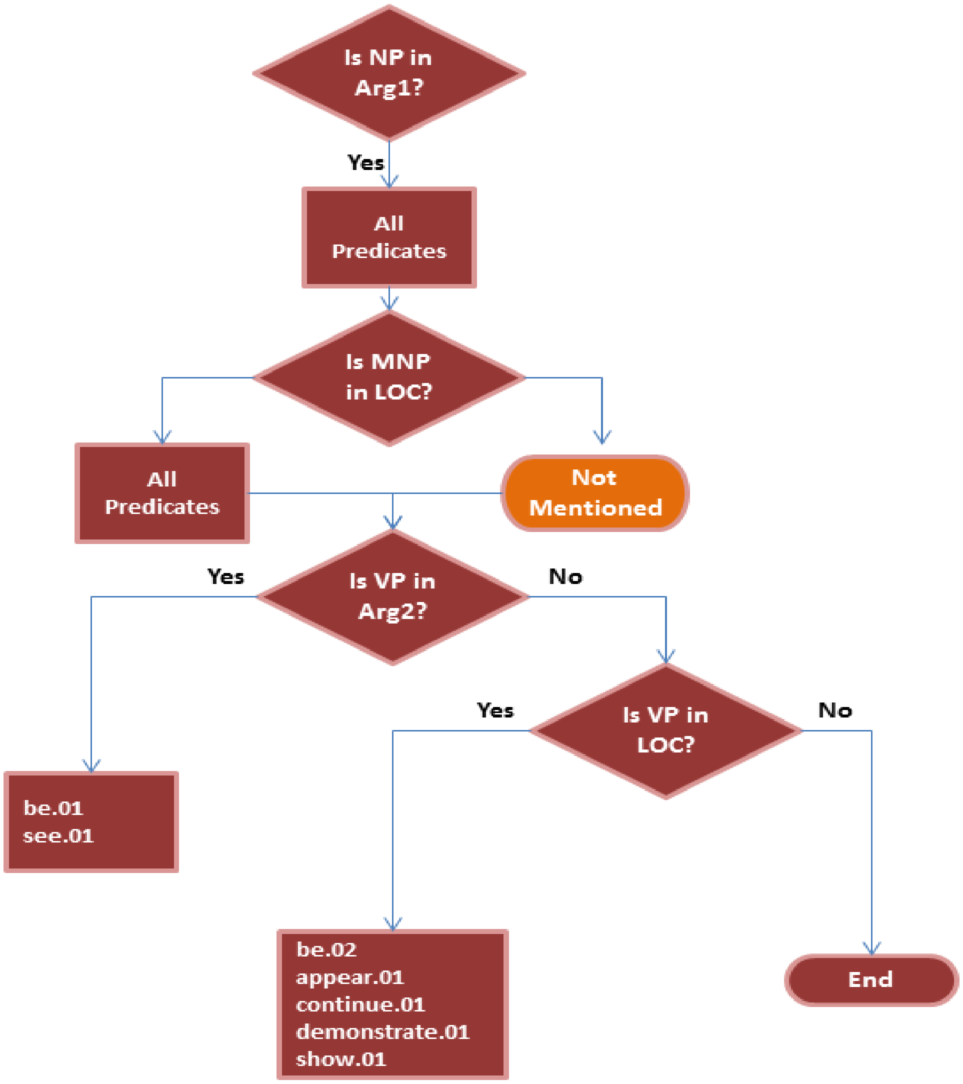


Figure 23— Synset 6 Argument Annotation Flowchart

Discussion:

A big decision for this sysnet was to make the NP subject inclusive of adjectives "grey matter/white matter" for differentiation. Only 7 sentences made reference to the MNP (location in brain). To be semantically similar, the location in the brain will have to be inferred through grey/white matter differentiation.

4.3.7 Synset 7

Synset Proposition: The intervertebral disc heights are normal.

NP VP

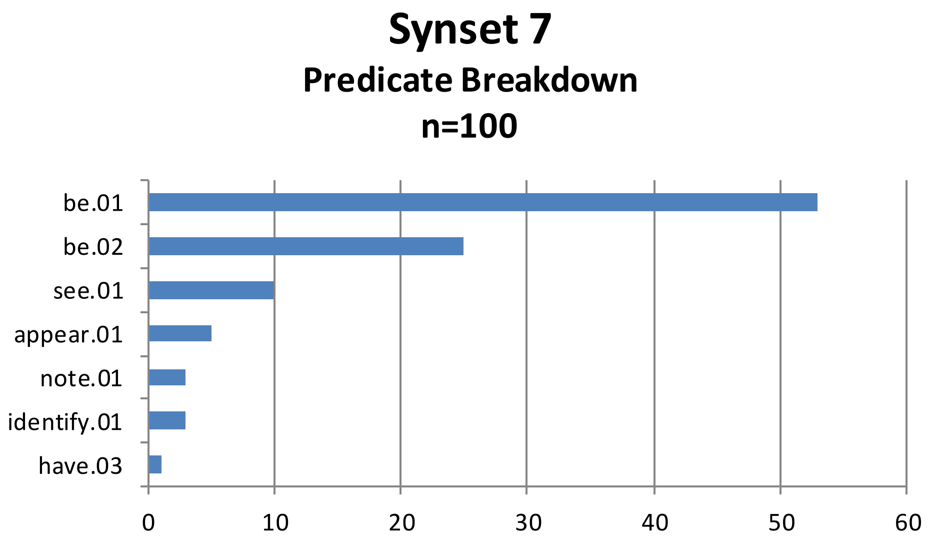


Figure 24— Synset 7 Predicates

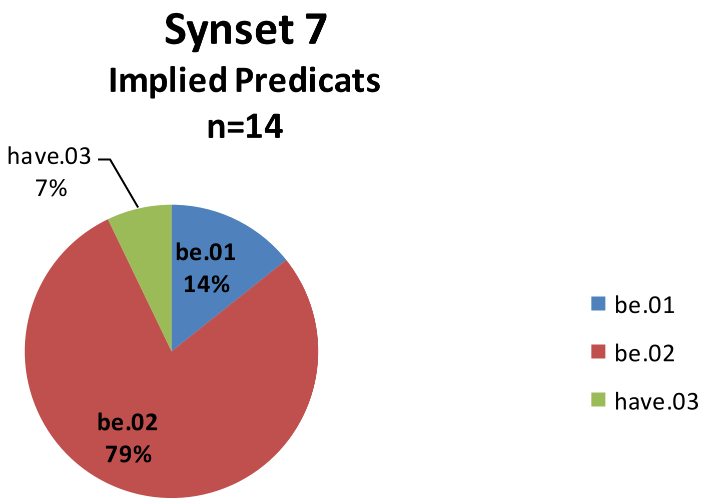


Figure 25— Synset 7 Implied Predicates

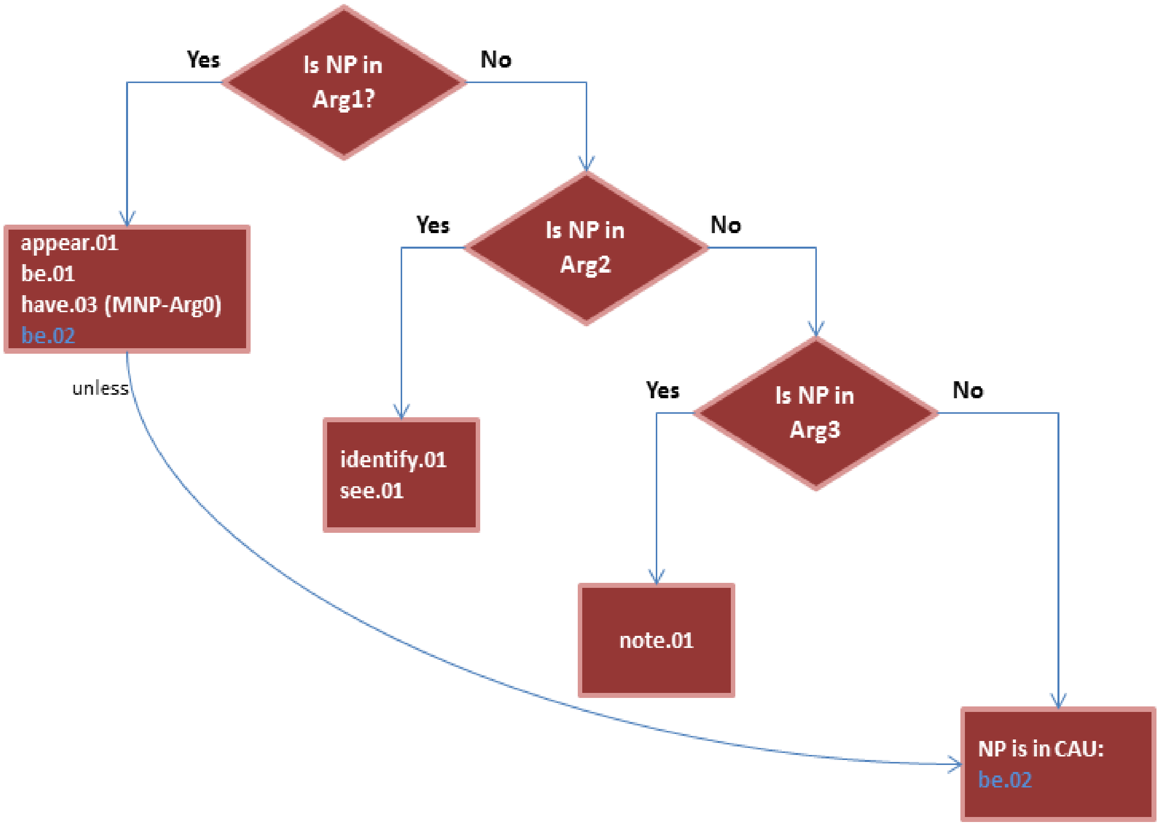


Figure 26— Synset 7 Argument Annotation of NP Flowchart

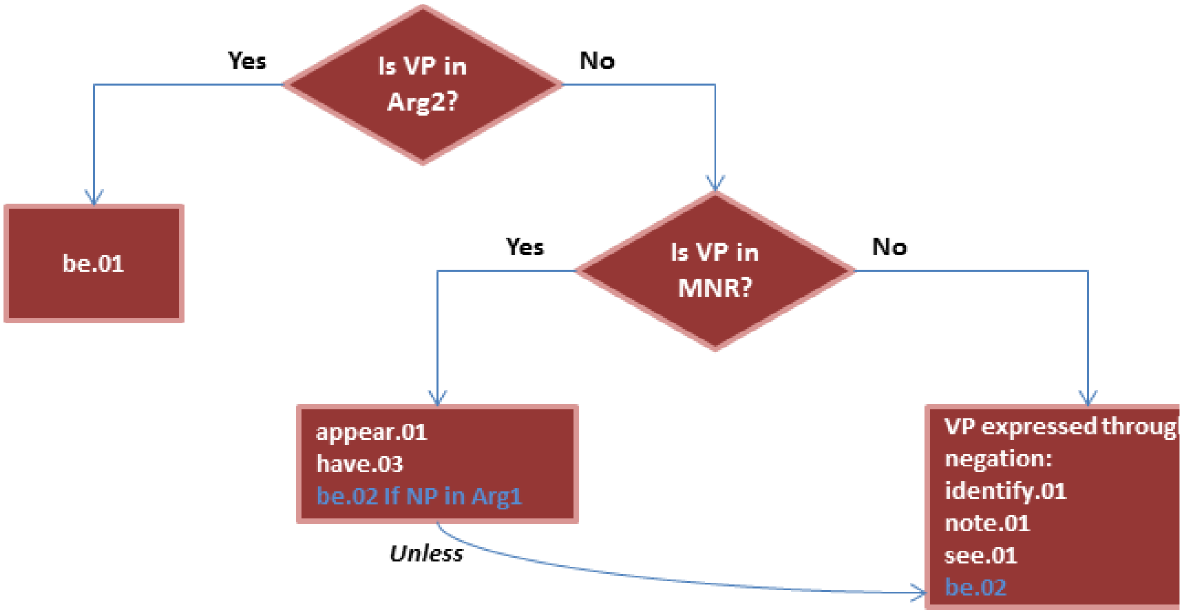


Figure 27— Synset 7 Argument Annotation of VP Flowchart

Discussion:

‘Intervertebral disc heights’ is the full NP. ‘Intervertebral’ is technically a MNP, but ‘disc heights’ is predominantly referring to space between vertebrae. Therefore, ‘intervertebral’ is expressing the same concept. Of all synsets, this synset presents the most diverse semantic challenge within the NP and VP. For the NP, terms are used such as ‘’disc space heights’, ‘vertebral body height’, ‘inner disc spaces’. For the VP, terms are used such as ‘maintained’, ‘preserved’, ‘within normal limits’, ‘no widening or narrowing’.

4.3.8 Synset 8

Synset Proposition: There are pelvic phlebolith(s).
MNP NP

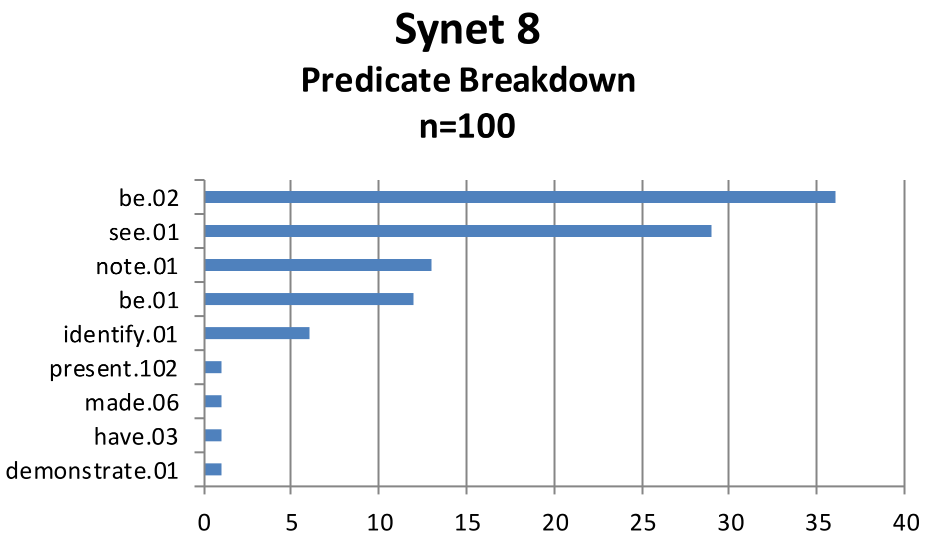


Figure 28— Synset 8 Predicates

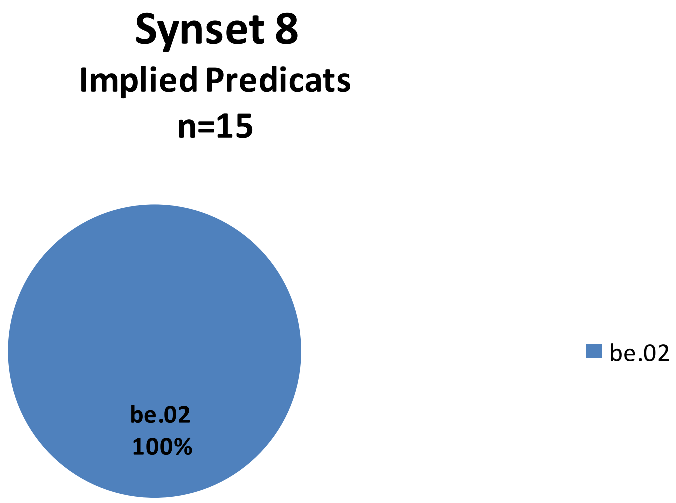


Figure 29— Synset 8 Implied Predicates

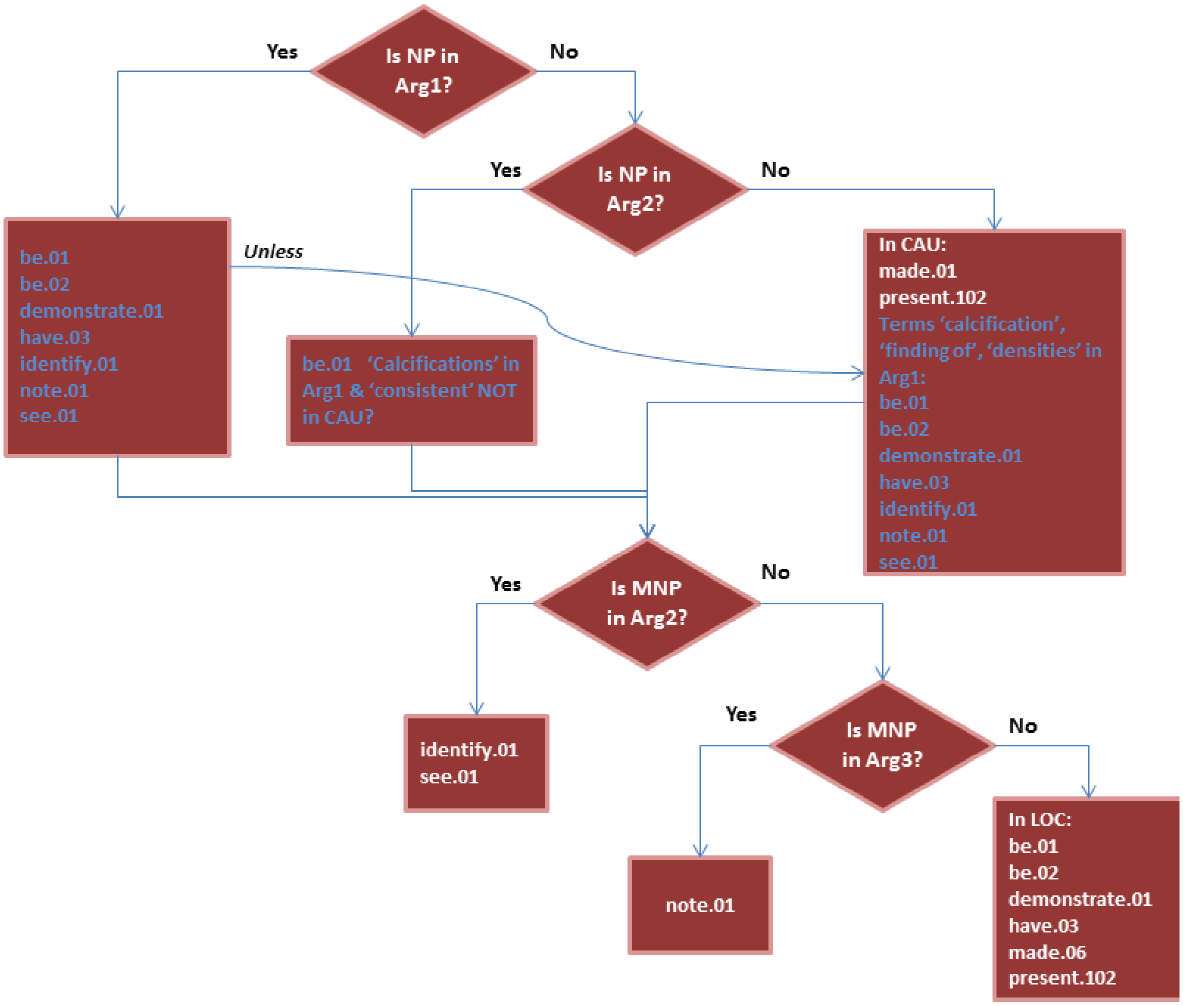


Figure 30— Synset 8 Argument Annotation Flowchart

Discussion:

Although phleboliths are most often found in the pelvis, 'pelvic' is a modifier since it is possible for phlebotliths to exist elsewhere. This synset shows the complication of formulating a term to findings. Pelvic phleboliths is a term for findings; a term to describe an image showing calcifications in a particular part of the body. When findings are part of a sentence but not part of the proposition sentence, the NP could be annotated to several arguments no matter what predicate is used. When the sentence has the image findings (typically in Arg1) in it, phleboliths will be in the argument modifier CAU.

