8-2016

Tool comparison of semantic parsers

Robert Hinh
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Robert Hinh

Entitled
TOOL COMPARISON OF SEMANTIC PARSERS

For the degree of Master of Science

Is approved by the final examining committee:

Julia Taylor-Rayz
Chair

John Springer

Victor Raskin

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Julia Taylor-Rayz

Approved by: Jeffrey Whitten 7/18/2016

Head of the Departmental Graduate Program Date

TOOL COMPARISON OF SEMANTIC PARSERS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Robert Hinh

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

August 2016

Purdue University

West Lafayette, Indiana

To my father, be.

You were right, two years went by pretty fast.

ACKNOWLEDGMENTS

I would like to, first and foremost, thank my major professor, Julia MT Rayz, for her guidance, patience, and support while working with me.

I would like to also extend my appreciation to my committee members, Victor Raskin and John Springer for their insightfulness and guidance, both for this thesis and my academics.

Lastly, I would like to thank my family and friends who've supported me in this endeavor these past 2(ish) years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Boxer DRT	Boxer Discourse Representation Theory
CCG	Combinatory Categorical Grammar
CoNLL	Conference on Computatiuonal Natural Language
EI	Event Identification
HAL	Hyperspace Analogue to Language
LSA	Latent Semantic Analysis
MUC	Message Understanding Conference
NEI	Named Entity Identification
NER	Named Entitiy Recognition
NLP	Natural Language Processing
OI	Object Identification
PR	Predicate Relationship
SEMAFOR	SEMantic Analyzer of Frame Representations
SemEval	Semantic Evaluation
XML	eXtensible Markup Language

ABSTRACT

Hinh, Robert M.S., Purdue University, August 2016. Tool Comparison of Semantic Parsers. Major Professor: Julia M. Taylor-Rayz.

Natural Language Processing (NLP) is a vital aspect for artificial intelligence systems to achieve integration into human lives, which has been a goal for researchers in this industry. While NLP focuses on an array of problems, semantic parsing will be specifically focused on throughout this paper. These parsers have been considerably targeted for improvement through the scientific community and demand for semantic parsers that achieve high accuracy has increased. There have been many approaches developed for this specific purpose and in this paper, a deep analysis was performed to compare the performance of semantic parsing systems. The implications of this comparison provides a viewpoint of how semantic parsers from different eras compare on a set of shared metrics.

CHAPTER 1. INTRODUCTION

Natural language is an sophisticated phenomena where a combination of phonetics, syntax, morphemes, pragmatics, semantics, etc., all form a complex system for communication between two users. While there are rules to help govern communication at each of these linguistic levels, these rules are loosely defined and are open to different interpretations. Moreover, there has been much interest for a push in automation tools to support various types of tasks that involve natural language, such as sentiment analysis, translation, etc. In combination of these approaches and advances in data storage and processing power, a push for processing natural language text computationally is a progressive fit.

There has been many types of automated tools developed within each linguistic level that have been developed to address the individual problems at each level. However, while many computational solutions exist for various problems at each linguistic level, which is overwhelming, only semantic properties will be examined. More specifically, semantic parsers are of interest in this thesis.

1.1 Scope

In this thesis, semantic properties extracted from available tools are primarily the focus. More specifically, semantic parsing tools is of particular interest as many emerging technologies utilize these types of systems however, many systems are inconsistent in accuracy and precision. The amount of information extracted from each parser on natural language text varies and will be compared.

1.2 Statement of Problem

Currently, there have been some collective initiatives that aim to bring researchers in this community together to focus on the issue of semantic parsing system. The most notable initiatives are the CoNLL (Carreras & Marquez, 2005) shared tasks, ranging from the late 1990's til now, and SemEval (Edmonds, 2002) shared tasks, ranging from the same time of CoNLL. These events provide an opportunity to showcase and evaluate new ideas for various problems in natural language processing by providing researchers a common purpose to train, test and present their systems to the industry. However, although there is an overwhelming number of semantic parsers commonly that are referred to as benchmarks in the community, these comparisons use quantifiable metrics such as accuracy and precision are used to determine the "better" systems. While accuracy and precision are helpful metrics to determine relative performance to other systems, there should be more focus on these systems extracting semantic information. Therefore, this thesis is focused on comparing semantic parsing tools from a variety of benchmark systems while using metrics focused more on the semantics of a sentence.

1.3 Significance

There have been many different types of approaches developed, specifically for semantic parsing as it is a particular interest to many industries, including security (Sheth et al., 2005), biomedical (Zhou & He, 2011) and many more. These approaches are comprised of different types of systems that support their process. The CoNLL and SemEval shared tasks provide an organized platform for comparison but for semantic parsers that are submitted to these events. Furthermore, there is a very low number of duplicate entries from the same semantic parsing tool. This shows that there has not been a collective side-by-side comparison of these tools currently. Therefore, by providing a comparison of these tools from different events, it provides a basic and comprehensive starting pointing to understand semantic parsing systems.

1.4 Assumptions

As mentioned, natural language contains many different linguistic levels (syntax, pragmatics, etc.) that are important in processing text. However, only semantic parsers are focused within this thesis.

1.5 Limitations

For each natural language tool at each linguistic level, there are various degrees of quality, quantity and completion to address the issues to processing text. There are some issues where more resources have been invested compared to other issues and semantic parsing tools are no exception. That being stated, there are a plethora of semantic parsing tools that have been, and always will be, developed. Some of these tools are available to the general public but many are not available. Therefore, while there are many semantic parsers that are desired in this study, these tools are not included in this comparison as the developers did not make their systems publicly available.

1.6 Delimitations

Due to the nature of this study, the only interest is comparing the performance of semantic parsers. Therefore, any external system that resides outside the scope of each parser will not be held accountable for that specific parser. For example, in pipeline systems (systems that are comprised of many smaller systems), the process of extracting semantic information from natural language text is of only interest and not the other separate systems that may be dependent.

1.6.1 Definitions of semantic parsers

While semantics parsers may have many different valid definitions as to what its abilities and limits are, when discussed in this thesis, semantic parsers only analyze sentences information about the specific metrics outlined in this thesis. There may very well be other types of "semantic parsers", which have different basis of theories, including the use of ontological semantics, and others. However, these will not be further explored.

1.7 Summary

This chapter provides the scope, significance, research question, assumptions, limitations, delimitations and other background information for this thesis. The next chapter provides a review of the literature relevant to for popular semantic parsing tools.