4.3.9 Synset 9

Synset Proposition: There is small vessel ischemic disease of the brain.
NP MNP

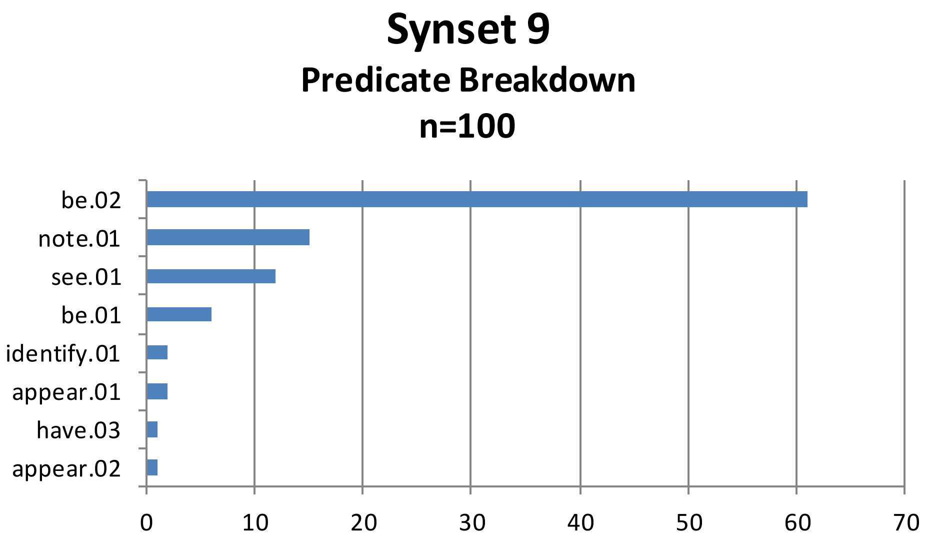


Figure 31— Synset 9 Predicates

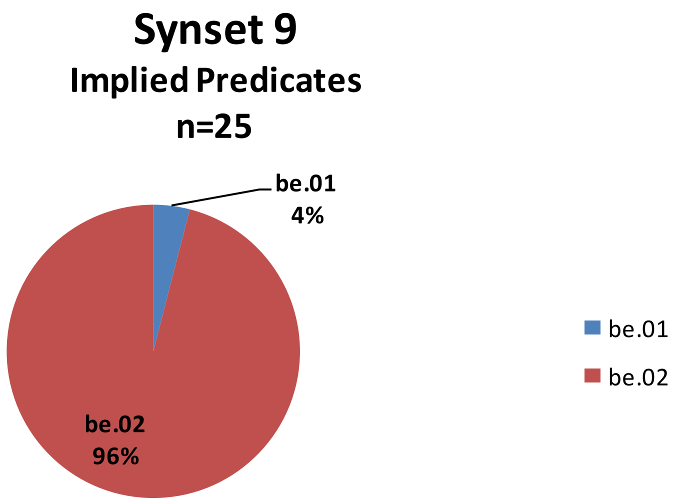


Figure 32— Synset 9 Implied Predicates

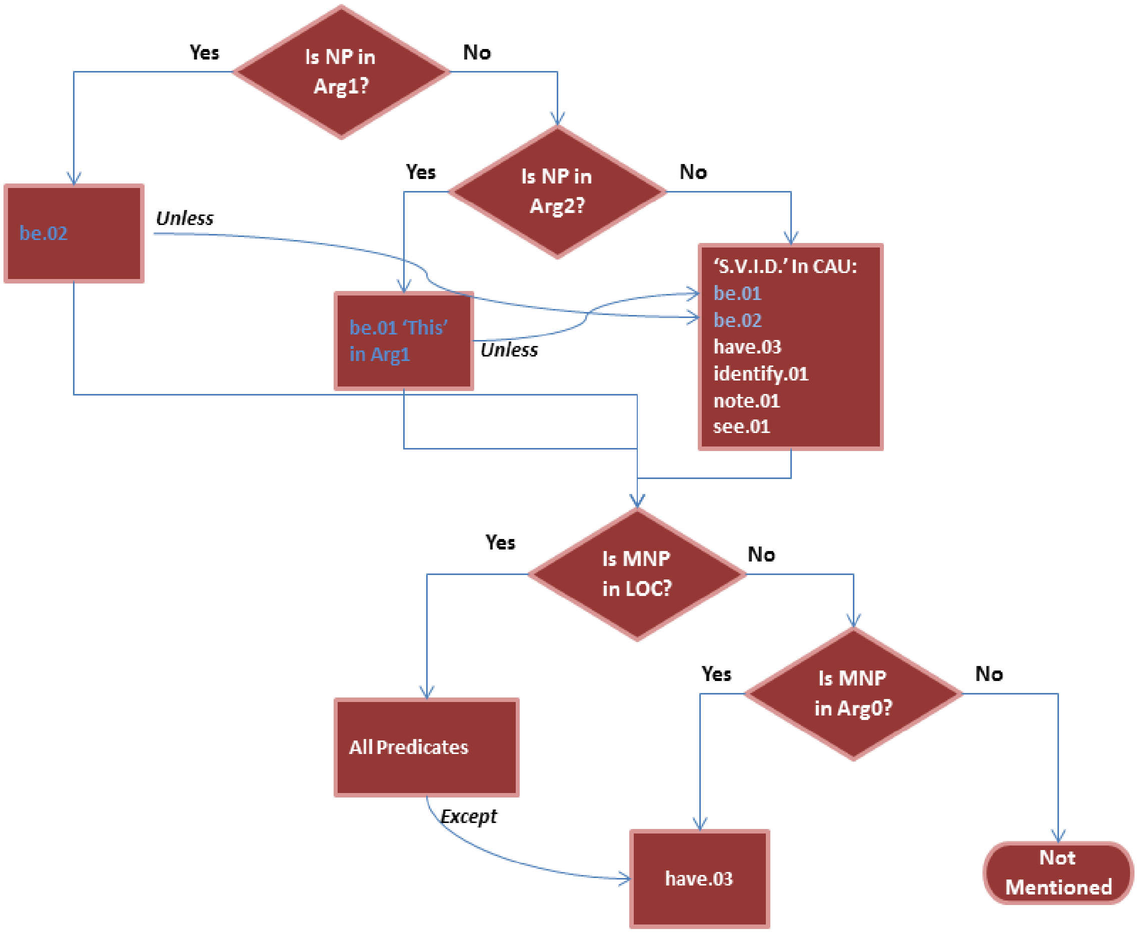


Figure 33— Synset 9 Argument Annotation Flowchart

Discussion:

'Ischemic' is a term acquainted mostly with the heart—'Ischemic heart disease'. Ischemic is also associated with strokes. While the term could technically have numerous modifiers, 'small vessel ischemic disease' is a term that stands on its own. For be.02 and a few be., if 'small vessel ischemic disease' is not expressed in the arg1, it is expressed in the argument modifier CAU because Arg1 has a finding annotated to it such as 'areas of hypodensity', 'decreased attentuation', or 'lucency'.

4.3.10 Synset 10

Synset Proposition: The lungs are diffusely hazy bilaterally.
NP M1VP VP M2VP

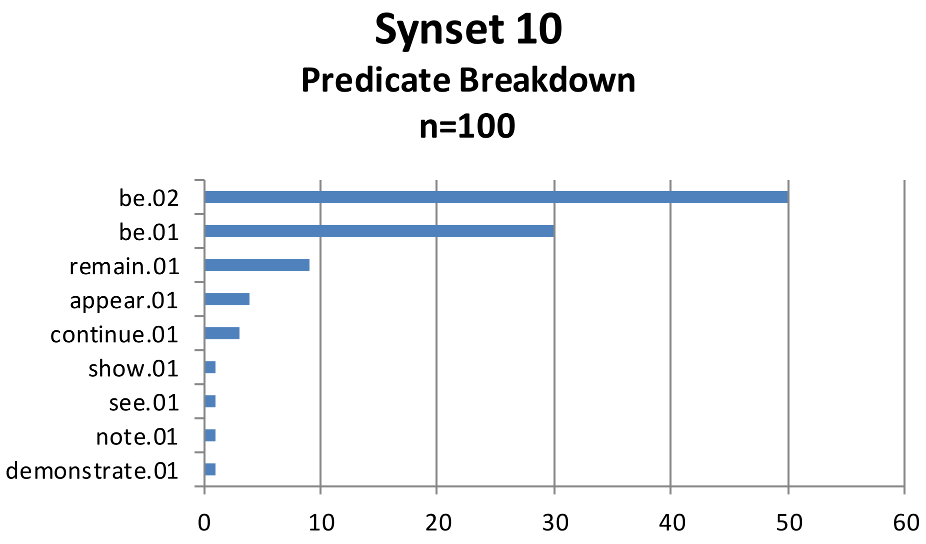


Figure 34— Synset 10 Predicates

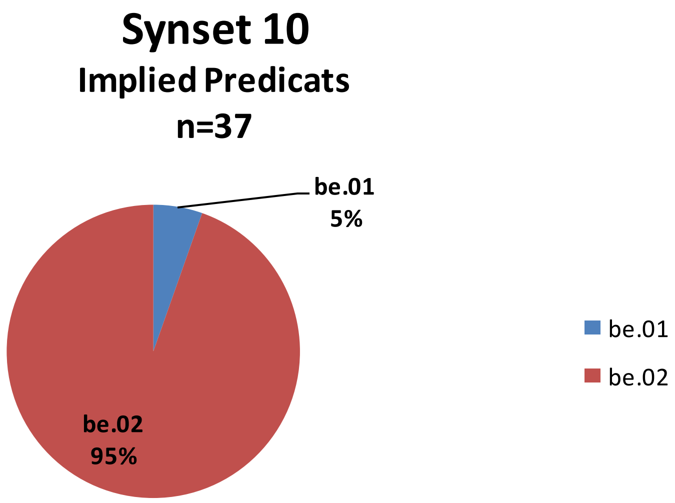


Figure 35— Synset 10 Implied Predicates

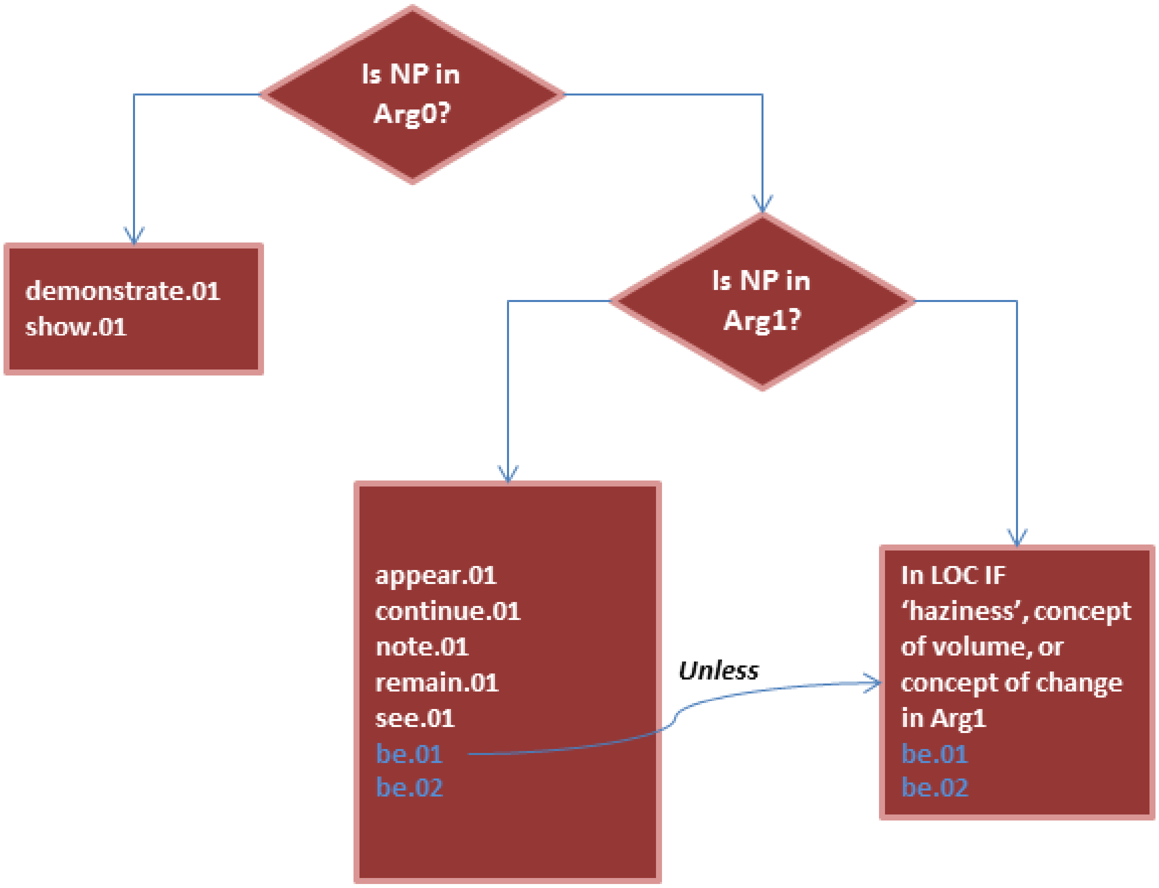


Figure 36— Synset 10 Argument Annotation of NP Flowchart

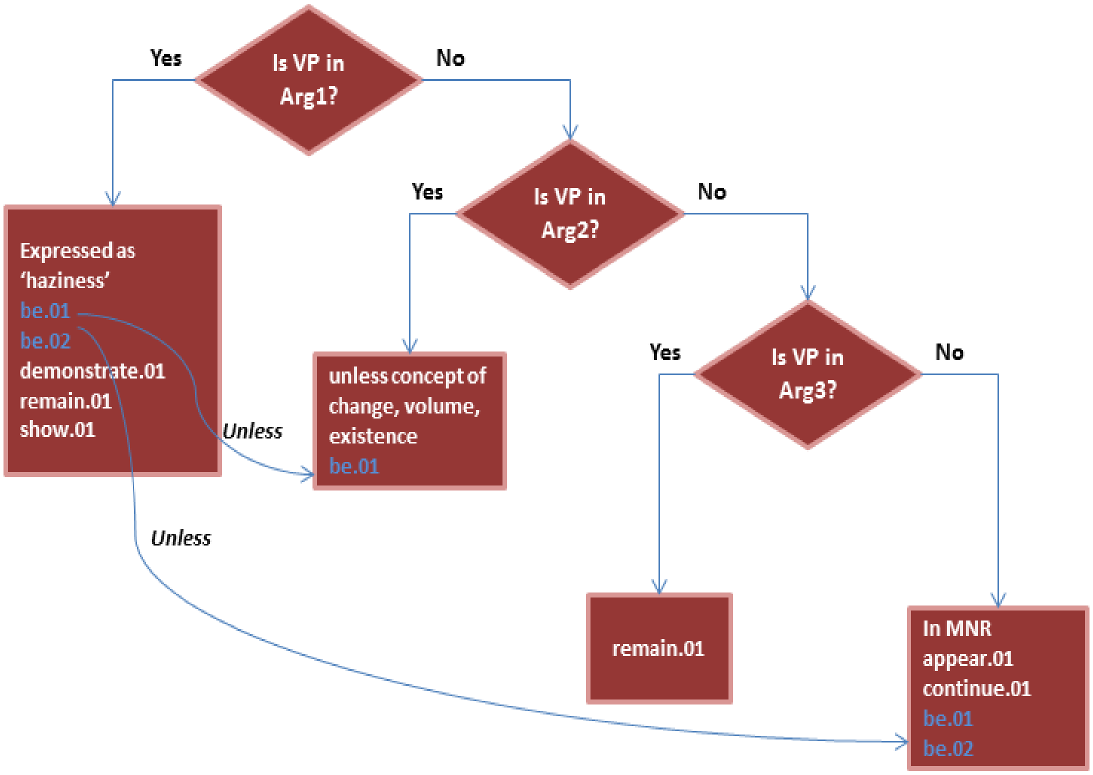


Figure 37— Synset 10 Argument Annotation of VP Flowchart

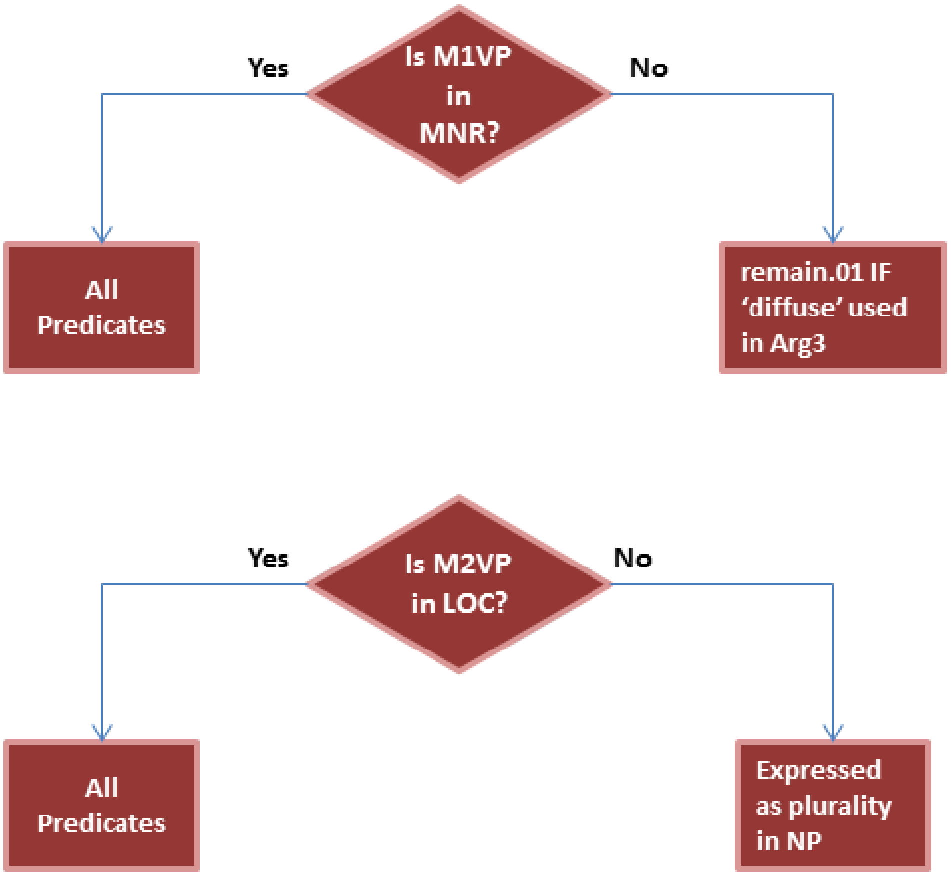


Figure 38— Synset 10 Argument Annotation of VP Modifiers Flowchart

Discussion:

The proposition sentence for this synset is unique in that it has two modifiers of the VP concept ‘hazy’. Typically, one concept is annotated to an argument. Annotation of some sentences in this sysnet showsthat the VP and M1VP are both expressed in the ArgM-MNR. When a phrase has multiple modifiers, future annotation may require subsets of arguments (ArgM-MNR1, ArgM-MNR2) to appropriately parse out the concepts. Also, the concept of ‘bilateral’ is some sentences is not explicitly expressed but implied through plurality of ‘lungs’. This implication will be an interesting test of semantics.