CHAPTER 2. REVIEW OF RELEVANT LITERATURE

In this section, the different types of semantic parsers and core features that make them distinct, metrics used, and the datasets used, are explored. Section 2.1 identifies the different types of semantic parsing families. Section 2.2 identifies the semantic parsers that were observed and will be used for further inspection. Section 2.3 targets the metrics that were tested on each parser. In addition, additional metrics were brought into account to combat the meaning of a semantic parser.

Semantics parsers are automated tools that focus on adding additional layers of semantic information to natural language. It is a difficult task with many approaches having different levels of success. Throughout this thesis, parsers are analyzed based on the semantic information that is generated on a given sentence. Interesting enough, out of all the many different types of parsers developed, many of these can be traced back to sets of core shared traits. These shared traits utilize frame semantics (Fillmore, 1982), distributional semantics (Lenci, 2008) and set theory logic (also known as predicate logic, first order logic, higher order logic, etc.) (Van Emden & Kowalski, 1976). The researcher will grade each semantic parser on the processing portion of parsing natural language text in English.

2.1 Semantic Parsing Families

2.1.1 Frame Semantics

Frame semantics is a concept originally developed by as an organized way to categorize generic definitions into their correct meanings and senses (Fillmore, 1982). To illustrate, in Fillmores classic example of breakfast, one may think some definitions could be:

- [Breakfast is] the first of three meals eaten in a given day or
- [Breakfast is] a meal that is usually eaten after a period of sleep

While these definitions are certainly valid and not the only definitions, Fillmore argues that there are many scenarios where the use of breakfast can contradict this definition. For example, a person wakes up at four in the afternoon, eats a meal and tells a friend that they ate breakfast at three in the afternoon. While it may be strange to some, ultimately, the word breakfast is used in this scenario that does not fit the definition and is understood despite contradicting a listed definition(s). Rather than explicitly listing all definitions of breakfast (and for that matter, all definitions for all words), frame semantics instead looks at the pattern of definitions. From all definitions of a particular word, there are elements that are shared common devices. And these shared elements can be used as a generic definition to define words instead of using explicit definitions.

A popular implementation of frame semantics is called FrameNet (Baker, Fillmore, & Lowe, 1998). This initiative is a database that contains frames and metadata information about each frame, such as its targets, semantic roles that it can support, lexical units, and more. In addition, each frame is a unique word sense of a given lemma. FrameNet is useful for applications that incorporate frame semantics into their parsing systems as this database is a growing project that contains hundreds of existing frames with many more lexical units. Lexical units are words that can force a frame to be evoked.

While there are benefits to using FrameNet show that it is an immensive lexical data source that contains considerable predefined semantic information, there are limitations with this approach as outlined by Palmer and Sporleder (Palmer & Sporleder, 2010). It was argued that while FrameNet has the ability to produce deeper knowledge compared to other frame-like systems (Ellsworth, Erk, Kingsbury,

& Padó, 2004), limitations of FrameNet include: insufficient training data of lexical units for frames and frames that are missing from the system (Palmer & Sporleder, 2010). In addition, FrameNet account for only 11,000 lexical units, which is lower compared to other lexical systems (Baker & Fellbaum, 2009). Due to these limitations, other approaches that rely on frame semantics have been developed. More specifically, the use of Propbank and nombank as an annotated lexical source.

Propbank is conceptually similar to FrameNet where this system utilizes frames, however, Propbank is another annotated database of frames that solely focuses on verbs from the Penn treebank corpus (Kingsbury, Palmer, & Marcus, 2002). It contains predicates identified and the types of arguments (roles) that the predicate can support. In addition, while Propbank concentrates only on verbs, nombank is a separate initiative that complements Propbank by focusing on nouns from the same Penn treebank corpus (Meyers et al., 2004).

2.1.2 Distributional Semantics

The next major distinction of semantic parsers is the use of distributional semantics, which is the concept of identifying the meaning of words based on the number of co-occurrences that the target word is associated (Harris, 1954). In addition to finding the physical co-occurrences of words, these systems also help by finding additional features that can be used to identify the strength of the relationship. One popular method in particular in this field is known as wordspace models. The example below helps illustrate the rationale with wordspace models (Erk & Padó, 2008):

- Catch a ball.
- Catch a disease.
- Attend a ball.

In the first two sentences, the intended meaning of both sentences is a person obtaining a physical entity (a ball and disease). The third sentence is a person that is being present at an event. In this small example, the system should be able to disambiguate ball as one of either two actions, a physical object or a social event, based on the event predicate within the sentence. In other words, information about a specific word or phrase can be derived from the context from which it was taken.

This branch of analyzing semantics has been utilized throughout the latter half of the 20th century; However, it has recently been the focus of attention. This is attributed to the copious amounts of data available to train systems, the accessibility to computing systems that can process this amount of data, etc. Moreover, depending on the training size of the data, distributive semantics is more robust compared to systems that rely on annotated databases, such as FrameNet and Propbank. This is because there is no dependency on external systems and because languages are constantly changing (Hickey, 2003), distributive semantics can identify those changes.

2.1.3 Set Theory and Predicate Logic

Finally, the last major grouping of semantics parsers employs the use of mathematical set theory (E.G. predicate, first, second, higher order logic). Set theory based parsers primarily focuses on converting natural language text into mathematical formulas that are represented with quantifiers, negations, variables, constants, and functions. One example is listed below with a sentence and its logical form.

Every girl smiles.

$$\forall x.girl(x)smile(x)$$

The above example (Francez, 2014) shows a representation of a natural language text being converted into a mathematical notation. These systems have been traditionally used for its ease of integration into a programmable format

(Van Emden & Kowalski, 1976), more specifically, in programming languages such as Prolog (Blackburn & Bos, 2005) and Lisp. Another tool that predicate logic parsers utilize are lambda calculus functions. This is a further implementation of predicate logic that uses the outputs from other functions within a system as an input into other functions.

Furthermore, lambda calculus has the ability to identify predicates and arguments that fit the predicates, which is very similar to frame semantics except that there is no annotated lexical database that can determine the correct roles associated for each predicate. With the combination of variables used within this system, ease of integration into programming languages and use of input and output functions, it is an advantageous system to parse natural language text.

2.2 Tools

In this section, a summary of the parsers that were used in this study are described in detail. Below is a graphic that summarizes the parsers categorized into families. Within each main family, the name of the specific tool is listed. With the exception of the Stanford CoreNLP tool, the parsers between the families are intersections of the tools where these are not available to the public. At first glance, the list of parsers may seem arbitrary but these parsers were chosen on the merit of being highly cited or referred as "golden standards". This loose restriction may result in parsers missing from this study.