CHAPTER FIVE: DISCUSSION

The primary objective in this project is to use the Propbank guidelines and Unified Verb Index resource to annotate radiology sentences into PAS frames and evaluate if this phrasal annotation schema can be used to semantically compare two sentences. An important by-product of annotation is the uncovering of traits embedded in the syntactical expression of the dataset that may be generalizable to the broader corpus. These crucial traits confirm the annotation schema, but also, illuminate potential adaptation needed for repeatability and better efficiency for NLP processing of the unstructured corpus. This section presents attributes unmasqued through the annotation process.

During the annotation some adaptations were made to schema and guidelines to assist in efficiency. Most sentences in this corpus were simple sentences—subject/verb with modifiers and phrases, but some sentences used conjunctions creating compound sentences. In all compound sentences, one side of the sentence pertained to the topic(s) of the propositional sentence and the other side was addendum information. The side of the sentence with addendum information was placed in ArgM-DIS. Typically, only the conjunctive word in the sentence is annotated to the ArgM-DIS with the compound sentence fully annotated. For the purposes of this study, it was felt that annotating a full compound sentence would produce a level of complexity hindering the goal of evaluating semantic equivalency. Therefore, ArgM-DIS typically contains a full sentence that has not been annotated.

Another modification to the guidelines involves the ArgM-CAU (Cause). The CAU modifier is used to annotate the reason for an action (i.e., 'why', 'as a result of'). Such explicit argument markers presented in Propbank guidelines (Babko-Malaya, 2005) were absent in this dataset. What was expressed frequently in the sentences of the dataset was objective findings and terms to describe these findings. Words frequently used in this kind of sentential structure were 'consistent with' and 'evidence of'. It was felt that for a radiology corpus the CAU modifier was more appropriately suited to represent the term used to describe findings. This modifier came to annotate not just the argument markers but the actually term as well. For example, the proposition sentence for synset 8 is "There are pelvic phlebolith(s)." However, most sentences in the synset were expressed in a structural form of findings and terms as in "Calcific densities in the pelvis consistent with phlebolith are identified." In this example, the entire phrase, 'consistent with phlebolith' would be annotated to ArgM-CAU. For this dataset, the definition for ArgM-CAU was broadened to 'cause & effect'.

The last modification to the annotation guidelines had to address missing predicates. 26% of the sentences in the dataset did not have a predicate. Without a predicate, annotating the structures of the incomplete sentences would result in erroneous assignment to arguments. In this study, each sentence with a missing predicate was transformed to a complete sentence by using first the predicate be.02—adding 'there is/are' at beginning of an incomplete sentence. If the transformed sentence expressed a cohesive thought with be.02, it was annotated with that predicate. If it did not express a cohesive thought, be.01 was used next. If be.01 did not result in a cohesive statement, the annotator selected a predicate that would complete the sentence. The following figure shows that 99% of incomplete sentences could be satisfied by using be.01 or be.02. This process for transforming incomplete sentences into complete sentences could have significant impact for future radiology annotation projects that use PAS because the prediction of which predicate to use is greatly weighted primarily weighted toward be.02, and secondarily, toward be.01.

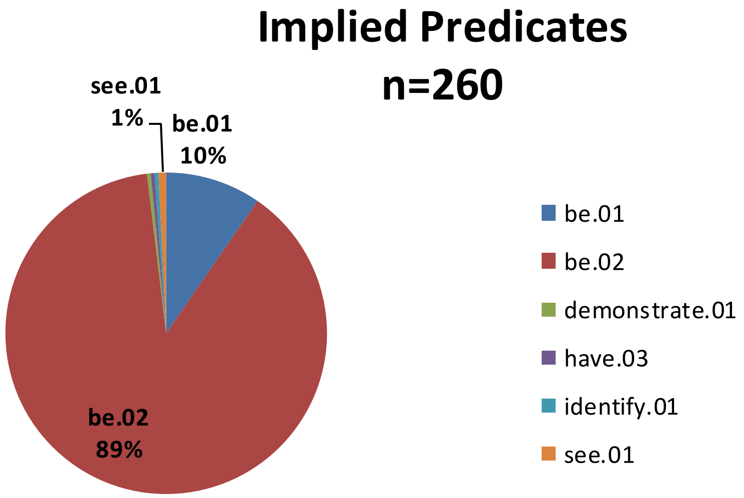


Figure 39--Dataset Summary of All Synsets

Adapting guidelines of annotation to needs of corpus provided a more a consistent structure allowing for inclusive analysis of data set. One of the most obvious variations in argument annotation relates to anatomical location. Most frequently, the anatomical location mentioned in a sentenced annotated to the ArgM-LOC. This is because of the frequency of the existential predicates be.01 and be.02. By default, the description of these predicates' arguments allows annotation of an anatomical location to the ArgM-LOC (exception sometimes being be.01). However, for non existing predicates, there could be varying annotations for the locative modifier depending on the argument definition for that predicate. In the first synset, location ('above the carina') could be annotated for non existenting predicates to Arg2, Arg3, or ArgM-LOC. Multi argument annotation of the same piece of locative information creates complexity for comparing semantic equivalence of anatomical location. Table 11 shows for this dataset all possible arguments (in italics) to which locative/anatomical information could be annotated. 'Attribute of arg1' frequently was a locative piece of information.

Predicate	**Arg2**	**Arg3**	**ArgM-LOC**
be.01	*comment*		*All Other Predicates/be.01*
see.01	*attribute of arg1, further description*		
note.01	hearer	*attribute of arg1*	
identify.01	*attribute*		
remain.01	benefactive, entity who gets the remainder	*attribute of arg1*	
reveal.01	hearer	*attribute of arg1*	
lie.01	*location*		
position.01	*location or position, or attribute of arg1*		
present.01	given to	*secondary attribute of thing given*	
resolve.01	comitative, resolved with	attribute, resolved to or at	
terminate.01	**Use ArgM-LOC-point at which something ends**		
feel.02	*attributive (whom/what the feeling is about)*		
locate.01	**Use ArgM-LOC**		
made.06	*attribute of arg1*		
see.02	*destination*		

Table 11--Propbank Annotation of Locative Argument

In future PAS annotations of radiology corpuses, it may be more beneficial to modify the Unified Verb Index argument descriptions to direct annotation of anatomical locations to the ArgM-LOC. This is exactly what Unified Verb Index does for predicates terminate.01 and locate.01 (bold in Table 11). The re-directive comment could say, “If attribute of arg1 is anatomical location, annotate to ArgM-LOC.” Furthermore, as presented in synset 5, multiple arguments for ArgM-LOC may be necessary for a radiology corpus because a primary concept could be modified by more than one locative descriptor. Multiple locative arguments will only work if it is clear which argument of the PAS frame it is modifying.

Since PAS centers on the predicate, looking at the predicates and predicate usage in the corpus may perhaps shed light on what general comparisons will be made between sentences. With over 2/3 of the predicates expressed in the dataset using an existential form of ‘to be’, it is evident that Arg0 is absent. In fact, with non-‘to be’ predicates, less than 1% annotated text to Arg0. Arg0 represents the role of an agent or someone doing the action. In contrast, all sentences annotated with Arg1. According to Babko-Malaya (2005), this indicates that predicates in this dataset are ‘externally caused’. With externally caused predicates, the explanation for the motive or stimulant in the sentence is implied. Therefore, the dataset for this study implies that actions and thoughts by medical staff and or patients are absent in a radiology corpus and that such thoughts and actions are already implied within the context of the corpus. The predicates in this dataset suggest that the sentences in a radiology corpus are expressions of statements and not expressions of action. Because the dataset is presented in ten sets of semantically equivalent sentences without access to the entire radiology report, these statements on one level appear to be made mysteriously and without authority. Although, the context of the corpus tells us that a professionally trained radiologist observed patterns in a medical image and expressed his or her conclusions about his or her observations. The predicates in this corpus tell us that these observations and conclusions are externally caused. For predicates not be.01 or be.02, 99% of their forms were expressed in the passive vs. active form. Any internally-caused predicates (predicates where text is annotated to Arg0) were predominantly related to the expression that the product of a radiological procedure (‘films’ ‘scan’ ‘images’ or ‘views’) projected a finding and was expressed through predicates demonstrate.01 and show.01.

The implication that predicates in a corpus are externally-caused vs. internally-caused suggests that the writer has a great deal of leniency to express a thought. A radiology report merely has to present observations and interpretation of those observations. It does not have to explain the actions of the disease process requiring use of internally-cause predicates. While predicates be.01 or be.02 and passive forms of predicates may not rank high on a level for creativity, they are simple predicates that allow for a great variation in language usage. This variation and flexibility in language use may be the difficulty for scalability in PAS (Friedlin et al, 2011). In the fourth synset in which the proposition is “The pulmonary vessels are prominent,’ annotation of arguments greatly varies even among identical predicates with different word forms such as ‘prominent’ vs. ‘prominence’.

Through the flowcharts of each synset analyzing annotation assignment of NP and VP arguments of proposition sentence and their modifiers to candidate sentences, it is clear that there is variability in the annotation of like content when based on predicates and way language is used in a sentence. This raises the questions of what importance semantic role labeling has in comparing semantic equivalence of two radiological sentences. If phrases that need to be semantically compared are dispersed among arguments with different semantic roles, how is it appropriate to proceed with PAS as method for comparison? Does the scalability problem prevent further experimentation?

In this project, 33 predicates were used in a dataset of 999 sentences. Arg1 is the argument that received the most annotated information. Traditionally, the semantic role of Arg1 is the patient (that which is acted upon). However, an argument has already been presented that the predicates in this dataset are externally-caused which intends to say that action upon the patient is implied. For a corpus void of real action, such as a radiology corpus, the role of Arg1 may need to be re-examined. In Table 12, all 33 predicates are listed with the description of Arg1 for that predicate. The root term describing Arg1 of most predicates is ‘thing’. For be.01, an inclusive descriptor, ‘topic,’ is used. It could be proposed that ‘thing’ & ‘topic’ is a very generalizable concept broad enough to allow that actual entity or term to cross semantic roles. A thing, such as ‘lungs’, could be annotated to Arg1 or to ArgM-LOC. A thing of ‘haziness’ annotated to Arg1 could be annotated as ArgM-MNR in another sentence in the form of ‘hazy’. It seems that externally-caused predicates are flexible enough to incorporate varying parts of speech and word forms that are not dependent on semantic roles. Cross semantic role comparison with PAS is feasible and scalability is no longer a problem as long as the synset shows a consistent pattern of concept expression intertwined among the predicates’ arguments.

Predicate	Arg1	Predicate	Arg1
be.02	**thing** that is	perform.01	performance, thing performed
be.01	topic	resolve.01	**thing** being resolved
see.01	**thing** viewed	terminate.01	**Thing** ending
note.01	utterance	appear.02	viewer, to whom does it seem like that?
identify.01	item being labelled	appreciate.02	**thing** appreciated
remain.01	Thing remaining	change.01	**thing** changing
appear.01	**thing** appearing	clear.04	**thing** becoming clean
demonstrate.01	**thing** demonstrated	complete.01	task, action coming to an end
show.01	**thing** seen/shown	feel.02	Belief
obtain.01	**thing** gotten	improve.01	**thing** improving
continue.01	**thing** continuing	locate.01	institution, thing located
have.03	possession	made.06	**thing** seeming
reveal.01	utterance, truth condition	persist.01	**thing** continuing, persisting
lie.01	entity in the position	present.102	Disorder presenting
position.01	**thing** positioned, often REC	see.02	entity in motion
present.01	**thing** given	suggest.01	Utterance (suggestion)
increase.01	**thing** increasing		

Table 12--Propbank Description of Arg1 Dataset Predicates

Since an argument has been made to support cross semantic role comparing, it is possible to address the overall question of the project: Is PAS frame an appropriate phrasal annotation pre-processing method from which to build the logical expression to semantically compare sentences? For each synset, a flowchart shows a process by which semantically equivalent sentences when annotated to a PAS frame direct the location of like content based on the proposition sentence (NP, MNP, VP, MVP). Across all synsets, it is estimated that the algorithm expressed in the flowchart will have a content coverage of no less than 96%. The flowcharts trace content based on predicates. If new sentences were added to the synsets, it is possible that the algorithm represented in the flowcharts could change, but seeing how out of 999 sentences, only 33 different predicates are used, it would be safe to assume that any new predicates would have minor changes to the underlying algorithm. The hypothesis is that the trend line represented in Figure 3 would result in no significant change.

Because of the generalizability of a predicate's argument descriptions for externally-caused predicates, it is possible to compare the content of an argument semantic role label to content of an argument of a different semantic role label. While the PAS method of phrasal annotation may be intricate and clunky, text of every sentence in the dataset was annotated to a role label indicating that the Probank knowledge source (Unified Verb Index) is sufficient to guide annotation of radiology sentences. Overall, the first research question presented in section 2.2 appears sufficiently satisfied.

While this annotation project has substantiated use of PAS to semantically compare two radiology sentences, some topics came to light during the annotation process that should be addressed before moving the study into the next phase of the second research question raised in section 2.2. As mentioned, the algorithm expressed in the synset flowcharts is rooted in a proposition sentence. The proposition sentences are basic and concrete (see table 9). While all candidate sentences in each synset maintain semantic equivalency to the proposition sentence, many candidate sentences elaborate or express concepts in addition to the concept in the proposition sentence. For example, in the sentences in Table 13, all express prominence of the pulmonary vessels, but each candidate sentence has content not expressed (underlined) in the proposition sentence. It is not clear when a sentence is annotated into phrases how additional content of the candidate sentences not present in the proposition may or may not affect the semantic directionality of the candidate sentences.

Proposition	The pulmonary vessels are prominent.
Candidates	There is cardiomegaly with prominent pulmonary vascularity , which appears stable from prior studies.
	Pulmonary vessels are prominent with mild edema.
	Cardiac silhouette and pulmonary vascularity are prominent.

Table 13--Candidate Sentences With Additional Concept

Another issue uncovered through the annotation process stems from the lack of semantic challenge posed by the dataset. The intent of the dataset is to present semantic challenges in order to determine if an annotation method of creating a logical expression based on an entire single sentence is better than creating a logical expression based on phrasal annotation. At the completion of annotation for this project, it was discovered that most synsets offered minimal variation of the NP and/or VP of the proposition sentence. For the synset with the proposition “There is small vessel ischemic disease of the brain,” 81 sentences contain the exact matching NP, “small vessel ischemic disease” in the ArgM-CAU role. The remaining 19 sentences have the exact NP in other arguments. This suggests that the dataset may be more appropriate for information retrieval (keyword) tests than information extraction. It appears that in many instances the difference between sentences was with spelling and punctuation rather than variations in expressions of concepts. Table 14 is the author’s personal evaluation as to the level of semantic challenge the synsets may present. While the study should move forward to trial the measurement of semantic directionality, future testing should create a dataset from a corpus where experts select sentences for semantic challenges rather than a random selection.

Synet	Propositional Sentence of Synset	Semantic Challenge Level
1.	The endotracheal tube is above the carina.	Low
2.	There is no pneumothorax.	Low
3.	There is a left lower lobe pulmonary infiltrate(s).	Low
4.	The pulmonary vessels are prominent.	Medium
5.	A posterior anterior (PA) chest x-ray was performed.	High
6.	The gray white matter differentiation of the brain is normal.	Low
7.	The intervertebral disc heights are normal.	High
8.	There are pelvic phlebolith(s).	Low
9.	There is small vessel ischemic disease of the brain.	Low
10.	The lungs are diffusely hazy bilaterally.	Low

Table 14--Semantic Challenge of Synsets

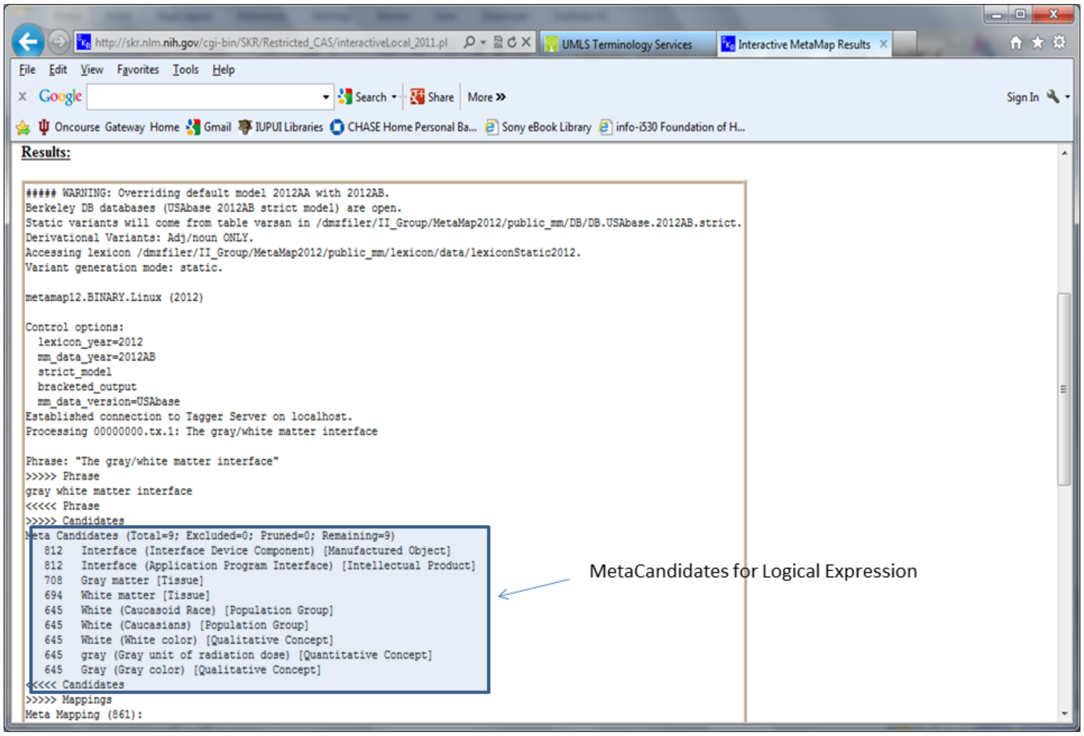
CHAPTER SIX: FUTURE PHASES

The next phase of this project will address the second research question: Does the logical expression of the propositional sentence represent the synset through a semantic similarity measurement that considers directionality? To address this question, the following processes will have to occur:

1) Create the logical expression of each PAS frame
2) Semantically compare the PAS frame of candidate sentences to proposition sentence using a scoring mechanism that measures directionality between two experts
3) Statistically analyze scores to determine if agreement is due to chance

6.1 Creating Logical Expression

In this step, each argument of each PAS frame will be entered into the UMLS (2012) MetaMap knowledge source to create a logical expression. The output of MetaMap provides MetaCandidates, which are the term labels of entries in the UMLS ontology, as well as, the semantic classification for that entry. An example of the output is shown in figure 41.



```
##### WARNING: Overriding default model 2012AA with 2012AB.
Berkeley DB databases (USAbase 2012AB strict model) are open.
Static variants will come from table varsan in /dmzfiler/II_Group/MetaMap2012/public_mm/DB/DB.USAbase.2012AB.strict.
Derivational Variants: Adj/noun ONLY.
Accessing lexicon /dmzfiler/II_Group/MetaMap2012/public_mm/lexicon/data/lexiconStatic2012.
Variant generation mode: static.

metamap12.BINARY.Linux (2012)

Control options:
  lexicon_year=2012
  mm_data_year=2012AB
  strict_model
  bracketed_output
  mm_data_version=USAbase
Established connection to Tagger Server on localhost.
Processing 00000000.tx.1: The gray/white matter interface

Phrase: "The gray/white matter interface"
>>>>> Phrase
gray white matter interface
<<<<< Phrase
>>>>> Candidates
Meta Candidates (Total=9; Excluded=0; Pruned=0; Remaining=9)
   812   Interface (Interface Device Component) [Manufactured Object]
   812   Interface (Application Program Interface) [Intellectual Product]
   708   Gray matter [Tissue]
   694   White matter [Tissue]
   645   White (Caucasoid Race) [Population Group]
   645   White (Caucasians) [Population Group]
   645   White (White color) [Qualitative Concept]
   645   gray (Gray unit of radiation dose) [Quantitative Concept]
   645   Gray (Gray color) [Qualitative Concept]
<<<<< Candidates
>>>>> Mappings
Meta Mapping (861):
```

MetaCandidates for Logical Expression

Figure 40—Screenshot of MetaMap Output—Meta Candidates

A more readable output of the MetaCandidates is:

```
Phrase: "The gray/white matter interface"
>>>>> Phrase
gray white matter interface
<<<<< Phrase
>>>>> Candidates
Meta Candidates (Total=9; Excluded=0; Pruned=0;
Remaining=9)
   812    Interface (Interface Device Component)
[Manufactured Object]
   812    Interface (Application Program Interface)
[Intellectual Product]
   708    Gray matter [Tissue]
   694    White matter [Tissue]
   645    White (Caucasoid Race) [Population Group]
   645    White (Caucasians) [Population Group]
   645    White (White color) [Qualitative Concept]
   645    gray (Gray unit of radiation dose) [Quantitative
Concept]
   645    Gray (Gray color) [Qualitative Concept]
```

```
<<<<< Candidates
>>>>> Mappings
```

In some cases, MetaMap will not be able to map all terms in argument. Notice in the MetaCandidates provided above that the semantic type (underlined) for interface is not appropriate for the brain. With such cases, a note will be made and conferences will address how to best create a logical expression for the unmapped terms. A simple logical expression for a PAS frame is show in the following table 15:

<table>
<tr><th colspan="4">The gray/white matter interface is preserved</th></tr>
<tr><td>Arguments</td><td>Desc</td><td>Text</td><td>Logical Expression
Term [Semantic Type]</td></tr>
<tr><td>Arg1</td><td>topic</td><td>The gray/white matter interface</td><td>Gray matter [tissue]
White matter [tissue]
Interface [no appropriate entry]</td></tr>
<tr><td>Arg2</td><td>comment</td><td>preserved</td><td>Preserving [functional concept]</td></tr>
</table>

Table 15--Transforming PAS Frame Into Logical Expression

6.2 Scoring PAS Frames

In the second step, the logical expression (LE) of PAS frames of candidate sentences will be scored against the PAS frame of proposition sentence of the synset. Coders will be asked to use a scoring system modified from Humphreys, McCray, & Cheh (1997). This scoring system is presented in table 16.

<table>
<tr><th colspan="2"><i>Score</i> = (<i>Arg (0 to x</i>)) of candidate PAS LE is (<i>category</i>) to corresponding propositional PAS LE Arg?</th></tr>
<tr><td>Score</td><td>Category</td></tr>
<tr><td>0</td><td>Does Not Match</td></tr>
<tr><td>1</td><td>Has Broader Meaning</td></tr>
<tr><td>2</td><td>Has Narrower Meaning</td></tr>
<tr><td>3</td><td>Is Complete Match</td></tr>
</table>

Table 16--Semantic Scoring Codes

In situations when one PAS frame has text annotated to an argument not present in the other PAS frame, annotators will mark a score of 9. Table 17 shows what the scoring will look like.

	Proposition	**Candidate**	**Rater 1 Score**	**Rater 2 Score**	**Success/ Failure**
Arg1	Gray matter [tissue] White matter [tissue] Differentiation(Histopathologic Grade differentiation) [Clinical attritube]	Gray matter [tissue] White matter [tissue] Interface [**no appropriate entry**]	3	3	1
Arg2	Normal [Qualitative Concept]	Preserving [Functional concept]	2	2	1
ArgM-LOC	Brain [Body Part, Organ, or Organ Component]		9	9	9
Success of Coder Agreement:					1
Success of PAS (True Synonymy):					1

Table 17--Scoring Process of Candidate PAS Frame to Propositonal PAS Frame

In table 17, scorers have said for Arg1 that the logical expression for the candidate sentence is a complete match (semantically the same) as the proposition sentence. For Arg2, the scorers have said that the logical expression of the candidate sentence has a narrower meaning than the proposition. The candidate sentence has no logical expression for ArgM-LOC and therefore, a score of 9. As alluded to in the discussion section, it is not clear how the score of 9 changes the semantic similarity of the candidate sentence. With sentential logical expressions, the candidate sentence inherits the locative property of ‘Brain’. However, when two sentences are annotated into phrases, it is not apparent if that trait is inherited. The last column of the table (shaded in gray) indicates the success

(1) or failure (0) of the scorers agreeing. Finally, a score is calculated to determine over all success of scorer agreement based on Boolean logic.

CHAPTER SEVEN: CONCLUSION

In this study, the science of annotation has been used to determine if the predicate-argument structure is a suitable annotation method to retain the completeness of sentences for generating logical expressions in preparation to do a measurement of semantic similarity between two sentences. Ten unique and semantically different synsets formed the dataset for testing an annotation schema that resulted in recommending modifications to the Propbank schema. One modification was to change the description of two argument modifiers. The ArgM-CAU description was modified to incorporate a statement of ‘findings’. It is recommended that any text/symbol representing an anatomical location be annotated to the ArgM-LOC even if it according to Propbank schema should be annotated to a primary Arg[0-4]. The most profound discovery that possibly could be inferred about a radiology corpus is that predicates are externally-caused minimizing the importance the semantic role labels and emphasizing more of the language usage that allows for cross comparison of different semantic role labels. The content analysis of the annotation output in this study showed a model of predictability for locating ‘like’ content in candidates sentences to the proposition’s NP & VP. This model may be useful to future NLP methods that need to pre-process sentences in a corpus to measure semantic similarity.

It should be noted that for a non-radiology corpus, one cannot assume predicates are externally-caused (Albright et al, 2013). Furthermore, the frequency distribution in this study (see Figure 3) may not result from annotation of other corpus domains. However, the methods of content analysis used in this study may still have relevancy if a synset exists by which the developed model can make predictions. Developing the synset may pose the biggest obstacle to the larger adoption of the findings in this study. This study had the benefit of using synsets from a larger annotation study (Friedlin et al, 2011). It is not practical to assume that such large scale annotations will be completed for all specialized domains, let alone the broader domain, in healthcare. A hierarchy of sentential propositions would require far too much time (years of annotation) to produce the needed synsets for a predictive model to provide basis for measurement of sentential semantic similarity. To incorporate the predictive model from PAS to the broader NLP community, what may be needed is collaboration with other NLP techniques as suggested by Chapman et al. (2011). More specifically, a possible merging with an NLP technique that could automate the generation of preliminary synsets would position the findings in this study to serve as a possible quality control. The quality control would serve a way to prepare to test the accuracy of the preliminary generated synset that could be done first by experts followed by automation which would enable ability to increase the size and accuracy of the synset.

REFERENCES

Albright, D., Lanfranchi, A., Fredriksen, A., Styler, W. F., Warner, C., Hwang, J. D., ... & Savova, G. K. (2013). Towards comprehensive syntactic and semantic annotations of the clinical narrative. Journal of the American Medical Informatics Association, doi:10.1136/amiajnl-2012-001317

Athenikos, S. J., & Han, H. (2010). Biomedical question answering: A survey. Computer Methods and Programs in Biomedicine, 99(1), 1-24.

Babko-Malaya, O. (2005). Propbank Annotation Guidelines Retrieved November 7, 2010, from http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf

Builtelaar, P., & Sacaleanu, B. (2001). Ranking and Selecting Synsets by Domain Relevance. Paper presented at the Proceedings of WordNet and Other Lexical Resources.

Campbell, J. R., Carpenter, P., Sneiderman, C., Cohn, S., Chute, C. G., & Warren, J. (1997). Phase II evaluation of clinical coding schemes. Journal of the American Medical Informatics Association, 4(3), 238-251. doi: 10.1136/jamia.1997.0040238

Campbell, K. E., Oliver, D. E., Spackman, K. A., & Shortliffe, E. H. (1998). Represnting thoughts, words, and things in the UMLS. Journal of American Medical Informatics Association, 5(5), 421-431.

Caviedes, J. E., & Cimino, J. J. (2004). Towards the development of a conceptual distance metric for the UMLS. Journal of Biomedical Informatics, 37(2), 77-85.

Chapman, W. W., Nadkarni, P. M., Hirschman, L., D'Avolio, L. W., Savova, G. K., & Uzuner, O. (2011). Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. Journal of the American Medical Informatics Association, 18(5), 540-543. doi: 10.1136/amiajnl-2011-000465

Chaves-González, J. M., & Martínez-Gil, J. (2013). Evolutionary algorithm based on different semantic similarity functions for synonym recognition in the biomedical domain. Knowledge-Based Systems, 37(0), 62-69. doi: http://dx.doi.org/10.1016/j.knosys.2012.07.005

Chou, W.-C., Tsai, R. T.-H., Su, Y.-S., Ku, W., Sung, T.-Y., & Hsu, W.-L. (2006). A Semi-Automatic Method for Annotating a Biomedical Proposition Bank. Paper presented at the Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora, Sydney, Australia.

Chute, C. G., Cohn, S. P., Campbell, K. E., Oliver, D. E., & Campbell, J. R. (1996). The content coverage of clinical classifications. Journal of the American Medical Informatics Association, 3(3), 224-233. doi: 10.1136/jamia.1996.96310636

Cohen, K. B., & Hunter, L. (2006). A critical review of PASBio's argument structures for biomedical verbs. BMC Bioinformatics, 7(Suppl 3), S5.

Cohen, K. B., Ogren, P. V., Fox, L., & Hunter, L. (2005). Corpus Design for Biomedical Natural Language Processing. Paper presented at the Proceedings of the ACL-ISMB Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics, Detroit, Michigan.

Delisle, S., & Szpakowicz, S. (1997). Extraction of predicate-argument structures from texts. Paper presented at the Proceedings of the 2nd Conference on Recent Advances in Natural Language Processing (RANLP-97), Tzigov Chark (Bulgaria).

Demner-Fushman, D., Chapman, W. W., & McDonald, C. J. (2009). What can natural language processing do for clinical decision support? Journal of Biomedical Informatics, 42, 760-772.

Elhadad, N., & Sutaria, K. (2007). Mining a Lexicon of Technical Terms and Lay Equivalents. Paper presented at the Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing, Prague, Czech Republic.

Fan, J.-W., & Friedman, C. (2007). Semantic classification of biomedical concepts using distributional similarity. Journal of the American Medical Informatics Association, 14(4), 467-477.

Friedlin, J., Mahoui, M., Jones, J., & Jamieson, P. (2011, 26-29 July 2011). Knowledge Discovery and Data Mining of Free Text Radiology Reports. Paper presented at the Healthcare Informatics, Imaging and Systems Biology (HISB), 2011 First IEEE International Conference on.

Friedman, C., & Hripcsak, G. (1999). Natural language processing and its future in medicine. Academic Medicine, 74(8), 890-895.

Gallier, J. H. (1986). Logic for computer science: Foundations of automatic theorem proving. New York: Harper & Row.

Godbert, E., & Royaute, J. (2010). PredXtract, A Generic Platform to Extract in Texts Predicate Argument Structures (PAS). Paper presented at the LREC 2010 Proceedings, Valleta, Malta.

Goulart, R., Strube de Lima, V., & Xavier, C. (2011). A systematic review of named entity recognition in biomedical texts. Journal of the Brazilian Computer Society, 17(2), 103-116. doi: 10.1007/s13173-011-0031-9

Gwet, K. (2008). Variance Estimation of Nominal-Scale Inter-Rater Reliability with Random Selection of Raters. Psychometrika, 73(3), 407-430. doi: 10.1007/s11336-007-9054-8

Holden, R. J. (2010). Physicians' beliefs about using EMR and CPOE: In pursuit of a contextualized understanding of health IT use and behavior. International Journal of Medical Informatics, 79(2), 71-80.

Hope, C. J., Garvin, J. H., & Sauer, B. C. (2012). Information extraction from narrative data. Am J Health Syst Pharm, 69(6), 455, 460-451. doi: 10.2146/ajhp110135

Humphreys, B. L., McCray, A. T., & Cheh, M. L. (1997). Evaluating the coverage of controlled health data terminologies: Report on the results of the NLM/AHCPR large scale vocabulary test. Journal of the American Medical Informatics Association, 4(6), 484-500. doi: 10.1136/jamia.1997.0040484

Johnson-Laird, P. (2010). Deductive reasoning. Wiley Interdisciplinary Reviews: Cognitive Science, 1(1), 8-17. doi: 10.1002/wcs.20

Jones, J., Newsom, E., & Delaney, C. (2009). Use of Nursing Management Minimum Data Set (NMMDS) for a Focused Information Retrieval. Paper presented at the 13th International Conference on Human Computer-Interaction, San Diego, CA.

Juckett, D. (2012). A method for determining the number of documents needed for a gold standard corpus. Journal of Biomedical Informatics, 45(3), 460-470. doi: http://dx.doi.org/10.1016/j.jbi.2011.12.010

Kilicoglu, H., Fiszman, M., Rosemblat, G., Marimpieti, S., & Rindflesch, T. (2010). Arguments of Nominals in Semantic Intepretation of Biomedical Text. Paper presented at the BioNLP 2010, Uppsala, Sweden.

Kim, J.-D., Ohta, T., & Tsujii, J. i. (2008). Corpus annotation for mining biomedical events from literature. BMC Bioinformatics, 9(1), 10.

Kogan, Y., Collier, N., Pakhomov, S., & Krauthammer, M. (2005). Towards Semantic Role Labeling & IE in the Medical Literature. Paper presented at the Annual AMIA Symposium.

Lin, D. (1998). An Information-Theoretic Definition of Similarity. Paper presented at the Proceedings of the Fifteenth International Conference on Machine Learning (ICML-98), Madison, Wisconsin.

Liu, K., Hogan, W. R., & Crowley, R. S. (2010). Natural language processing methods and systems for biomedical ontology learning. [doi: DOI:

10.1016/j.jbi.2010.07.006]. Journal of Biomedical Informatics, In Press, Corrected Proof.

Madnani, N., & Dorr, B. J. (2010). Generating phrasal and sentential paraphrases: A survey of data-driven methods. Computational Linguistics, 36(3), 341-387. doi: doi:10.1162/coli_a_00002

Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: the penn treebank. Comput. Linguist., 19(2), 313-330.

McCrae, J., & Collier, N. (2008). Synonym set extraction from the biomedical literature by lexical pattern discovery. BMC Bioinformatics, 9(159).

Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-Based and Knowledge-Based Measures of Text Semantic Similarity. Paper presented at the Proceedings of the 21st national conference on Artificial intelligence, Boston, Massachusetts.

Millar, N., & Budgell, B. (2008). The language of public health—a corpus-based analysis. Journal of Public Health, 16(5), 369-374. doi: 10.1007/s10389-008-0178-9

Miller, S., Guiness, J., & Zamanian, A. (2004). Name Tagging with Word Clusters and Discriminiative Training. Paper presented at the Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics.

Miltsakaki, E., Robaldo, L., Lee, A., & Joshi, A. (2010). Sense annotation in the penn discourse treebank. In A. Gelbukh (Ed.), Computational Linguistics and Intelligent Text Processing (Vol. 4919, pp. 275-286). Berlin: Springer.

Miyao, Y., Ohta, T., Masuda, K., Tsuruoka, Y., Yoshida, K., Ninomiya, T., & Tsujii, J. i. (2006). Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases. Paper presented at the Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, Sydney, Australia.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551. doi: 10.1136/amiajnl-2011-000464

Nieswiadomy, R. M. (2008). Foundations of Nursing Research (5th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Ogden, C. K., & Richards, I. A. (1943). The meaning of meaning (8 ed.). New York: Harcourt, Brace & World, Inc.

Pedersen, T., Pakhomov, S. V. S., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. Journal of Biomedical Informatics, 40(3), 288-299.

Pugeault, F., Saint-Dizier, P., & Monteil, M.-G. (1994). Knowledge extraction from texts: A method for extracting predicate-argument structures from texts. Paper presented at the Proceedings of Fifteenth International Conference on Computational Linguistics, Kyoto, Japan.

Pyysalo, S., Ginter, F., Heimonen, J., Bjorne, J., Boberg, J., Jarvinen, J., & Salakoski, T. (2007). BioInfer: A corpus for information extraction in the biomedical domain. BMC Bioinformatics, 8(1), 50.

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics, 19(1), 17-30.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. eprint arXiv:cmp-lg/9511007.

Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research, 11, 95-130.

Roberts, A., Gaizauskas, R., Hepple, M., Demetriou, G., Guo, Y., Roberts, I., & Setzer, A. (2009). Building a semantically annotated corpus of clinical texts. Journal of Biomedical Informatics, 42(5), 950-966.

Roberts, A., Gaizauskas, R., Hepple, M., Demetriou, G., Guo, Y., Setzer, A., & Roberts, I. (2008). Semantic annotation of clinical text: The CLEF corpus. Paper presented at the Proceedings of Building and evaluating resources for biomedical text mining: workshop at LREC.

Rost, T. B., Huseth, O., Nytro, O., & Grimsmo, A. (2008). Lessons from developing an annotated corpus of patient histories. Journal of Computing Science and Engineering, 2(2), 162-179.

Roth, D., & Yih, W.-t. (2001). Relational learning via propositional algorithms: An information extraction

case study. Paper presented at the Proceedings of the 15th International Conference On Artificial Intelligence (IJCAI).

Samsonovic, A. V., & Ascoli, G. A. (2010). Principal semantic components of language and the measurement of meaning. PLoS ONE, 5(6), e10921.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. Journal of the American Medical Informatics Association, 17(5), 507-513. doi: 10.1136/jamia.2009.001560

Sevenster, M., Ommering, R., & Qian, Y. (2012). Automatically Correlating Clinical Findings and Body Locations in Radiology Reports Using MedLEE. Journal of Digital Imaging, 25(2), 240-249. doi: 10.1007/s10278-011-9411-0

Shi, L., & Mihalcea, R. (2005). Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing. In A. Gelbukh (Ed.), Computational Linguistics and Intelligent Text Processing (Vol. 3406, pp. 100-111). Berlin: Springer.

Sinacore, J. M., Connell, K. J., Olthoff, A. J., Friedman, M. H., & Gecht, M. R. (1999). A Method for Measuring Interrater Agreement on Checklists. Evaluation & the Health Professions, 22(2), 221-234. doi: 10.1177/01632789922034284

Surdeanu, M., Harabagiu, S., Williams, J., & Aarseth, P. (2003, July 2003). Using Predicate-Argument Structures for Information Extraction. Paper presented at the Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan.