2.2.1 SEMAFOR

SEMantic Analyzer of Frame Representations, or commonly referred to as SEMAFOR, is a semantic parser that was developed on the basis of Fillmores frame semantics (Dipanjan, Schneider, Desai, & Smith, 2010). More specifically, it harnesses FrameNet as its underlying database reference for obtaining annotated frames. The system uses a *pipeline* design to streamline the process starting from raw text to semantic frame parsing. SEMAFOR's pipeline components is comprised of three major units to achieve semantic parsing.

Prior to the first component, a prepossessing step is included to tokenize and perform part-of-speech tagging on raw sentences to formula a list of lexical units. The first major component is to determine which lexical units in the sentence have the ability to evoke a frame. While it may seem intuitive to identify, lexical units can span across multiple tokens. The result of this process produces a list of possible candidate targets, which are then narrowed down further by applying a set of rules (Dipanjan et al., 2010)(Johansson & Nugues, 2008).

The second major component uses the target list compiled in the previous component to find frames from FrameNet. It employs the use of machine learning algorithms, WordNet and custom defined rules to identify the correct frame associated with the targets.

The final major component of this system is argument identification, which is to identify the arguments of each frame chosen in the second component. Similar to the second component, this section uses machine learning algorithms and a set of custom defined rules to restrict and label non-frame tokens as frame elements. Figure 2.1 is an example output from the system.

2.2.2 SHALMANESER

In addition to SEMAFOR, another state of the art parser that utilizes both frame semantics and FrameNet as its annotated frame source is SHALMANESER (Erk & Pado, 2006). The SHALMANESER system is part of a suite of NLP tools from the SALSA II project. The SHALMANESER system performs the semantic parsing and is split into three components.

The first component - dubbed FRPREP - performs the prepossessing screen of raw text. FRPREP uses a variety of open source and available software to

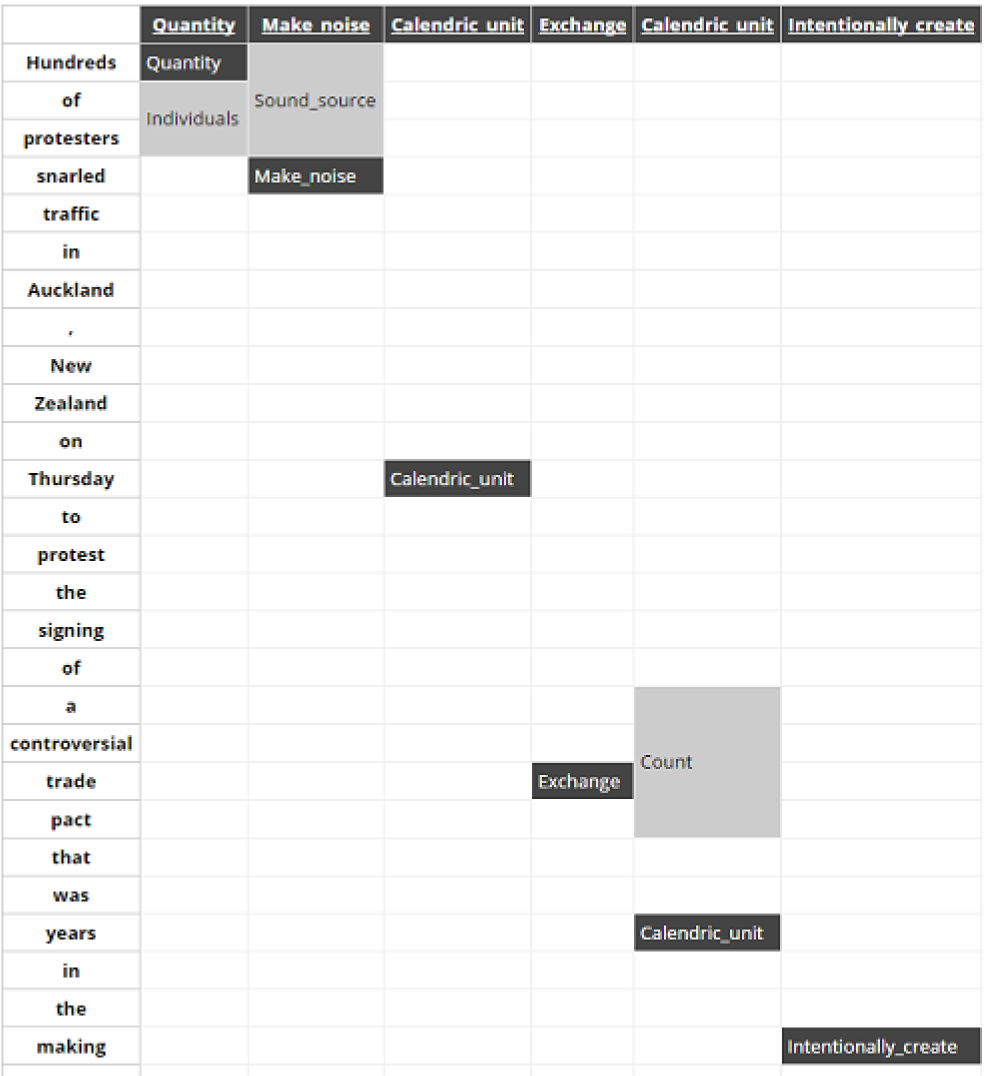

Figure 2.1.: Example of SEMAFOR output

perform tokenization and part-of-speech tagger. Some of the third party tools used are: the COLLINS Parser, Mallet, Minipar, and TNT.

The second component - dubbed FRED - applies commonly used NLP techniques, such as: bag-of-words, n-grams, and Naive Bayes to identify all available frames in a sentence.

The third component - dubbed ROSE - provides frame elements for the chosen frames from FRED. It uses each identified frame as a starting (root) node and employs a supervised machine learning technique, with 30 features predefined from the CoNLL shared task (Carreras & Marquez, 2005), to find the frame elements associated with the frame. The last output of the SHALMANESER system produces a custom file format (SALSA/TIGER XML). While it is possible to read the file output and decipher information from it manually, it is more feasible to use the SALTO tool (a part of the same NLP suite of tools) to generate intuitive graphical images (Burchardt et al., 2006). Figure 2.2 is an example output from the system.