Tan, H., Kaliyaperumal, R., & Benis, N. (2012). Ontology-Driven Construction of Domain Corpus with Frame Semantics Annotations. Computational Linguistics and Intelligent Text Processing, 7181, 54-65. doi: 10.1007/978-3-642-28604-9_5

Tsai, R., Chou, W.-C., Su, Y.-S., Lin, Y.-C., Sung, C.-L., Dai, H.-J., . . . Hsu, W.-L. (2007). BIOSMILE: A semantic role labeling system for biomedical verbs using a maximum-entropy model with automatically generated template features. BMC Bioinformatics, 8(1), 325.

Unified Verb Index. (2012). Retrieved December 12, 2012, from http://verbs.colorado.edu/verb-index/index.php

United Medical Language Server--Semantic Knowledge Resource; MetaMap. (2012). Retrieved January 6, 2013, from http://skr.nlm.nih.gov/interactive/metamap.shtml

Wang, R., & Zhang, Y. (2009). Recognizing Textual Relatedness with Predicate-Argument Structures. Paper presented at the Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore.

Ware, H., Mullett, C. J., Jagannathan, V., & El-Rawas, O. (2012). Machine learning-based coreference resolution of concepts in clinical documents. Journal of the American Medical Informatics Association, 19(5), 883-887. doi: 10.1136/amiajnl-2011-000774

Wattarujeekrit, T., Shah, P., & Collier, N. (2004). PASBio: Predicate-argument structures for event extraction in molecular biology. BMC Bioinformatics, 5(1), 155.

Wee, L. C., & Hassan, S. (2008). Exploiting Wikipedia for Directional Inferential Text Similarity. Paper presented at the Fifth International Conference on Information Technology: New Generations.

Wilbur, W. J., Rzhetsky, A., & Shatkay, H. (2006). New directions in biomedical text annotation: Definitions, guidelines and corpus construction. BMC Bioinformatics, 7(1), 356.

Xia, F., & Yetisgen-Yildiz, M. (2012). Clinical Corpus Annotation: Challenges and Strategies. Paper presented at the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining Workshop Programme, Istanbul, Turkey.

Xu, Y., Tsujii, J., & Chang, E. I.-C. (2012). Named entity recognition of follow-up and time information in 20 000 radiology reports. Journal of the American Medical Informatics Association, 19(5), 792-799. doi: 10.1136/amiajnl-2012-000812

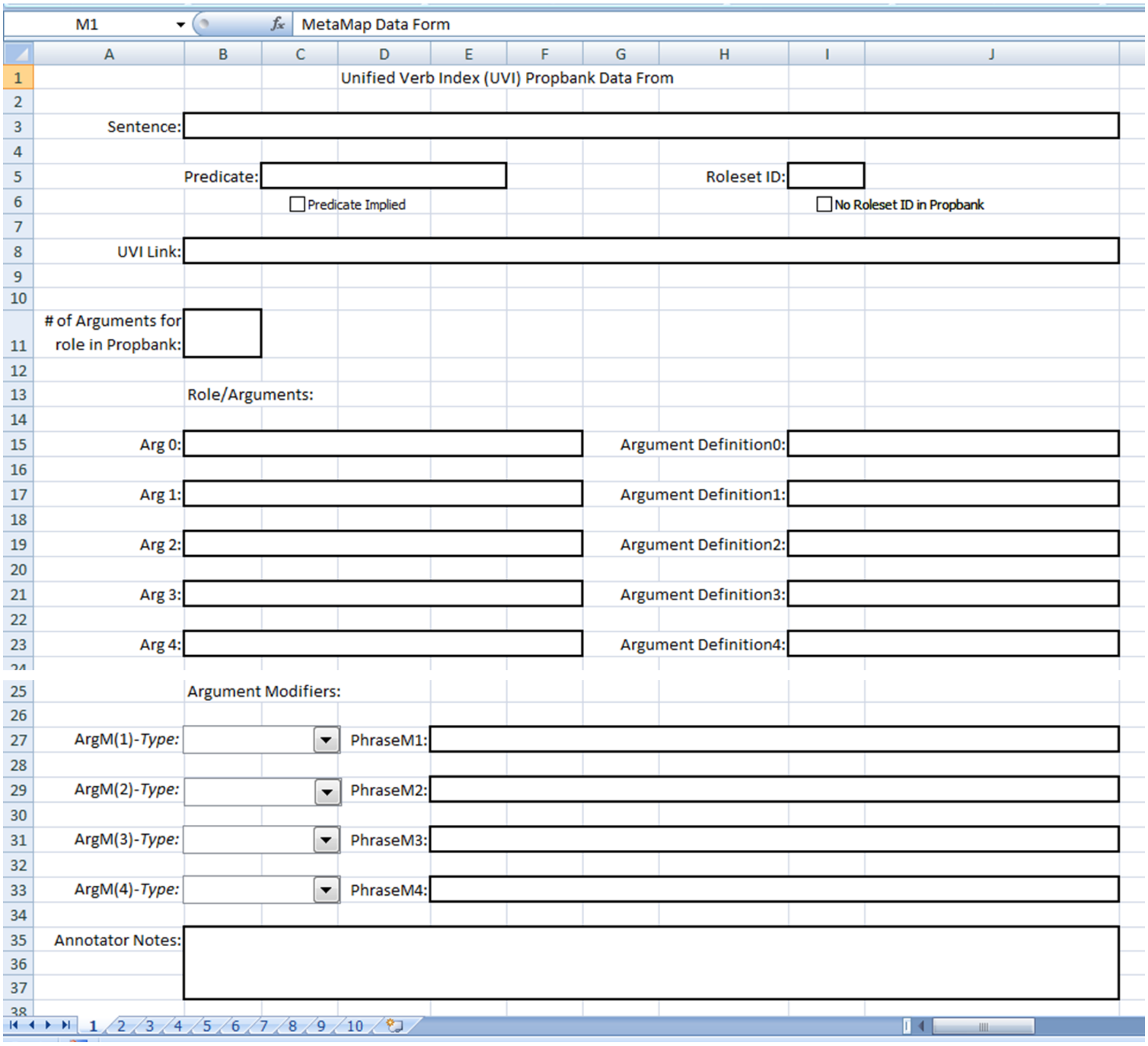
Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. The Qualitative Report, 16(3), 730-744.

Zielstorff, R. D., Tronni, C., Basque, J., Griffin, L. R., & Welebob, E. M. (1998). Mapping nursing diagnosis nomenclatures for coordinated care. Journal of Nursing Scholarship, 30(4), 369-373. doi: 10.1111/j.1547-5069.1998.tb01335.x

Zweigenbaum, P., & Demner-Fushman, D. (2009). Advanced literature-mining tools. In D. Edwards, J. Stajich & D. Hansen (Eds.), Bioinformatics (pp. 347-380). New York: Springer.

APPENDICES

Appendix A—Excel Worksheet Annotation Tool



Unified Verb Index (UVI) Propbank Data From

Sentence:

Predicate: Roleset ID:

☐ Predicate Implied ☐ No Roleset ID in Propbank

UVI Link:

# of Arguments for role in Propbank:

Role/Arguments:

Arg 0: Argument Definition0:

Arg 1: Argument Definition1:

Arg 2: Argument Definition2:

Arg 3: Argument Definition3:

Arg 4: Argument Definition4:

Argument Modifiers:

ArgM(1)-*Type:* PhraseM1:

ArgM(2)-*Type:* PhraseM2:

ArgM(3)-*Type:* PhraseM3:

ArgM(4)-*Type:* PhraseM4:

Annotator Notes:

Figure 41--Excel Worksheet Notation Aid Tool

Appendix B—Annotation Training/Guideline Manual

There are three phases to the data collection of this study:

- I. Annotation of Predicate Argument Structure
- II. Development of Logical Expression of Predicate Argument Structure (MetaMap)
- III. Scoring of Semantic Similarity

This manual serves to assist in first phase of data collection—annotation of radiology sentences into a Predicate Argument Structure (PAS). The focus will be on explaining PAS and dissecting sentences into PAS . This manual is sectioned into the following sections:

1. Study abstract, general introduction, and Propbank through the Unified Verb Index.
2. Step by Step examples for annotating sentences.
3. Working with the Excel workbook.
4. Implied predicate, No Matching Roleset ID, Verb 'be', Argument Modifiers
5. Training Sample Set and Process
6. Study Sample Set and Process

Section 1

Abstract:

The amount of information produced in the form of electronic free text is increasing to levels incapable of being processed by humans. This is certainly the case in the field of healthcare where professional journals, clinical research data, and electronic medical records hold an abundance of unprocessed knowledge potentially beneficial to improvement of health. Information extraction is a sub-field of natural language processing with the goal of data reduction. It specifically has a focus on understanding the semantics of text. Pertinent to IE is an annotated corpus that holds clues to how information extraction methods should process human decisions with meaning of sentences. Annotating a corpus requires considerable investment by domain experts, and most approaches to annotation have investigated corpus styles that involve minimal annotation. These studies consistently have problems addressing semantics and none have addressed the issue of semantic similarity which is a measurement necessary for reliable data reduction. It is not known if there is a benefit to maximal annotation. This study attempts to explore a method for assessing a sentential annotation style to determine if there is a need for larger annotation corpora. Predicate-argument structures will be the method of annotation and domain experts will use a scoring mechanism to rate semantic similarity between two unique structures. A convenient sample from a prior study with ten synsets of 100 unique sentences deemed by domain experts to mean the same thing will be the text from which predicate-argument structures are formed.

General Introduction:

PAS is a form of annotation that centers on the predicate of a sentence. Underlining the theory of PAS is the premise that a predicate assumes in a sentence a specific role. The roles that a verb may play in a sentence are a subfield of research in linguistics based on the sense or usage of the predicate. For the verb 'bag', there are two possible roles in the sense, or predicate, of 'to bag':

1) to gain something and/or place in location, and 2) discard or abandoned. For the verb 'lie', there are two senses, or predicates, of the verbs use, 'to lie' and 'to lie down'. For the predicate 'to lie', there are two possible roles: 1) assume a horizontal position, more generally 'exist', and 2) tell a false hood. For the predicate, 'to lie down', there are two possible roles: 1) a more explicit horizontal position, and 2) install or establish. Both predicates for the verb 'lie' have a role inferring some type of horizontal positioning, but one role is deemed to be more explicit than the other. For instance, in the sentence, 'After the robbery, the intruders lie low', the predicate's role is more specific than just a position in space. How to interpret which sense of a predicate is being used in a sentence is dependent on an individual's own experience and knowledge about language. This study explores the assumption that radiologists share a common interpretation of language used in the domain of radiology reports.

Any changes, deletions, or additions to a predicate's role are dependent on the use of the verb in a culture's language. A predicate's role will be determined by using Propbank referenced through the Unified Verb Index (http://verbs.colorado.edu/verb-index/index.php), which is maintained by the linguistics department at the University of Colorado. Probank will index these roles as a 'Roleset id'. For the verb 'bag', 'Roleset id: bag.01' and 'Roleset id: bag.02' are the two indexes. Propbank uses the label 'Rel' to label the predicate's role.

Centered on a role, the predicate role determines what semantic roles, or arguments, the remaining pieces of text of the sentence are assigned. The argument roles of a predicate can be agent (that which does the action), patient (that which receives the action), locative (where action is done), temporal (when action is done), manner (how action is done), and/or cause (what happened when action applied). Not all roles will be present in a sentence, and arguments may have modifiers, such as a negative (not) or temporal (whenever, when, or before). Modifiers are labeled as 'ArgM'. Propbank uses a dynamic argument structure based on the verb's role. This structure begins with Arg0 (typically the patient) all the way up to the needed number of arguments based on roles use in language (Arg1, Arg2). Each 'Roleset id' will define what the arguments are. Below is an example of the dynamic argument structure defined for the two 'Roleset ids' for the verb 'bag':

Roleset id: bag.01 Definition: to gain something and/or place in a location	Roleset id: bag.02 Definition: discard or abandon
Arguments: Arg0: gainer/placer Arg1: thing gained/placed Arg2: bag, location	Arguments: Arg0: abandoner Arg1: thing abandoned
Example1: Mark bags 10 lbs of potatoes before leaving store.	Example1: The nursing students bagged last Thursday's clinical.

Rel: bags Arg0: Mark Arg1: 10 lbs of potatoes Arg2: ArgM-temp: before leaving store	Rel: bagged Arg0: nursing students Arg1: last Thursday's clinical

Some roles may have more arguments than other roles, and arguments have different descriptions between the predicate roles. Also, a sentence may not have all arguments in it. Predicate roles in Propbank do not have a role for every 'verb' definition that may be found in a dictionary, such as Websters. For this study, the use of a verb or its definition will be limited to the Propbank definitions**. If Propbank does not offer a suitable role to the sense of the predicate in a study's sample sentence, annotators should note this**.

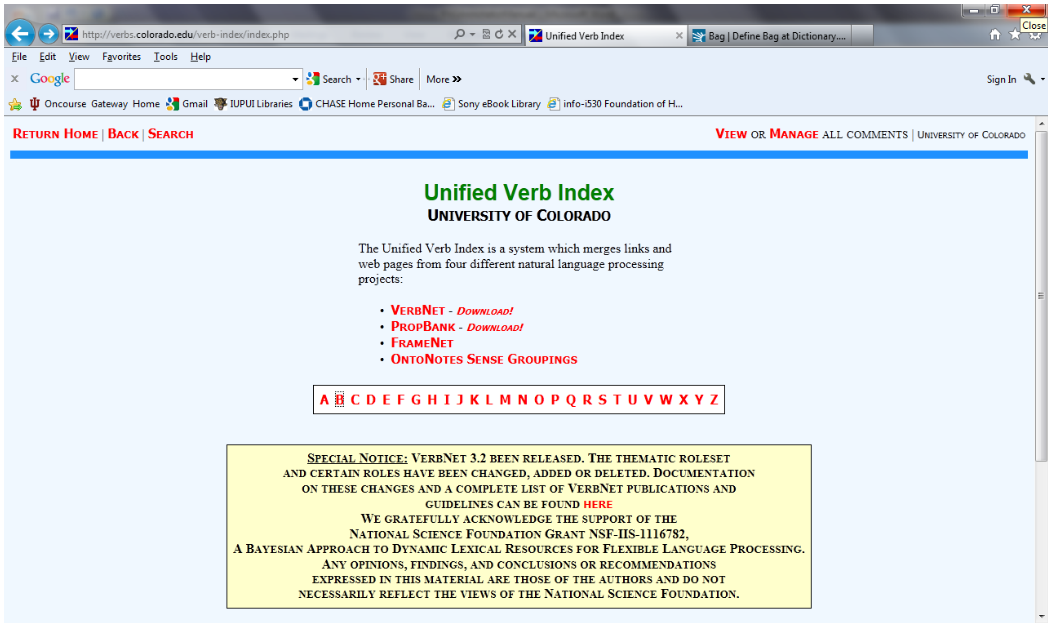
It is up to the annotator to accurately decipher the meaning of the text (determine the role the predicate plays in the sentence) and map the text phrases to the appropriate arguments. This is, by no means, an easy task. This manual will help annotators become familiar with Propbank and provide numerous examples of radiology sentences to train the 'thinking' process for sufficient sentential annotation.

Section 2

This sections is a point and click follow along guide for using Propbank in the Unified Verb Index. The example sentence for this exercise is:

The tip of the Corpak stylet lies behind the stomach in the duodenum.

1. Goto http://verbs.colorado.edu/verb-index/index.php. The webpage should look like the following: (Hint: Create a shortcut/favorite to this page!!)



RETURN HOME | BACK | SEARCH

VIEW OR MANAGE ALL COMMENTS | UNIVERSITY OF COLORADO

Unified Verb Index

UNIVERSITY OF COLORADO

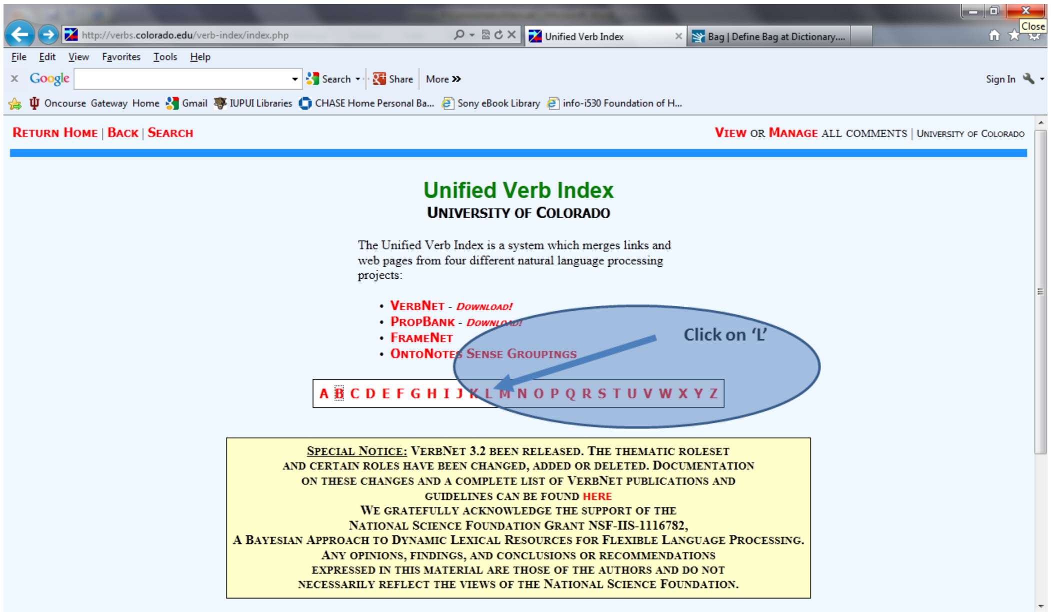
The Unified Verb Index is a system which merges links and web pages from four different natural language processing projects:

- VERBNET - *DOWNLOAD!*
- PROPBANK - *DOWNLOAD!*
- FRAMENET
- ONTONOTES SENSE GROUPINGS

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

SPECIAL NOTICE: VERBNET 3.2 BEEN RELEASED. THE THEMATIC ROLESET AND CERTAIN ROLES HAVE BEEN CHANGED, ADDED OR DELETED. DOCUMENTATION ON THESE CHANGES AND A COMPLETE LIST OF VERBNET PUBLICATIONS AND GUIDELINES CAN BE FOUND HERE
WE GRATEFULLY ACKNOWLEDGE THE SUPPORT OF THE NATIONAL SCIENCE FOUNDATION GRANT NSF-IIS-1116782,
A BAYESIAN APPROACH TO DYNAMIC LEXICAL RESOURCES FOR FLEXIBLE LANGUAGE PROCESSING.
ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS MATERIAL ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE NATIONAL SCIENCE FOUNDATION.

2. The verb in this sentence is ‘lies’. Click on the letter ‘L’:



RETURN HOME | BACK | SEARCH

VIEW OR MANAGE ALL COMMENTS | UNIVERSITY OF COLORADO

Unified Verb Index

UNIVERSITY OF COLORADO

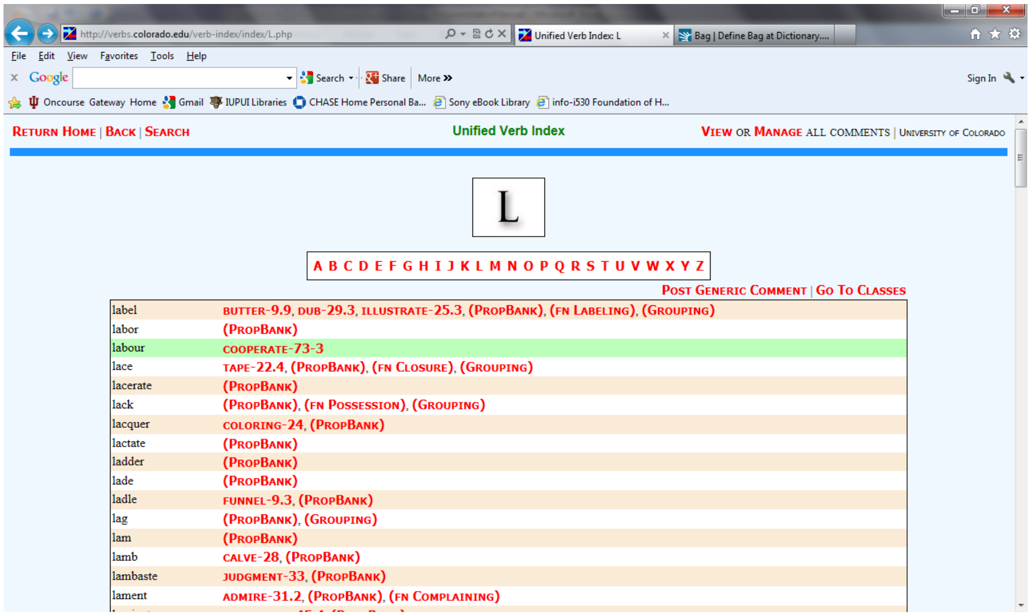
The Unified Verb Index is a system which merges links and web pages from four different natural language processing projects:

- VERBNET - *DOWNLOAD!*
- PROPBANK - *DOWNLOAD!*
- FRAMENET
- ONTONOTES SENSE GROUPINGS

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

SPECIAL NOTICE: VERBNET 3.2 BEEN RELEASED. THE THEMATIC ROLESET AND CERTAIN ROLES HAVE BEEN CHANGED, ADDED OR DELETED. DOCUMENTATION ON THESE CHANGES AND A COMPLETE LIST OF VERBNET PUBLICATIONS AND GUIDELINES CAN BE FOUND HERE
WE GRATEFULLY ACKNOWLEDGE THE SUPPORT OF THE NATIONAL SCIENCE FOUNDATION GRANT NSF-IIS-1116782,
A BAYESIAN APPROACH TO DYNAMIC LEXICAL RESOURCES FOR FLEXIBLE LANGUAGE PROCESSING.
ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS MATERIAL ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE NATIONAL SCIENCE FOUNDATION.

3. The following page will display listing all of the verb beginning with letter ‘L’:



Return Home | Back | Search

Unified Verb Index

View or Manage all comments | University of Colorado

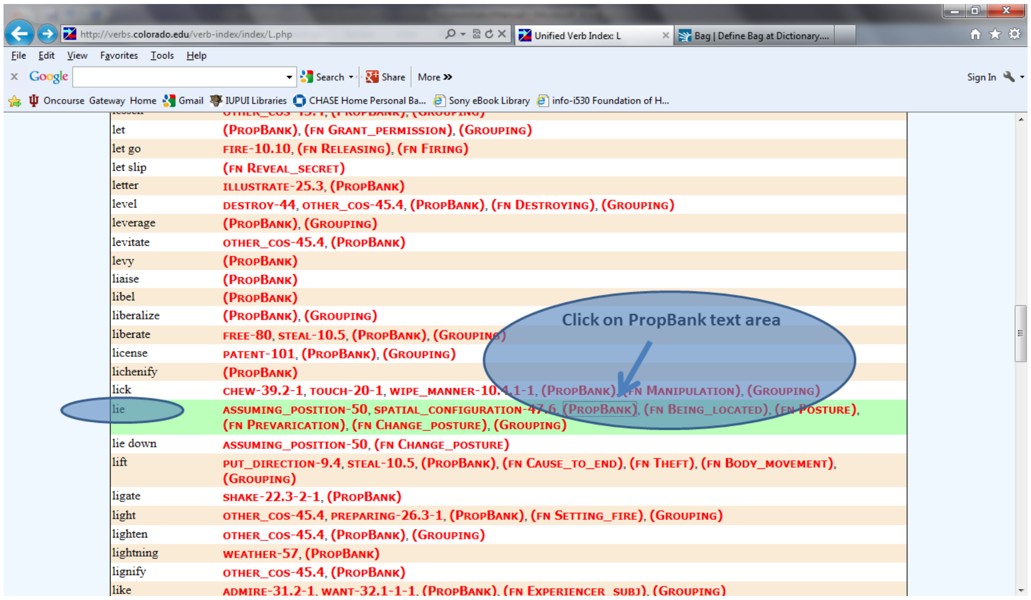
L

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Post Generic Comment | Go To Classes

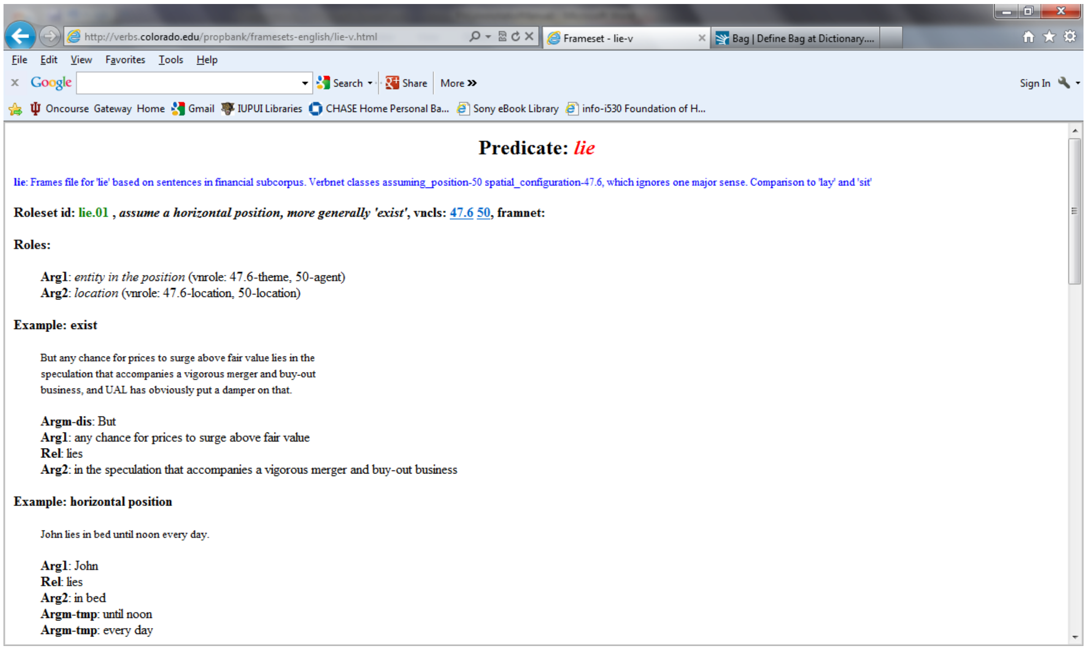
label	butter-9.9, dub-29.3, illustrate-25.3, (PropBank), (fn Labeling), (Grouping)
labor	(PropBank)
labour	cooperate-73-3
lace	tape-22.4, (PropBank), (fn Closure), (Grouping)
lacerate	(PropBank)
lack	(PropBank), (fn Possession), (Grouping)
lacquer	coloring-24, (PropBank)
lactate	(PropBank)
ladder	(PropBank)
lade	(PropBank)
ladle	funnel-9.3, (PropBank)
lag	(PropBank), (Grouping)
lam	(PropBank)
lamb	calve-28, (PropBank)
lambaste	judgment-33, (PropBank)
lament	admire-31.2, (PropBank), (fn Complaining)

4. Use scroll bar to find verb ‘lie’ (non conjugated form) and click on (PropBank) text area:



let	(PropBank), (fn Grant_permission), (Grouping)
let go	fire-10.10, (fn Releasing), (fn Firing)
let slip	(fn Reveal_secret)
letter	illustrate-25.3, (PropBank)
level	destroy-44, other_cos-45.4, (PropBank), (fn Destroying), (Grouping)
leverage	(PropBank), (Grouping)
levitate	other_cos-45.4, (PropBank)
levy	(PropBank)
liaise	(PropBank)
libel	(PropBank)
liberalize	(PropBank), (Grouping)
liberate	free-80, steal-10.5, (PropBank), (Grouping)
license	patent-101, (PropBank), (Grouping)
lichenify	(PropBank)
lick	chew-39.2-1, touch-20-1, wipe_manner-10.4.1-1, (PropBank), (fn Manipulation), (Grouping)
lie	assuming_position-50, spatial_configuration-47.6, (PropBank), (fn Being_located), (fn Posture), (fn Prevarication), (fn Change_posture), (Grouping)
lie down	assuming_position-50, (fn Change_posture)
lift	put_direction-9.4, steal-10.5, (PropBank), (fn Cause_to_end), (fn Theft), (fn Body_movement), (Grouping)
ligate	shake-22.3-2-1, (PropBank)
light	other_cos-45.4, preparing-26.3-1, (PropBank), (fn Setting_fire), (Grouping)
lighten	other_cos-45.4, (PropBank), (Grouping)
lightning	weather-57, (PropBank)
lignify	other_cos-45.4, (PropBank)
like	admire-31.2-1, want-32.1-1-1, (PropBank), (fn Experiencer_subj), (Grouping)

5. The following page will display (called ‘Frameset’):



Predicate: *lie*

lie: Frames file for 'lie' based on sentences in financial subcorpus. Verbnet classes assuming_position-50 spatial_configuration-47.6, which ignores one major sense. Comparison to 'lay' and 'sit'

Roleset id: lie.01 , *assume a horizontal position, more generally 'exist'*, vncls: 47.6 50, framnet:

Roles:

Arg1: *entity in the position* (vnrole: 47.6-theme, 50-agent)
Arg2: *location* (vnrole: 47.6-location, 50-location)

Example: exist

But any chance for prices to surge above fair value lies in the
speculation that accompanies a vigorous merger and buy-out
business, and UAL has obviously put a damper on that.

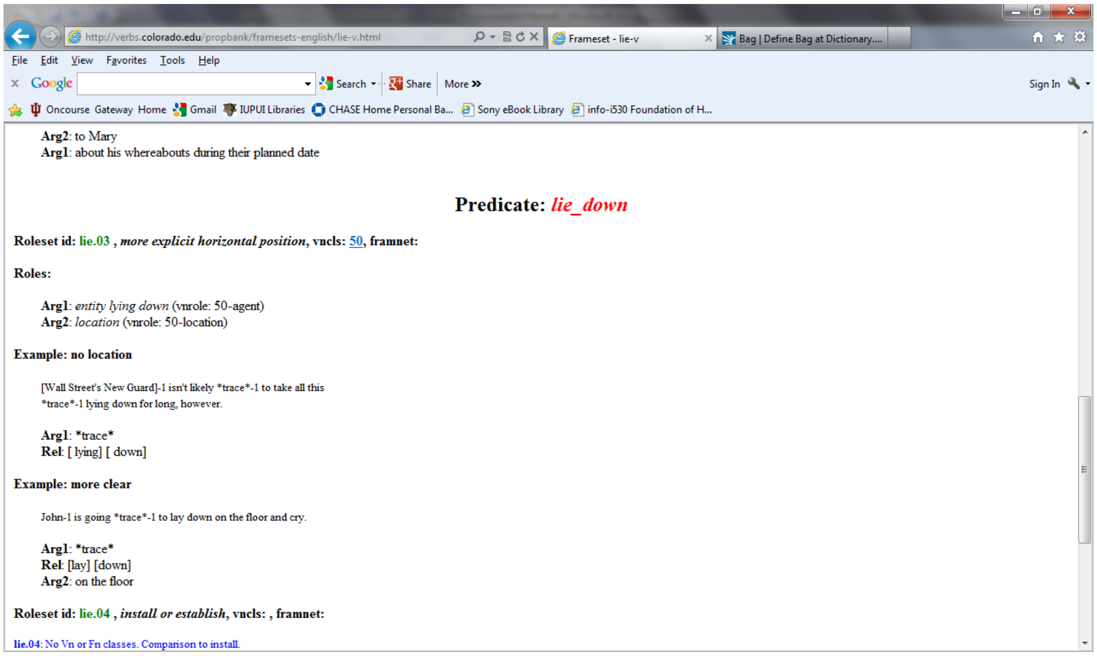
Argm-dis: But
Arg1: any chance for prices to surge above fair value
Rel: lies
Arg2: in the speculation that accompanies a vigorous merger and buy-out business

Example: horizontal position

John lies in bed until noon every day.

Arg1: John
Rel: lies
Arg2: in bed
Argm-tmp: until noon
Argm-tmp: every day

Use the scroll bar to see other predicate, 'to lie down' for the verb 'lies' as shown in following Frameset screen shot:



Arg2: to Mary
Arg1: about his whereabouts during their planned date

Predicate: *lie_down*

Roleset id: lie.03 , *more explicit horizontal position*, vncls: 50, framnet:

Roles:

Arg1: *entity lying down* (vnrole: 50-agent)
Arg2: *location* (vnrole: 50-location)

Example: no location

[Wall Street's New Guard]-1 isn't likely *trace*-1 to take all this
trace-1 lying down for long, however.

Arg1: *trace*
Rel: [lying] [down]

Example: more clear

John-1 is going *trace*-1 to lay down on the floor and cry.

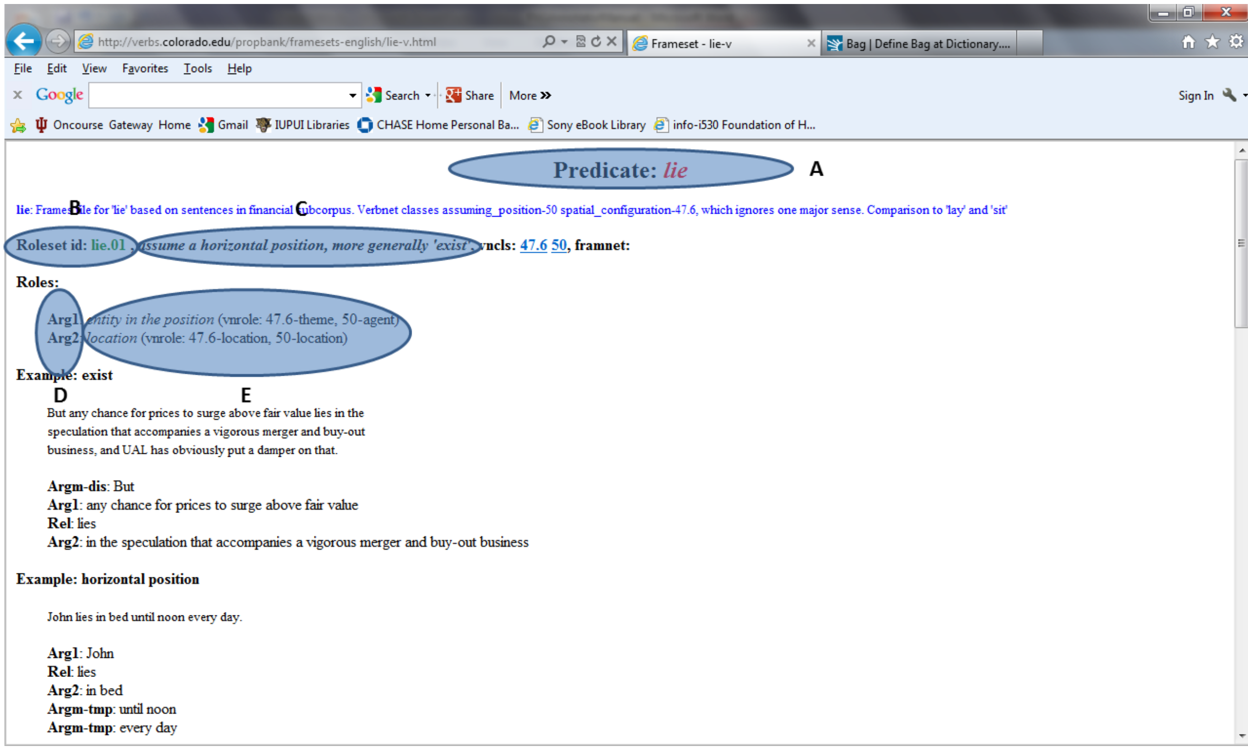
Arg1: *trace*
Rel: [lay] [down]
Arg2: on the floor

Roleset id: lie.04 , *install or establish*, vncls: , framnet:

lie.04: No Vn or Fn classes. Comparison to install.

Hint: Always scroll through entire page to overview predicates and roles before annotating sentence!

6. The following screen shot highlights general layout of a Propbank Frameset page:



Predicate: *lie*

lie: Frames file for 'lie' based on sentences in financial subcorpus. Verbnet classes assuming_position-50 spatial_configuration-47.6, which ignores one major sense. Comparison to 'lay' and 'sit'

Roleset id: lie.01 , ***assume a horizontal position, more generally 'exist'*** **, vncls: 47.6 50, framnet:**

Roles:

Arg1: *entity in the position* (vnrole: 47.6-theme, 50-agent)
Arg2: *location* (vnrole: 47.6-location, 50-location)

Example: exist

But any chance for prices to surge above fair value lies in the speculation that accompanies a vigorous merger and buy-out business, and UAL has obviously put a damper on that.

Argm-dis: But
Arg1: any chance for prices to surge above fair value
Rel: lies
Arg2: in the speculation that accompanies a vigorous merger and buy-out business

Example: horizontal position

John lies in bed until noon every day.