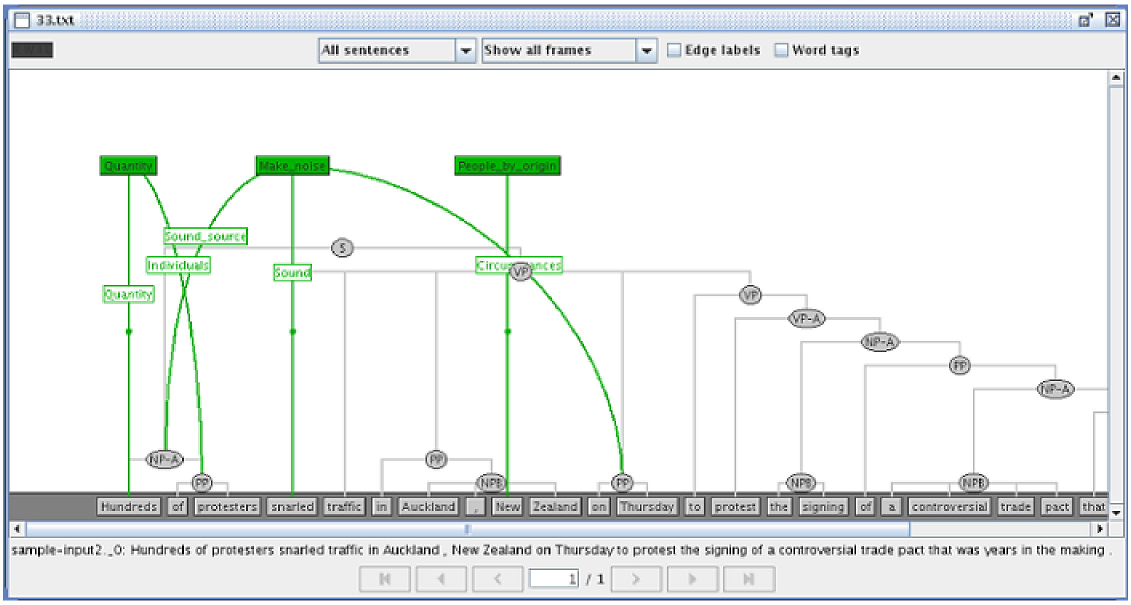

Figure 2.2.: SHALMANESER parser output

2.2.3 Punyakanok, Roth and Yihs approach

Punyakanok, Roth, and Yih present another semantic parser that also utilizes semantic frames but, instead of using annotated data from the FrameNet project, their approach applies the Propbank database for its annotated data of frames (Punyakanok, Roth, & Yih, 2008). This system is broken into two major components.

The first component preprocesses the data by parsing the sentence and locating the verbs (predicates). Once the verbs have been identified, the arguments for each verb are identified. Due to the nature of Propbank, predicates can have arguments ranging from A0-A5, AA and special argument types predefined in Propbank, the system is to identify correct arguments for each predicate by a reduction method. In other words, all non-predicate words are candidates to become arguments but are reduced further and further until a shorter list is compiled of candidate arguments. It reduces the number of arguments by using supervised machine learning methods with a variety of features specifically to minimize candidate arguments.

Since arguments within Propbank are not all equal (A0 in one predicate is not the same as another A0 in a different predicate) and the order of the labeling type can be different depending on the predicate, the second component proceeds with the candidate list of arguments for each predicate to label the types of arguments. Similar to the first component, this system uses a variety of supervised machine learning approaches to determine the type for each predicate's argument. The features used for both the first and second components machine learning approaches is detailed in (Punyakanok et al., 2008). Figure 2.3 is an example output from the system.

2.2.4 Johansson and Nugues approach

Another approach that utilizes frame semantics as well as the Propbank annotated database is a parser developed by Johansson and Nugues (Johansson & Nugues, 2008). Similar to the parsers described above, this system also contains multiple components in a pipeline fashion to process and parse semantic information from text.

The first component breaks a given sentence down by identifying all possible verbs (predicates) and arguments. Therefore, the resulting output from the first

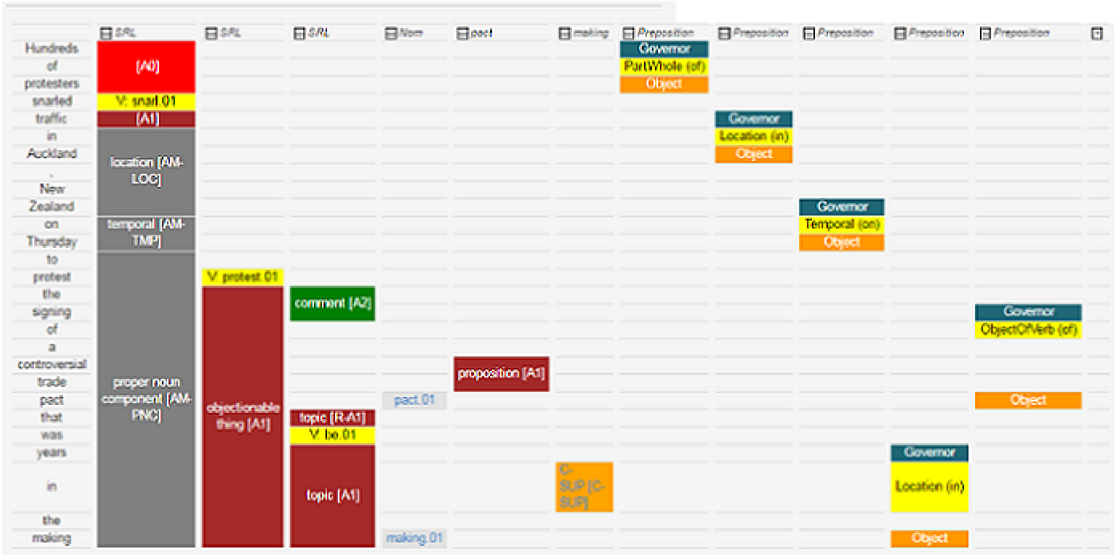

Figure 2.3.: Punyakanok, Roth and Yihs SLR parser output

component is a list of these predicate-argument structures. The second component then reduces the list by removing each item if it does not satisfy a linguistic rule. The three component applies a ranking score for each of the remaining items on the list and applies the highest ranking predicate-argument structure.

While the benefits of frame semantics has been well received by the industry, there are other sets of methods that are becoming ever more increasingly popular, specifically distributive semantics. This is due to copious amounts of data that is generated from natural language text and also technological advances that have giving computing systems to process this massive amount of data. Figure 2.4 is an example output from the system.