Arg1: John
Rel: lies
Arg2: in bed
Argm-tmp: until noon
Argm-tmp: every day

A – Verb's use in language, predicate
B – Roleset id index label, lie.01
C – Definition of predicate's role
D – Possible argument roles for roleset id
E – Definitions for argument roles

7. Review roles on Frameset (summarized below):

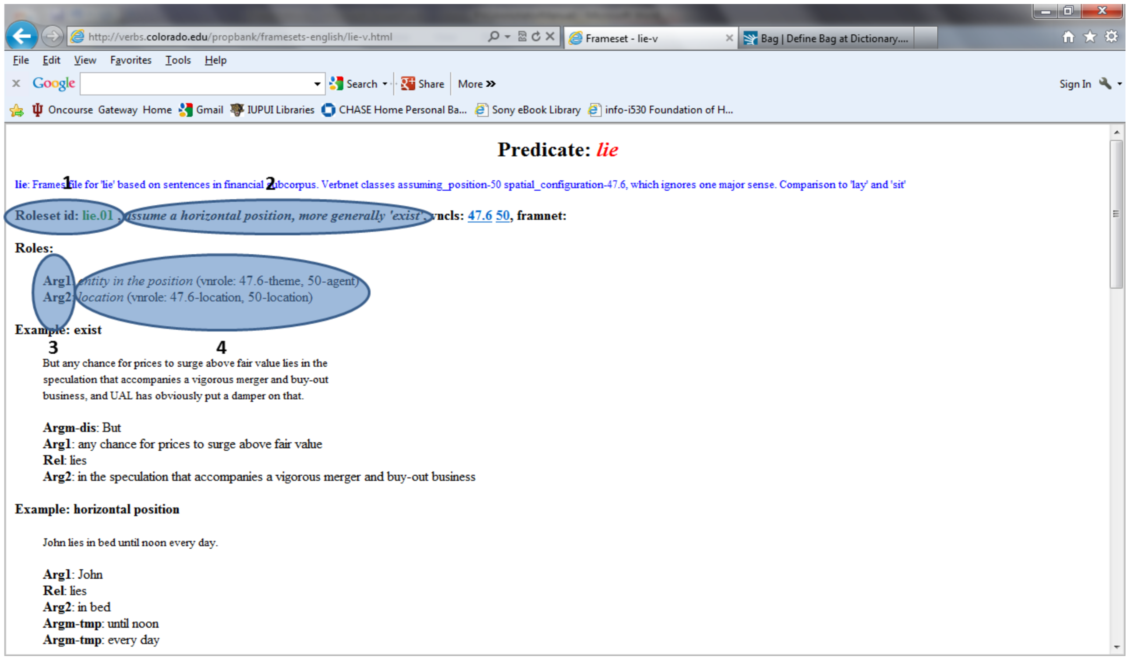
Verb: Lie (non-conjugated form)		
Predicate	Roleset ID	Definition
To lie	lie.01	assume a horizontal position, more generally 'exist'
	lie.02	tell a false hood
To lie down	lie.03	a more explicit horizontal position
	lie.04	install or establish

8. Select appropriate role.

In the example of our sentence, “The tip of the Corpak stylet lies behind the stomach in the duodenum,” the usage of ‘lies’ refers to a position or existence of a particular thing. Lie.01 is the appropriate role. Arguments roles for lie.01 will be used to annotate sentence.

9. Annotate sentence according to Roleset id arguments.

The tip of the Corpak stylet lies behind the stomach in the duodenum		
Argument	Definition	Corresponding phrase
Rel (lie.01[1]):	exist[2]	lies
Arg1[3]:	entity in the position[4]	‘The tip of the Corpak stylet’
Arg2[3]:	location[4]	‘behind the stomach in the duodenum’
[1,2,3,4]See screen shot below to see where information came from in the frameset		



Predicate: *lie*

lie: Frames file for 'lie' based on sentences in financial subcorpus. Verbnet classes assuming_position-50 spatial_configuration-47.6, which ignores one major sense. Comparison to 'lay' and 'sit'

Roleset id: lie.01 , *assume a horizontal position, more generally 'exist'*, **vncls:** 47.6 50, **framnet:**

Roles:

Arg1: *entity in the position* (vnrole: 47.6-theme, 50-agent)
Arg2: *location* (vnrole: 47.6-location, 50-location)

Example: exist

But any chance for prices to surge above fair value lies in the speculation that accompanies a vigorous merger and buy-out business, and UAL has obviously put a damper on that.

Argm-dis: But
Arg1: any chance for prices to surge above fair value
Rel: lies
Arg2: in the speculation that accompanies a vigorous merger and buy-out business

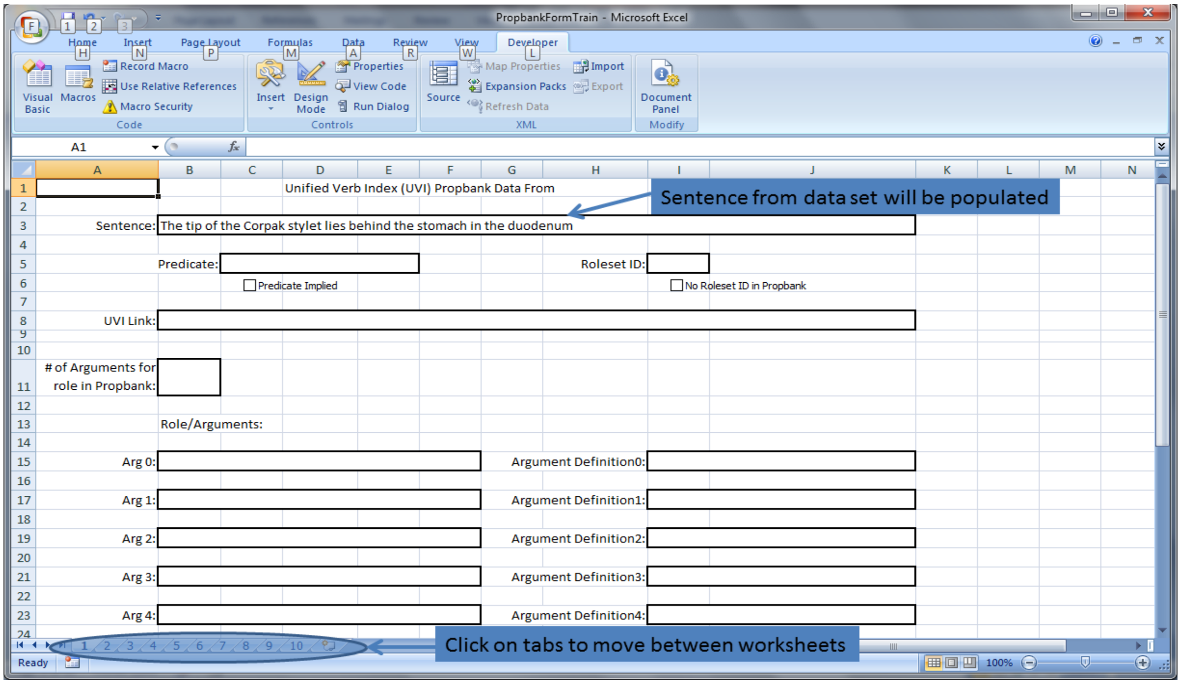
Example: horizontal position

John lies in bed until noon every day.

Arg1: John
Rel: lies
Arg2: in bed
Argm-tmp: until noon
Argm-tmp: every day

Section 3

This section explains the data collection form. An Excel workbook will be used that will contain multiple sheets, and each worksheet will provide a sentence from the study's sample set. The following two screen shots of the worksheet show the fields used to annotate Propbank data. The first screen shot shows the top part of the worksheet and the second screen shot shows the bottom part. Open up the file: PropbankFormTrain.xlsx to view the sheet. Notice that a sentence is already provided in the 'Sentence' field. Each tab at bottom represents a worksheet for one sentence in data sample set.



Unified Verb Index (UVI) Propbank Data From

Sentence: The tip of the Corpak stylet lies behind the stomach in the duodenum

Sentence from data set will be populated

Predicate:

☐ Predicate Implied

Roleset ID:

☐ No Roleset ID in Propbank

UVI Link:

# of Arguments for role in Propbank:

Role/Arguments:

Arg 0: Argument Definition0:

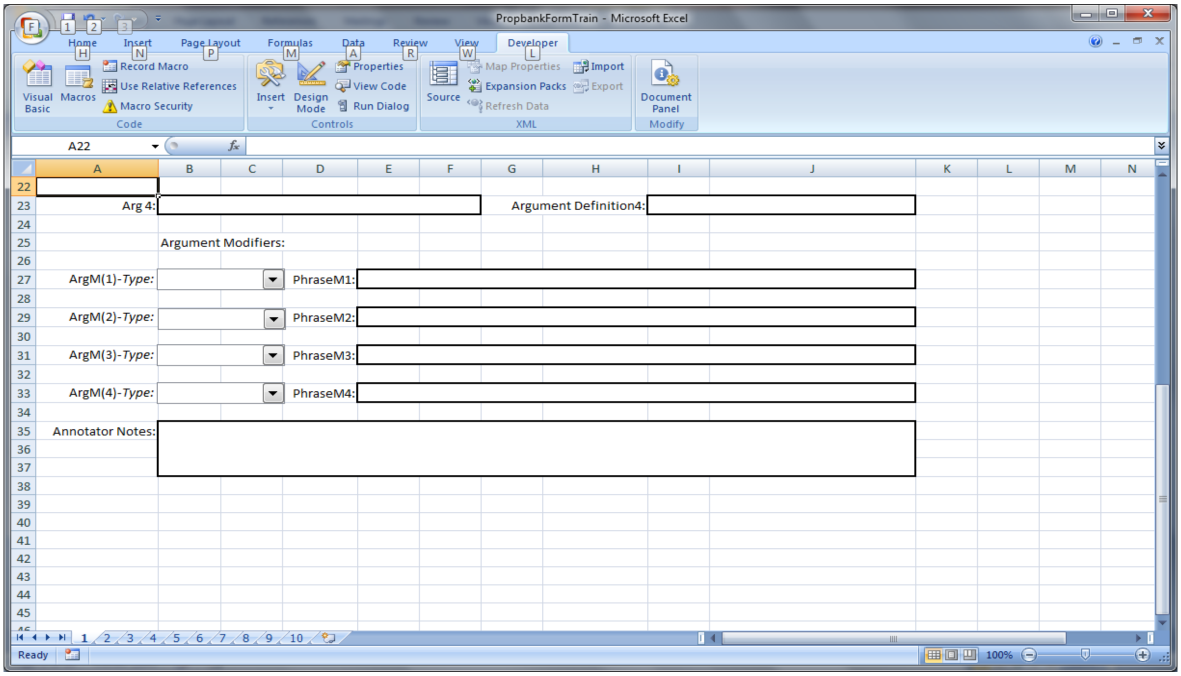
Arg 1: Argument Definition1:

Arg 2: Argument Definition2:

Arg 3: Argument Definition3:

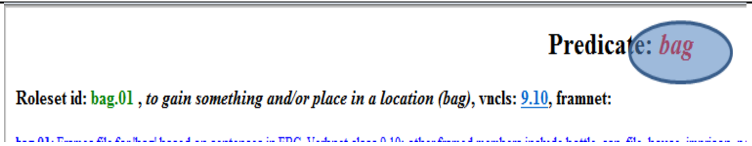
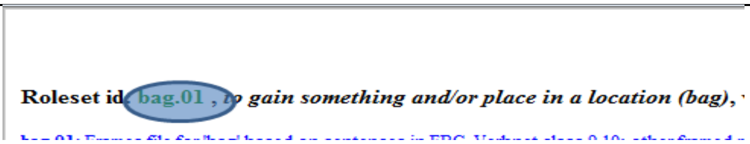
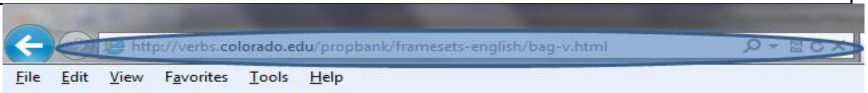
Arg 4: Argument Definition4:

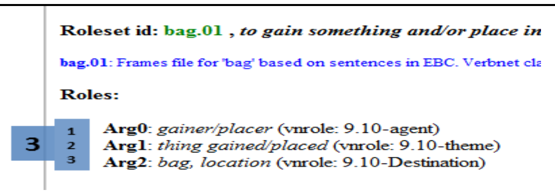
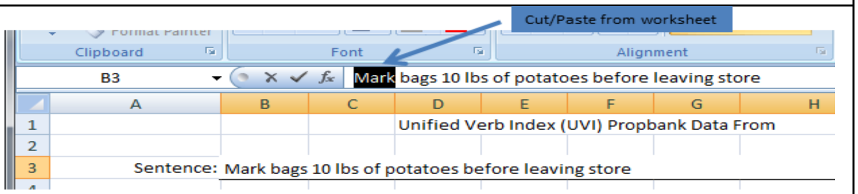
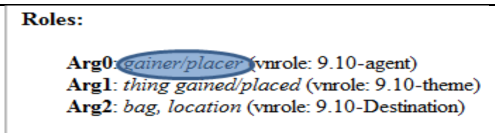
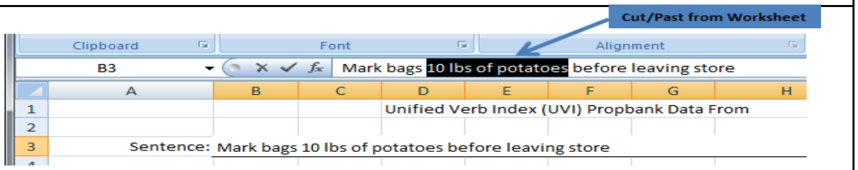
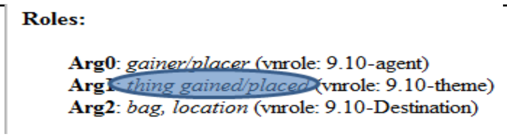
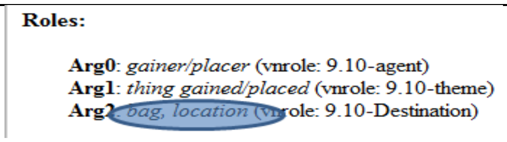
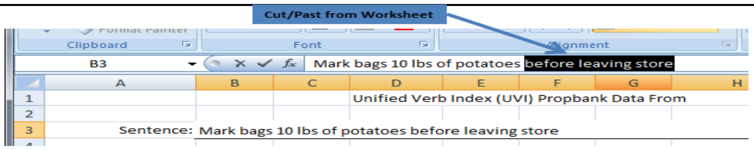
Click on tabs to move between worksheets



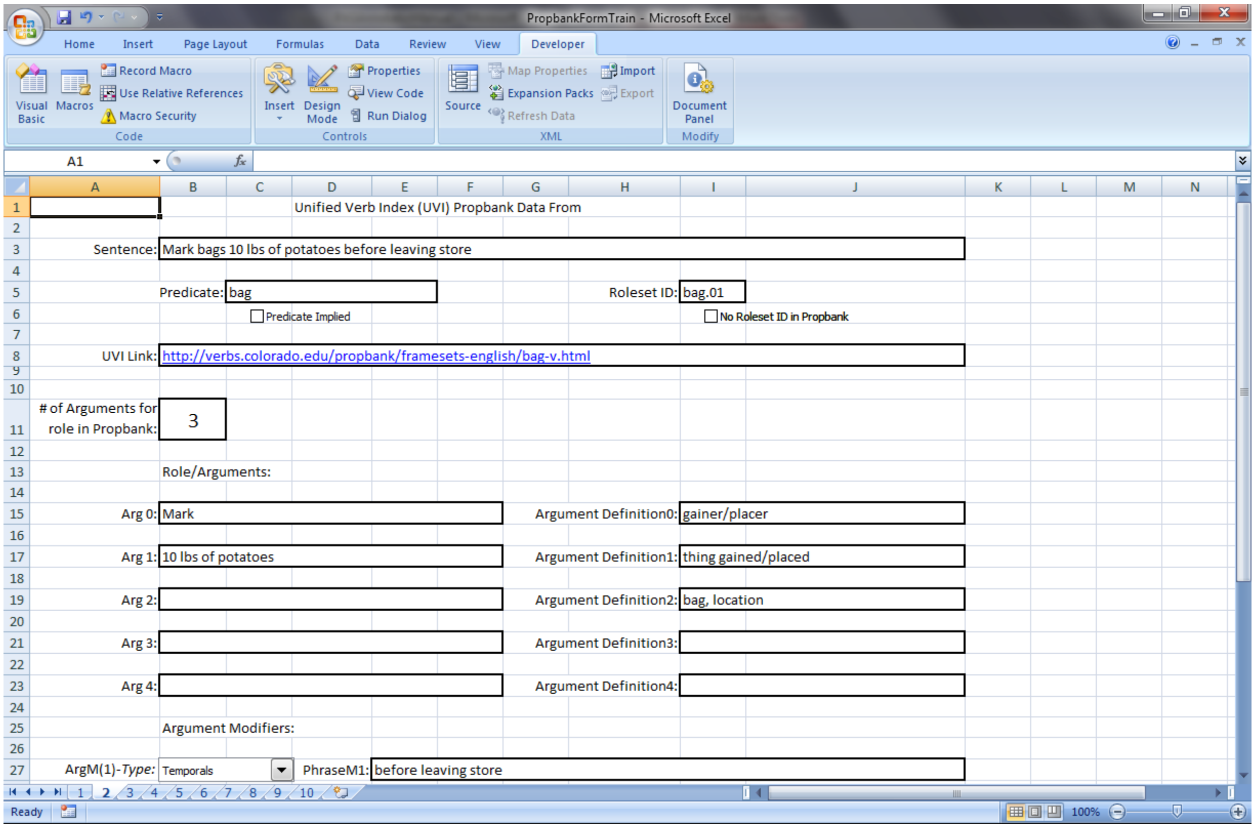
Arg 4:
Argument Definition4:
Argument Modifiers:
ArgM(1)-*Type*: PhraseM1:
ArgM(2)-*Type*: PhraseM2:
ArgM(3)-*Type*: PhraseM3:
ArgM(4)-*Type*: PhraseM4:
Annotator Notes:

Generally, an annotator should be able to cut paste from Propbank the 'pieces' of data to complete the data collection form. Click on tab '2' of PropbankFormTrain work book for a review of how the form is completed using the prior example sentence, "Mark bags 10lbs of potatoes before leaving the store." Fill in the form on tab 2 as you work through the fields in following table. Go to Unified Verb Index website to the Propbank Frameset for the verb 'bag' (linke: http://verbs.colorado.edu/propbank/framesets-english/bag-v.html).