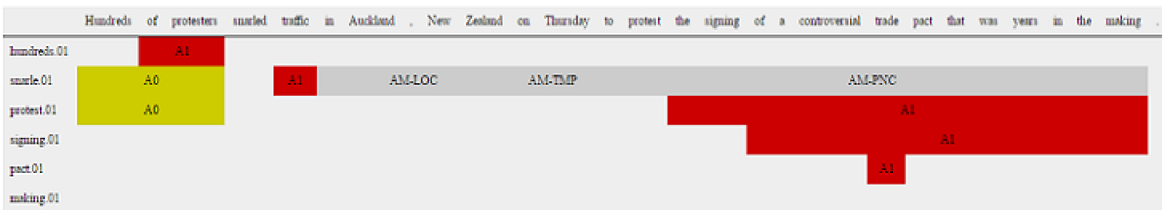

Figure 2.4.: Johansson and Nugues parser output

2.2.5 Hermann, Das, Weston and Ganchevs approach

An example of a semantic parser that utilizes both frame semantics and distributional semantics is outlined in this paper (Hermann, Das, Weston, & Ganchev, 2014). This approach uses a combination of both frame and distributive semantics to parse for semantics. The system uses the part-of-speech tags of a given sentence to find common, reoccurring syntactical patterns over many sentences for each available predicate. This identifies which predicates are most likely to appear in an unknown sentence. Unfortunately, access to this system is not available to the public and therefore, cannot be included in this tool comparison.

2.2.6 Word space models

Although, the intersection between frame and distributive semantics is not available for public use, there are also other approaches within this industry that use pure distributional semantics to obtain the meaning of a given text. The first popular approach within distributional semantics are wordspace models (Schtze, 1993). The model described in the paper focuses on context-group discrimination, which groups occurrences of an ambiguous word into clusters and those clusters have information regarding what words, contexts and clusters are represented. There are two popular approaches that utilize wordspace models: Hyperspace Analogue to Language (HAL) (Lund & Kevin, 1997) and Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997).

This approach is useful by identifying words and partially disambiguating word sense with little to no supervised training data. However, the researcher claims that word sense disambiguation can be divided into two separate tasks, specifically sense discrimination and sense labeling. Sense discrimination is the process of identifying clusters whereas labeling is mapping a cluster to a specific sense of the word. An approach that utilizes distributional semantics is the S-Space framework (Jurgens & Stevens, 2010). This packages framework contains wordspace models.

While using distributional semantics is a powerful tool, it has its limitations. These limitations are particularly based on the the dataset that is used to train the system (I.E. - the size of the training data, the type of training data, etc.). In other words, the more text that a distributive system has availability to, the more accurate the results will be produced, in theory. In the case where a reduced amount of data is available for a system to use, other more robust methods are introduced, which introduces the concept of first order logic based tools.

2.2.7 Beltagy, Erk and Mooneys approach

An approach that utilizes a combination of both distributional semantics and first order logic was developed by Beltagym and others (Beltagy, Erk, & Mooney, 2014). There are three components to this system, the logical conversion, the weight identification and the construction of an ontology. For the logical conversion, this is simply converting the input text into a logical form. The next step is to apply weights to each relationship. These weights describe how strong a relationship is between two words based on factors, such as: synonyms, antonyms, hyponyms, etc. The final component is to obtain the most optimal combination of words that have the highest valued weight pairs to declare as the meaning representation of the sentence. However, this system is not available for public download and thus, this parser will not be compared. While this system is not used in this study, it gives a segue into the final family parser, set theory (or predicate logic).

2.2.8 Boxer DRT

Boxer is a semantic parser that relies on Combinatory Categorial Grammars (CCG) and Discourse Representation Theory (Bos, 2008). CCG‘s are grammars that are predefined valid rules that the system must abide. CCG‘s also have a combination of both syntax and semantics (as logical representations) in the predefined rules. The figure below is a representation of CCG. Discourse

Representation Theory contains a structure that is compatible with first order logic equations and is comprised of objects being represented as variables and functions that utilize variables. In addition, the figure 2.5 shows a presentation of the Boxer semantic parser output.

2.2.9 Stanford CoreNLP tools suite

The last system that will be analyzed is the Stanford CoreNLP toolkit (Manning et al., 2014). This system is comprised of separate functioning modules that interact with each other to process natural language text, not only semantics. Moreover, there is no direct component that process text for semantic information, instead, this task is broken between different modules within this system, namely the named entity recognizer (Finkel, Grenager, & Manning, 2005) and Open Information Extractor (Angeli, Premkumar, & Manning, 2015). The Stanford NER tagger uses the original 7 MUC NER tags (Grishman & Sundheim, 1996a) when disambiguating words and their tags. The Stanford OpenIE system identifies the core elements of a sentence by identifying the predicate and its two arguments (because most sentences dwindle down to those three major components). In other words, the sentence is chucked into three parts where one part is the predicate and each argument is further reduced into the most atomic essence of the sentence. Figure 2.6 is a screenshot of the Stanford OpenIE tool.

2.3 Metrics

Now that the semantic parsers have been identified, the next important part in this process is to identify the metrics that will grade how well these parsers perform. Below are the metrics that have been defined for this study and listed below.

- Identification of events

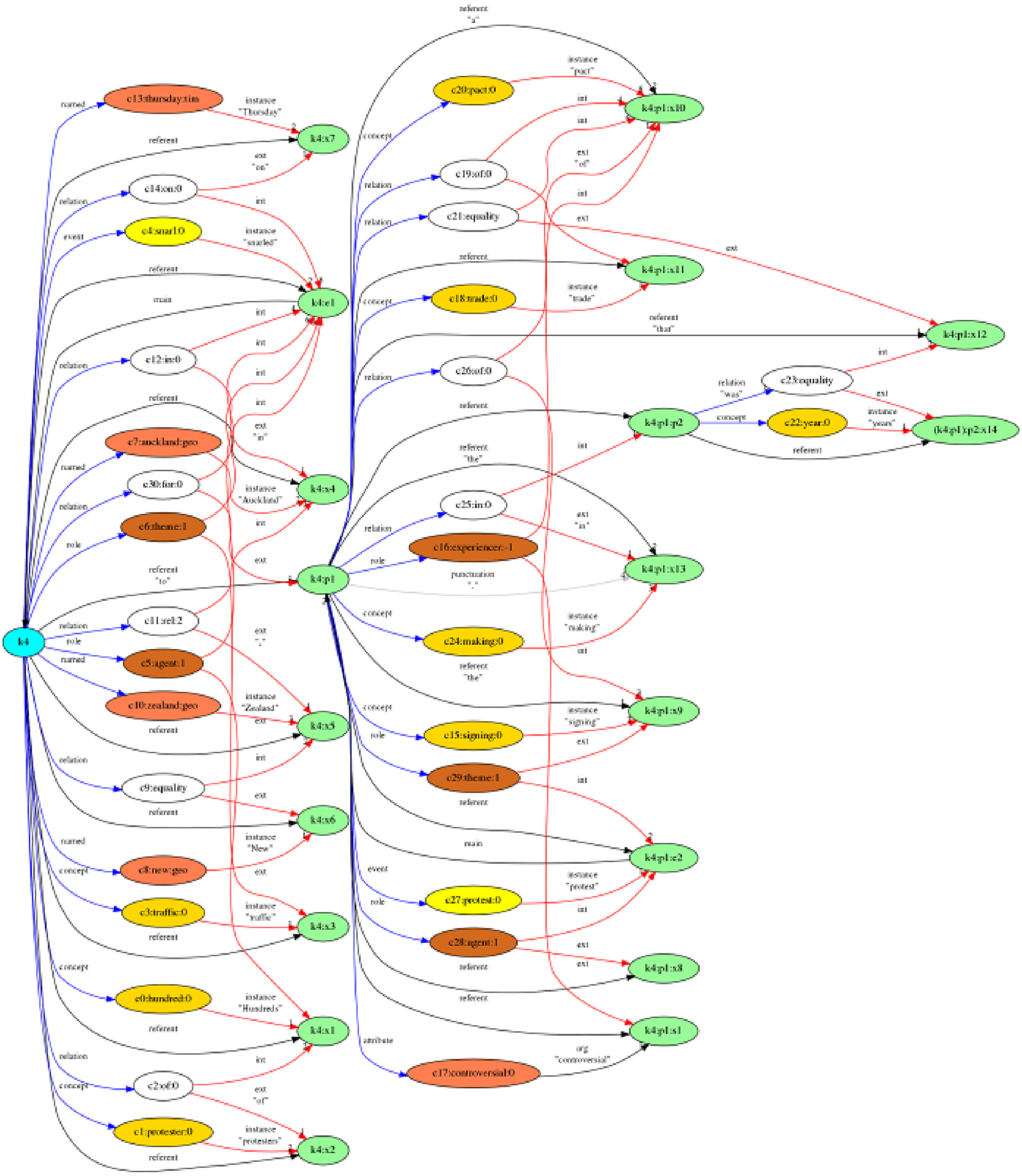

Figure 2.5.: Boxer parser output

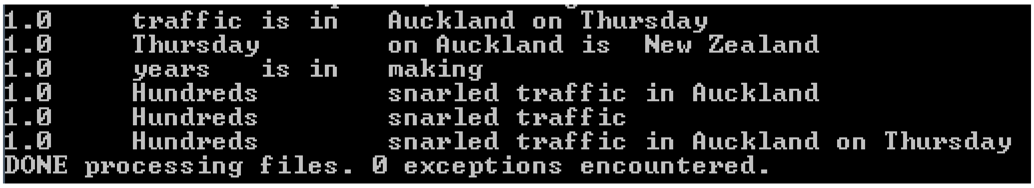

Figure 2.6.: Stanford OpenIE output

- Named entity identification
- Identification of objects
- Predication-entity relationship identification
  - Accuracy of arguments for each predicate
  - Number of arguments for each predicate
- Word Sense Disambiguation
  - Identifying the correct sense of predicates
  - Number of senses

2.3.1 Identification of Events

Events within a sentence describe *what* the interactions are between objects. It is a very common issue in NLP (Allan, Papka, & Lavrenko, 1998) (Yang, Pierce, & Carbonell, 1998) and has been a focus of a SemEval task (Verhagen et al., 2007). Many past studies have looked at media feeds as its dataset to identify events (Doddington et al., 2004) (Ritter, Etzioni, Clark, et al., 2012). The purpose of the first metric is to identify key words within the sentence that trigger an event. Events are words that describe a sequence(s) of a interactions between it's objects.

Objects (as described later) in a sentence illustrate who, what, where, etc. is involved. Without the event identified, it becomes unclear about the interaction between objects within a sentence (at this point, the sentence is a list of objects). Furthermore, the identification of events provides tremendous information about the the tone and feelings for the reader. As a simple example, if two sentences had identical words except for the event where the events are replaced with help and kill, both sentences illustrate different effects on the same objects.Therefore, it is clear that events are important and generally expressed as the ”glue”, which illustrates the connection between these very objects. Events appear in most sentences and represent an important cornerstone for the sentence. As a metric, events are words that are identified as the an action. While syntactical information may be useful (and used in many semantic parsers) in determining key actions, events can also be identified as other lexical categories.

2.3.2 Named Entity Identification

The second metric is focused on the task of Named Entity Identification (NER) (Chinchor, Brown, Ferro, & Robinson, 1999). Named entities are the most atom elements of a sentence which can be represented with words or phrases. These words or phrases can also be classified further with tags and some of the original NER tags are: named entities (person, location, organization), time and measurement units (Grishman & Sundheim, 1996b). While this may seem like a trivial task, this is a complex problem with the industry making several attempts to promoting for a better solution (Tjong Kim Sang, 2002)(Tjong Kim Sang & De Meulder, 2003)(Nadeau & Sekine, 2007). While the classic NER tags are generally used in many information extraction systems, semantic parsers, etc. These NER tags are limited in identifying words that match these tags. For example, deciphering word sense disambiguation and identifying other artifacts. That is, if the word, *Obama* appears in a sentence, should *Obama* be classified as a person (President of the U.S.) or a location (a city in Japan). As for artifacts, if the phrase, *World War II* appears in a sentence, World War II is not a time, unit measurement or named entity. Thus, additional NER tags have been proposed to support more phrases that span across more domains (Sekine, Sudo, & Nobata, 2002) by creating more specific subsets of categories to accommodate other artifacts and is used in this study.

2.3.3 Identification of Objects

The third metric is the identification of objects within a sentence. These can also be referred to arguments for predicates, which can be used in conjunction with objects. This provides detail to questions about what, whom, where, etc. of the sentence. While named entities have similar functionalists as objects, in terms of being used as arguments for predicates, objects are not specific but are needed within relationships to provide more understanding about sentences.

2.3.4 Predicate-entity Relationship Identification

The identification of events only provides information about the actions of objects within the sentence. However, as a realistic possibility, events alone do not specify *which* of the objects are associated to sentences with multiple events. Therefore, event phrases with their objects (also known as predicates), provides deeper information about a sentence and examines *how* the objects are related to the predicate.

The fourth determines how many objects the parser identifies for each predicate that matches the objects identified manually. In other words, if two objects within a sentence are arguments for one predicate and the parser identifies two different objects for the same manually-identified predicate, the parser scores a 1 for matching one object. Arguments can also be shared between predicates. In the scenarios where parser-identified arguments contain the ”correct” arguments but with more words than the manually classified, punctuation, articles, and words that do not intersect between other arguments in both the manually and parser classified are treated correctly. As a simple example below, the top and bottom relationship represents a manually classified relationship and a parser classified relation, respectfully. In this case, the parser would receive a score of 2.