Form Field	Data Comes From	Where Data Comes From
Predicate	Propbank Frameset	 **Predicate: *bag*** **Roleset id:** bag.01 , *to gain something and/or place in a location (bag)*, **vncls:** 9.10, **framnet:**
Roleset id	Propbank Frameset	 **Roleset id: bag.01 , *to gain something and/or place in a location (bag),***
Predicate Implied	Annotator Judgment	Discussed in next section
No Roleset ID in Propbank	Annotator Judgment	
UVI Link	Propbank Frameset	 http://verbs.colorado.edu/propbank/framesets-english/bag-v.html

# of Arguments for role in Propbank	Propbank Frameset	 **Roleset id:** bag.01 **, *to gain something and/or place in*** **bag.01**: Frames file for 'bag' based on sentences in EBC. Verbnet cla **Roles:** 3 1 2 3 **Arg0**: *gainer/placer* (vnrole: 9.10-agent) **Arg1**: *thing gained/placed* (vnrole: 9.10-theme) **Arg2**: *bag, location* (vnrole: 9.10-Destination)
Arg0	Sample Sentence	 B3 Mark bags 10 lbs of potatoes before leaving store 1 Unified Verb Index (UVI) Propbank Data From 2 3 Sentence: Mark bags 10 lbs of potatoes before leaving store
Argument Definition 0	Propbank Frameset	 **Roles:** **Arg0**: *gainer/placer* (vnrole: 9.10-agent) **Arg1**: *thing gained/placed* (vnrole: 9.10-theme) **Arg2**: *bag, location* (vnrole: 9.10-Destination)
Arg1	Sample Sentence	 B3 Mark bags 10 lbs of potatoes before leaving store 1 Unified Verb Index (UVI) Propbank Data From 2 3 Sentence: Mark bags 10 lbs of potatoes before leaving store
Argument Definition 1	Propbank Frameset	 **Roles:** **Arg0**: *gainer/placer* (vnrole: 9.10-agent) **Arg1**: *thing gained/placed* (vnrole: 9.10-theme) **Arg2**: *bag, location* (vnrole: 9.10-Destination)
Arg2	Sample Sentence	Blank, no data for this argument, but still fill in corresponding definition field!
Argument Definition 2	Propbank Frameset	 **Roles:** **Arg0**: *gainer/placer* (vnrole: 9.10-agent) **Arg1**: *thing gained/placed* (vnrole: 9.10-theme) **Arg2**: *bag, location* (vnrole: 9.10-Destination)
ArgM(1)-Type	Dropdown	Temporal (Modifiers will be explained in more depth in next section)
PhraseM1	Sample Sentence	 B3 Mark bags 10 lbs of potatoes before leaving store 1 Unified Verb Index (UVI) Propbank Data From 2 3 Sentence: Mark bags 10 lbs of potatoes before leaving store

When completed, the worksheet should look like the following screen shot:



Unified Verb Index (UVI) Propbank Data From

Sentence: Mark bags 10 lbs of potatoes before leaving store

Predicate: bag

Roleset ID: bag.01

☐ Predicate Implied

☐ No Roleset ID in Propbank

UVI Link: http://verbs.colorado.edu/propbank/framesets-english/bag-v.html

# of Arguments for role in Propbank: 3

Role/Arguments:

Arg 0: Mark

Argument Definition0: gainer/placer

Arg 1: 10 lbs of potatoes

Argument Definition1: thing gained/placed

Arg 2:

Argument Definition2: bag, location

Arg 3:

Argument Definition3:

Arg 4:

Argument Definition4:

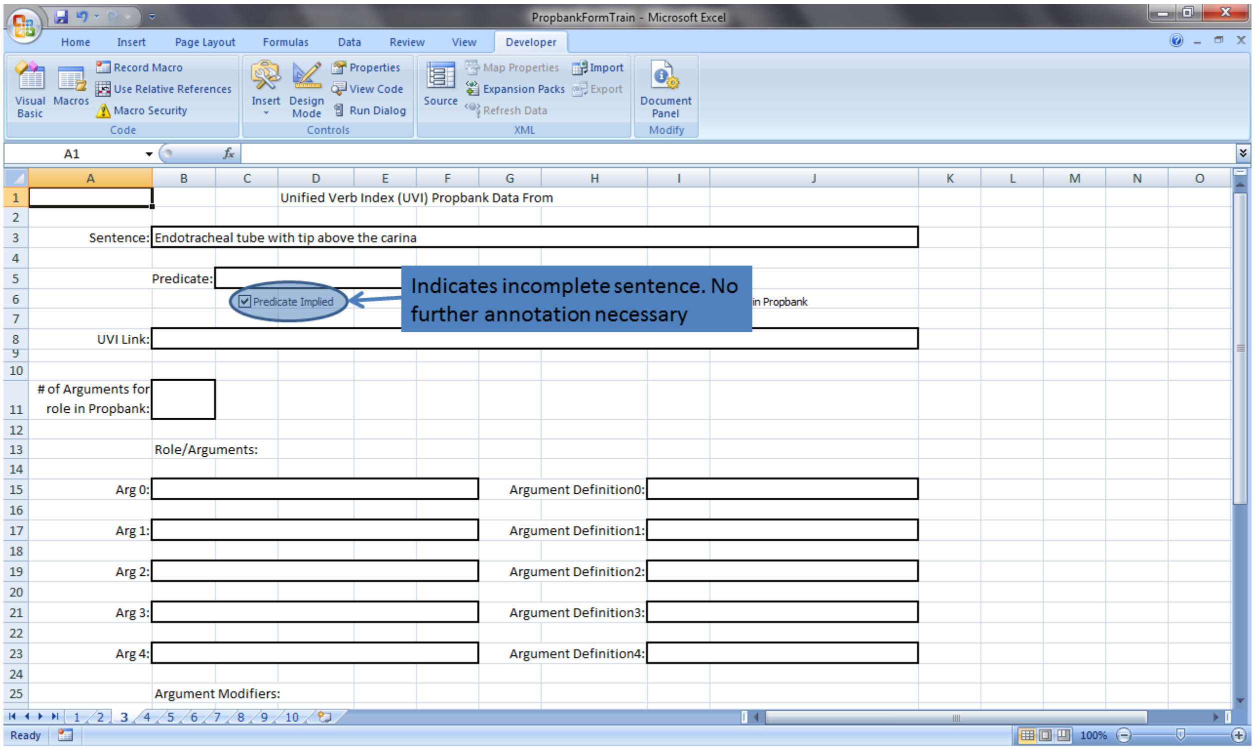
Argument Modifiers:

ArgM(1)-*Type:* Temporals

PhraseM1: before leaving store

Implied Predicate

By definition for this study, an incomplete sentence will be one that lacks a verb. If this is the case, the original author intended for the reader to fully understand the meaning of the sentence without the use of a verb. More than likely, the previous sentence or sentences helped frame a context for clear understanding. For this study, if a sentence lacks a predicate, all the annotator needs to do is check the box for the field ‘Predicate Implied’. These sentences will be reviewed for later processing. It may be necessary to provide a more complete passage of the radiology note for annotator to infer and annotate the implied predicate. The screen shot below shows show the worksheet should be completed for the sample sentence ‘Endotracheal tube with tip above the carina’:



Unified Verb Index (UVI) Propbank Data From

Sentence: Endotracheal tube with tip above the carina

Predicate:

☑ Predicate Implied

Indicates incomplete sentence. No further annotation necessary

in Propbank

UVI Link:

# of Arguments for role in Propbank:

Role/Arguments:

Arg 0: Argument Definition0:

Arg 1: Argument Definition1:

Arg 2: Argument Definition2:

Arg 3: Argument Definition3:

Arg 4: Argument Definition4:

Argument Modifiers:

No Match Roleset ID

As mentioned in the beginning of this manual, Propbank will be the reference from which to construct a PAS. It may be possible that an annotator will review a frameset for a verb and determine that there is no appropriate roleset ID for the predicate’s use in the sentence. If this is the case, the annotator should check the field ‘No Roleset ID in Propbank’. It is very important for the annotator to make a comment in the ‘Annotator Notes’ field at bottom of worksheet which reflect the thought process as to why Propbank does not have an appropriate roleset id for the verb. The intent is to have in the comments some explanation as to why the use of the verb in a radiology corpus does not fit the sense of predicates defined in Propbank.

Verb 'be'

Undoubtedly, the most used predicate will be some conjugated form of the verb"to be"—'be' in Propbank. Two important points should be made about certain predicate uses of the verb 'be'.

One: A common phrase is 'There are' or 'There is'. Such forms are an existential use of the predicate 'be', which is Roleset id: be.02. Be.02 has one argument, Arg1: thing that is but may have multiple argument modifiers (discussed following). If an annotator sees the predicate use, 'There are' or 'There is', he or she should automatically use roleset id: be.02 and complete the worksheet accordingly.

Two: Some sentences may be passive—using 'was' or 'were' before the verb (An MRI of the head was obtained). In passive sentences, an agent or an unnamed/implied agent does the action on the subject of the sentence. This means that the subject of the passive sentence becomes Arg1 in Propbank and the annotator should select the predicate/roleset id from the active form of the verb in the sentence. The annotation would be the following:

Rel: obtained (Roleset id: obtain.01)
Arg1: An MRI of the head
Arg0:

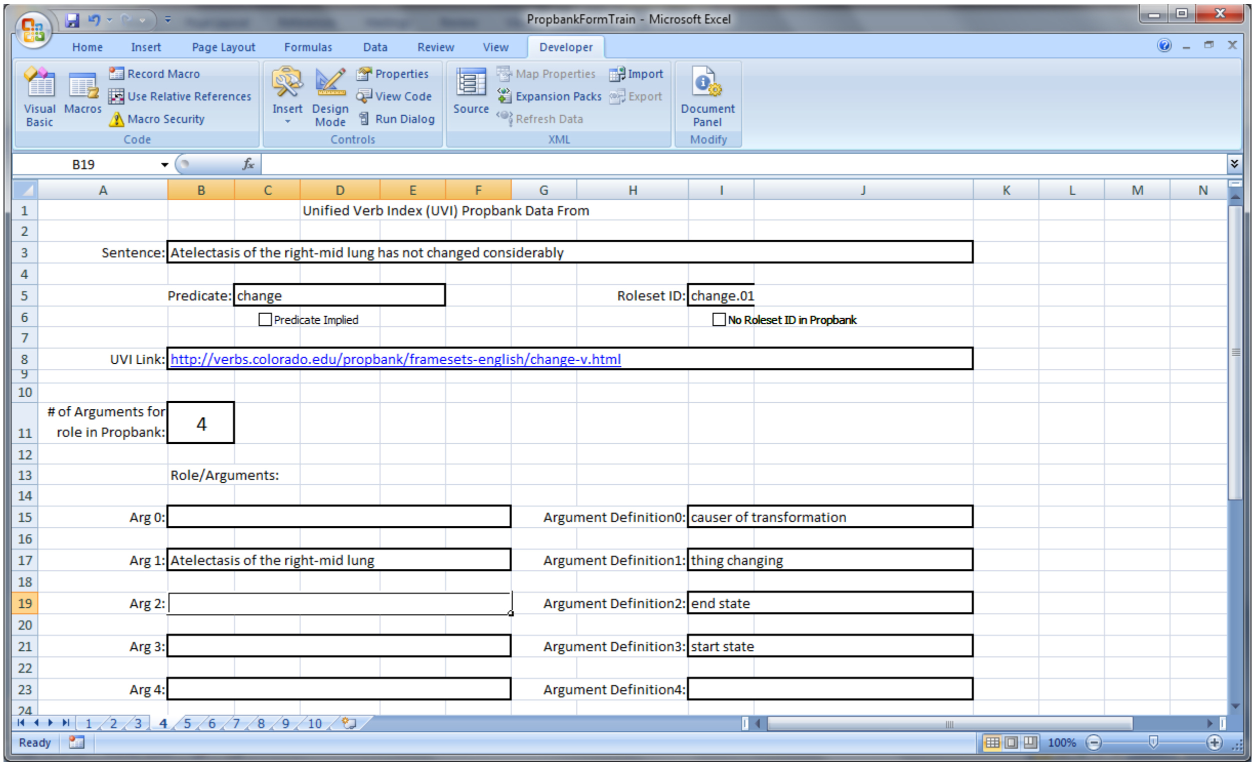
Argument Modifiers

Sentences may have modifiers—words/phrases that clarify meaning of arguments. Annotation of modifiers requires a high level competency of the sentential structure and of grammar. For this study, annotators are not expected to have such a competency level. The primary focus should be on correct interpretation of the sentence for appropriate roleset id and argument annotation. However, the worksheet provides for annotation of modifiers. One important modifier to annotate is negation (is not, does not), but the annotator is welcome to try to annotate other modifiers. Post conferences to annotation of PASs will address modifiers.

A following is a description of modifiers used in Propbank:

Directionals—modifiers show motion along some path
Locatives—modifiers indicate where some action takes place
Manner—adverbs specify how an action is performed
Temporal—show when an action took place
Extent—indicate amount of change occurring from an action (numerical adjuncts, quantifiers-a lot, and comparatives)
Reciprocals—reflexives and reciprocals (himself, itself, themselves)
Secondary Predication—show that an adjunct of a predicate is in itself capable of carrying some predicate structure
Purpose—used to show the motivation for some action
Cause—reason for an action
Discourse—connect a sentence to a preceding sentence
Adverbials—syntactic elements which clearly modify the event structure of the verb
Modals—"will, may, can, must, shall, might, should, could, would"
Negation—not, n't, never, no longer, etc.

For the follow sentence, “Atelectasis of the right-mid lung has not changed considerably”, correct annotation of arguments and modifiers is shown if following screen shots:



Unified Verb Index (UVI) Propbank Data From

Sentence: Atelectasis of the right-mid lung has not changed considerably

Predicate: change

Roleset ID: change.01

☐ Predicate Implied

☐ No Roleset ID in Propbank

UVI Link: http://verbs.colorado.edu/propbank/framesets-english/change-v.html

# of Arguments for role in Propbank: 4

Role/Arguments:

Arg 0:

Argument Definition0: causer of transformation

Arg 1: Atelectasis of the right-mid lung

Argument Definition1: thing changing

Arg 2:

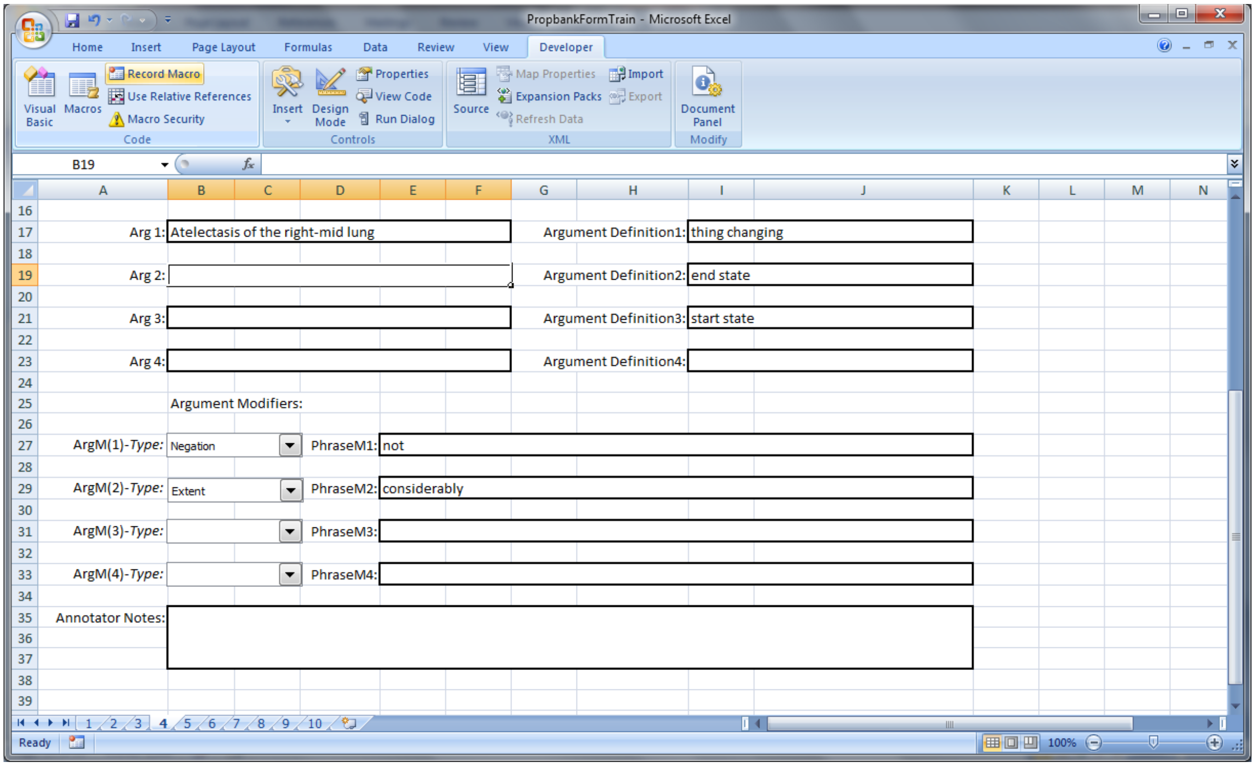
Argument Definition2: end state

Arg 3:

Argument Definition3: start state

Arg 4:

Argument Definition4:



Arg 1: Atelectasis of the right-mid lung

Argument Definition1: thing changing

Arg 2:

Argument Definition2: end state

Arg 3:

Argument Definition3: start state

Arg 4:

Argument Definition4:

Argument Modifiers:

ArgM(1)-*Type*: Negation

PhraseM1: not

ArgM(2)-*Type*: Extent

PhraseM2: considerably

ArgM(3)-*Type*:

PhraseM3:

ArgM(4)-*Type*:

PhraseM4:

Annotator Notes:

Section 5

After completion of this manual and introduction to study, annotators will be provided with a series of Excel workbooks for training purposes. Each workbook will have multiple worksheets with each worksheet corresponding to a sentence to be annotated. Once all sentences have been annotated, annotators will email completed workbook back to study contact who will review and return to annotator with comments. Conference calls may be scheduled to address important points. The purpose of the training is to familiarize annotator with the annotation process using Propbank and data collection worksheet. It is important for annotators to develop the habit of making comments on worksheet concerning questions, problems, and issues—especially thoughts pertaining to knowledge of radiology sentential structure. Comments will be part of the study's data analysis and analysis will focus on particular traits a corpus of radiology sentences have using PAS to develop a logical expression.

For each sentence, follow these steps to develop a good annotation skill:

1. Read sentence in its entirety without making entries in worksheet
2. Read sentence 2nd time and enter predicate in worksheet
3. Look up predicate in Propbank and enter appropriate Roleset ID into worksheet
4. Copy corresponding argument descriptions from Propbank into corresponding entry in worksheet
5. Go through sentence and copy text phrase in sentence that corresponds to argument descriptions and paste on worksheet
6. Go through sentence and copy text phrase in sentence that corresponds to a modifier and paste on worksheet selecting in drop down box type of modifier
7. Review annotation that all text of sentence has been annotated into worksheet
8. Record any questions, uncertainties, ambiguities, problems, or issues in comment section of worksheet.

Section 6

Once training data set is completed, annotators will begin receiving workbooks with 10 worksheets (10 sentences). The process will not be much different except that follow-up to each workbook will be dependent on review of both annotators (inter-rater reliability). A due date will be set for completion of each workbook so that a timely conference can be scheduled to review and address annotation problems. Once a PAS has been approved for all sentences in the workbook, annotators will be emailed the following workbook. Again, comments reflective of the annotator’s thoughts during the annotation are imperative to this study!!

Again, following these steps for good annotation:

1. Read sentence in its entirety without making entries in worksheet
2. Read sentence 2nd time and enter predicate in worksheet
3. Look up predicate in Propbank and enter appropriate Roleset ID into worksheet
4. Copy corresponding argument descriptions from Propbank into corresponding entry in worksheet
5. Go through sentence and copy text phrase in sentence that corresponds to argument descriptions and paste on worksheet
6. Go through sentence and copy text phrase in sentence that corresponds to a modifier and paste on worksheet selecting in drop down box type of modifier
7. Review annotation that all text of sentence has been annotated into worksheet
8. Record any questions, uncertainties, ambiguities, problems, or issues in comment section of worksheet.

Eric Tyner Newsom

enewsom@iupui.edu
(317) 313-8397
Indianapolis, IN

Education

Master of Science in Healthcare Informatics, Expected May 2013
School of Informatics, Indiana University Purdue University at Indianapolis (IUPUI)
Thesis: An Exploratory Study Using The Predicate-Argument Structure To Develop Methodology For Measuring Semantic Similarity of Radiology Sentences
Advisor: Josette Jones

Bachelor of Science in Nursing, May 2009
School of Nursing, Indiana University Purdue University at Indianapolis (IUPUI)
Honors Project: Use of Nursing Management Minimum Data Set (NMMDS) for a Focused Information Retrieval

Bachelor of Arts in English Teaching, December 1988
School of Education/Liberal Arts, University of Indianapolis
Indianapolis, IN

Research Interests

- Natural Language Processing
- Patient Portals

Experiences

Indiana University Health
Indianapolis, IN

- Clinical IS Educator, February 2013 to current
- Associate Partner, RN, May 2009 to February 2013

Publications

Jones, J., Newsom, E., & Delaney, C. (2009). Use of Nursing Management Minimum Data Set (NMMDS) for a Focused Information Retrieval. Paper presented at the 13th International Conference on Human Computer-Interaction, San Diego, CA.

Monsen, K. A., & Newsom, E. T. (2011). Feasibility of Using the Omaha System to Represent Public Health Nurse Manager Interventions. Public Health Nursing, 28(5), 421-428.