Crash(ship, island)

Crash(the ship, the large island)

The fifth metric sums the number of arguments identified by the parser for the manually identified predicates. This provides a quick comparison between manual and parser identifying arguments.

The fourth and fifth metrics analyze the predicates. It has particularly evolved as an important task in this industry (Gildea & Jurafsky, 2002)(Matsubayashi, Okazaki, & Tsujii, 2014) and has been commonly integrated with open information extraction.

2.3.5 Word Sense Disambiguation count

Word Sense Disambiguation (WSD) is an important (Laorden, Santos, Sanz, Alvarez, & Bringas, 2012) and fundamental problem within this field. Languages, such as English and many others, contain multiple definitions and meaning tied to shared lemmas and in most cases, can cause confusion within text and even dialogs (if the lemmas also share the same phonetics between different definitions). Therefore, WSD is the generic problem to decipher the correct definition when a word contains multiple definitions (Ide & Véronis, 1998). The process of WSD can be split into different two stages.

The first stage and as the sixth metric is identifying that a specific word has the correct sense associated for a specific predicate. In the scenario that "meet" is manually identified as a verb but a parser identified it as a noun, then the parser has no scoring for that predicate. The second stage and the seventh metric is counting the number of senses available for each predicate.

To help verify and strengthen the quality of manual annotations, Wordnet will be used as a reference to count the total number of senses a verb may contain. Wordnet (Miller, 1995) is a database of synonyms that can differentiate lemmas into different senses.

2.3.6 Additional metrics

Throughout the literature review, there are other types of metrics that were used to grade the success of the system not described above. For example, systems such as SEMAFOR and Shalmaneser (and others) use precision, recall and F scores to show the effectiveness of their systems (Dipanjan et al., 2010)(Erk & Pado, 2006). While these metrics provide a powerful insight on the accuracy performance of these systems, it is difficult to apply the same metric across multiple systems as some are more dependent on resources, such as training data, lexical datasets and more. Therefore, the metrics above have the ability to be applied the same across multiple systems.

2.4 Summary

In this chapter, a extensive description of the semantic parsing tools that will be used are described. This provides the necessary background knowledge needed to understand the mechanics of each tool and why each tool performs differently compared with other semantic parsers. The figure below shows a summarized view of each semantic parsing tool used separated into groups, which contain unique characteristics. In the next chapter, the experiment design will be described in detail.

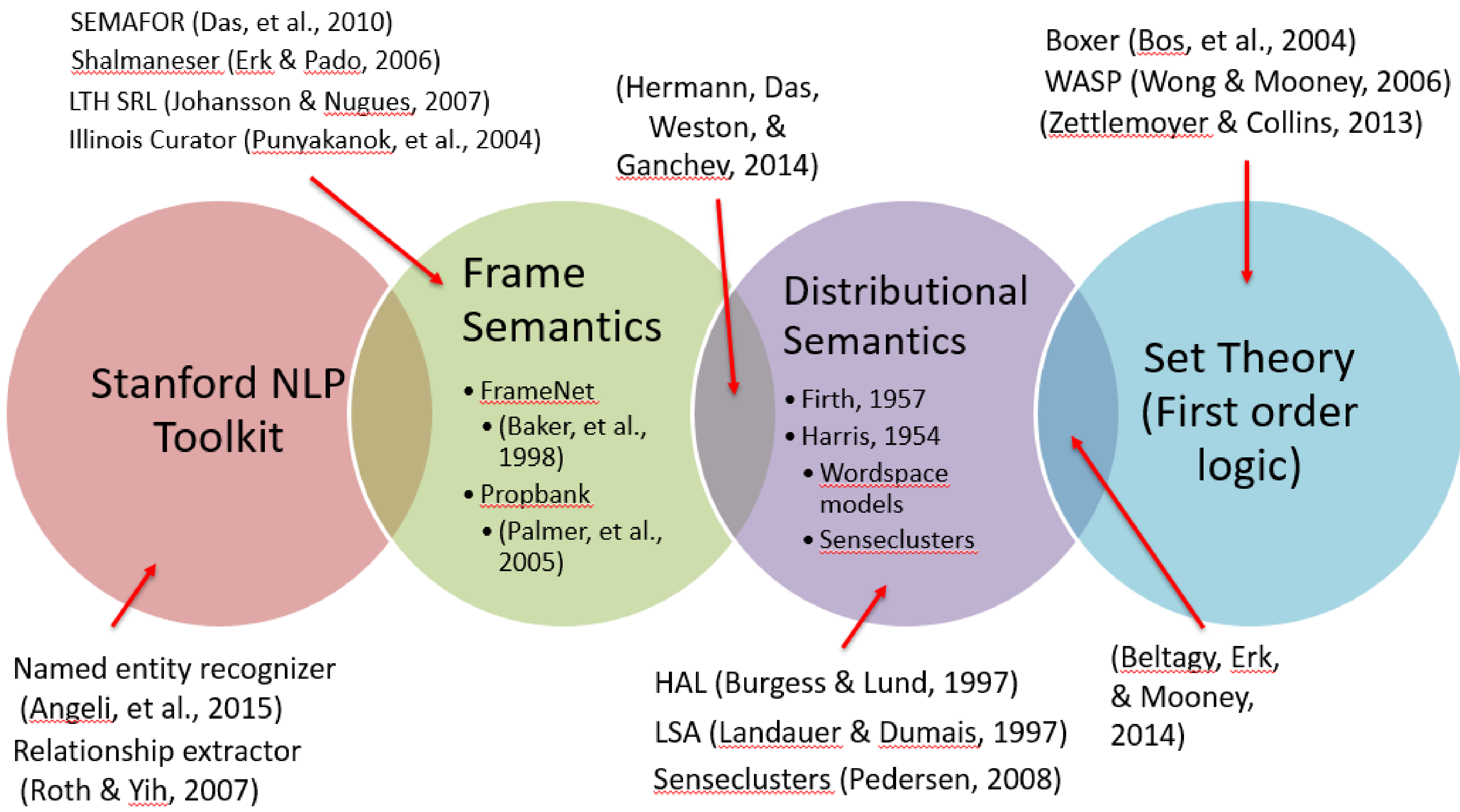

Figure 2.7.: Summarized diagram of semantic parsing tools

CHAPTER 3. FRAMEWORK AND METHODOLOGY

In this section, an extensive outline of the dataset, experiment design and results are described. As an introduction, the initial step was to obtain a dataset and manually annotate each sentence by identifying what the researcher believes is correct. Next, the same annotated sentences will be used as inputs into the semantic parsers that are under investigation. Third, each resulting output from each parser was analyzed to determine if the parser at hand identifies the same metric as the manually annotated sentences. Finally, the results are organized in a tablular view for easy digestion.

3.1 Participating Semantic Parsing tools

The parsing tools used in this experiment are listed in the figure 3.1 below.

3.2 Tool configuration settings

Some participating semantic parsing tools used this study can function out-of-the-box with minimal configuration. While the installation of other semantic parsers are more involved and require additional files (such as models, training data, etc.) to operate. Therefore, to reduce the amount of discrepancy on how each parser was trained, all parsers used default settings and/or configurations that were recommended by the authors of each system. As an example, participating distributive parsers require corpora in order to create and train its models. In this specific case, to minimize the discrepancy between training training data for each parser and the given the fact that frame semantic tools are based on annotated

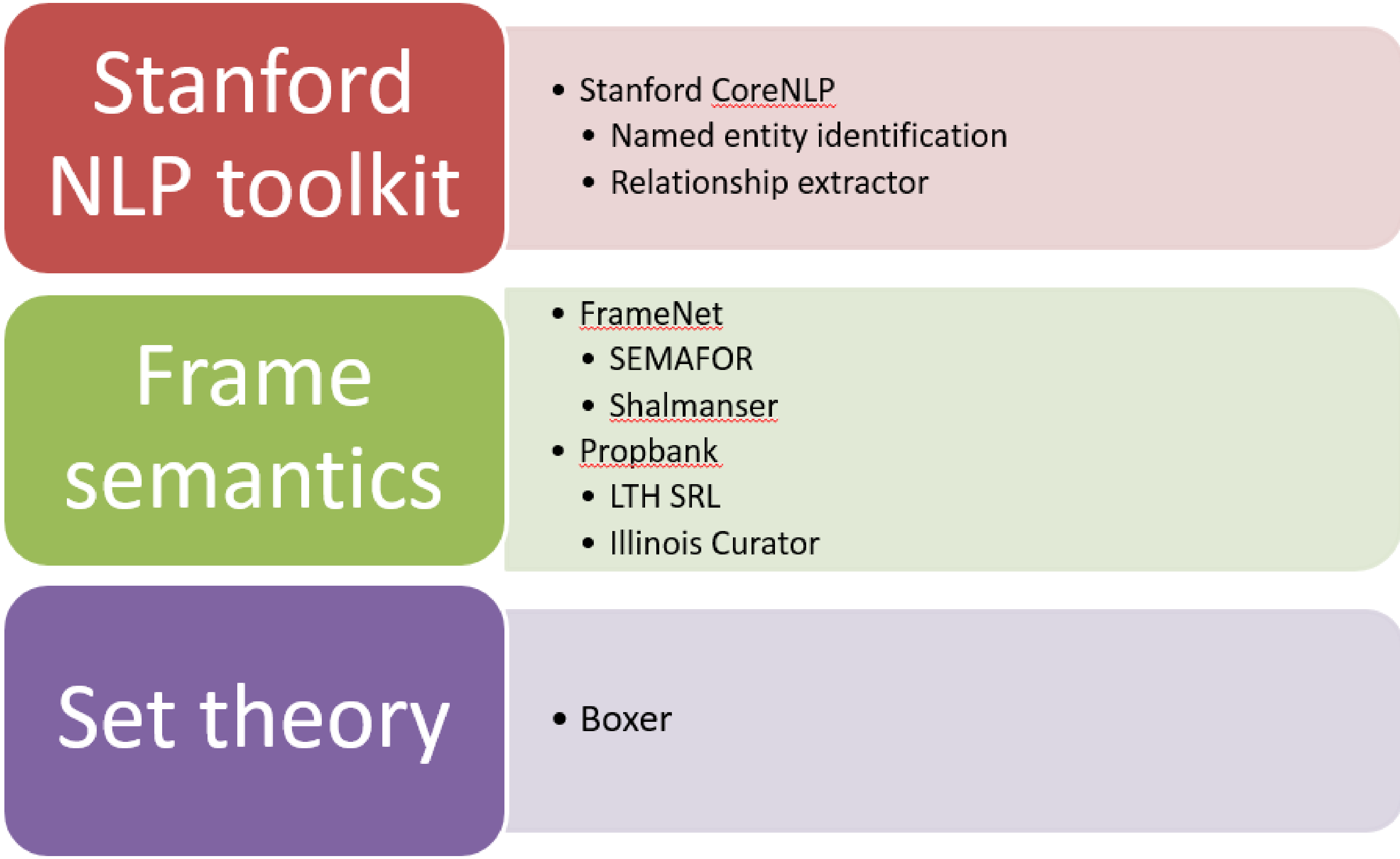

Figure 3.1.: Parsers used in this study

sources (such as FrameNet and Propbank), the corpora was derived from the same combined annotated sources.

3.3 Dataset

In this section, the selection process of the dataset that will be used in this study is described.

3.3.1 Data Source

The dataset for study contains 95 randomly chosen natural language sentences in English from the various news sites. The rationale for choosing 95 sentences is to show statistically significant results and to validate the overall performance of the participating parsers. The reasoning for choosing sentences from various news sites is because these sentences were written in structured, formal English text, which is the primary focus of testing the parsing systems described in Chapter 2. Some of these news sites include text snippets from the New York Times, CNN, USA Today, and more. Furthermore, these news sites contain snippets of how English is currently being used and news sites contain text in an open domain setting.

3.3.2 Data Selection Criteria

Since the data source of the sentences will be extracted from news sites, a list of restrictions are imposed on both the articles and individual sentences. Since the default components in each of parser were used, the purpose of diversifying articles and news sources was to validate how robust each parser performs on data that may not have been used as apart of the parser's training (if any) or testing.

Restrictions on articles

- Written in formal English - Some parsers have difficulties in parsing text in non-English characters.
- Taken from a recent news articles - Provides sentences that these parsers have not been trained or tested on
- Contains a mixture of different news sources - Provides different writing styles, which validates how robust each parser performs

Restrictions on sentences

- No quotations can appear
- Sentences must contain named entities - To identify and classify a NER tag for each named entity
- Sentences must have predicate words with multiple word senses - To disambiguate word senses for each predicate identified
- Sentences must have multiple events - To decipher which arguments are associated for each predicate

3.4 Execution

Of the 95 randomly selected sentences, these will be used as an input into the six parsers that were chosen based on their availability for public use. Once every sentence has run through the six parsers, the results will be analyzed for each sentence. The goal of this execution plan was to determine if the semantic metrics described above appears within the results of the parsing systems. Figure 3.2 shows an outline of the execution process for this experiment.

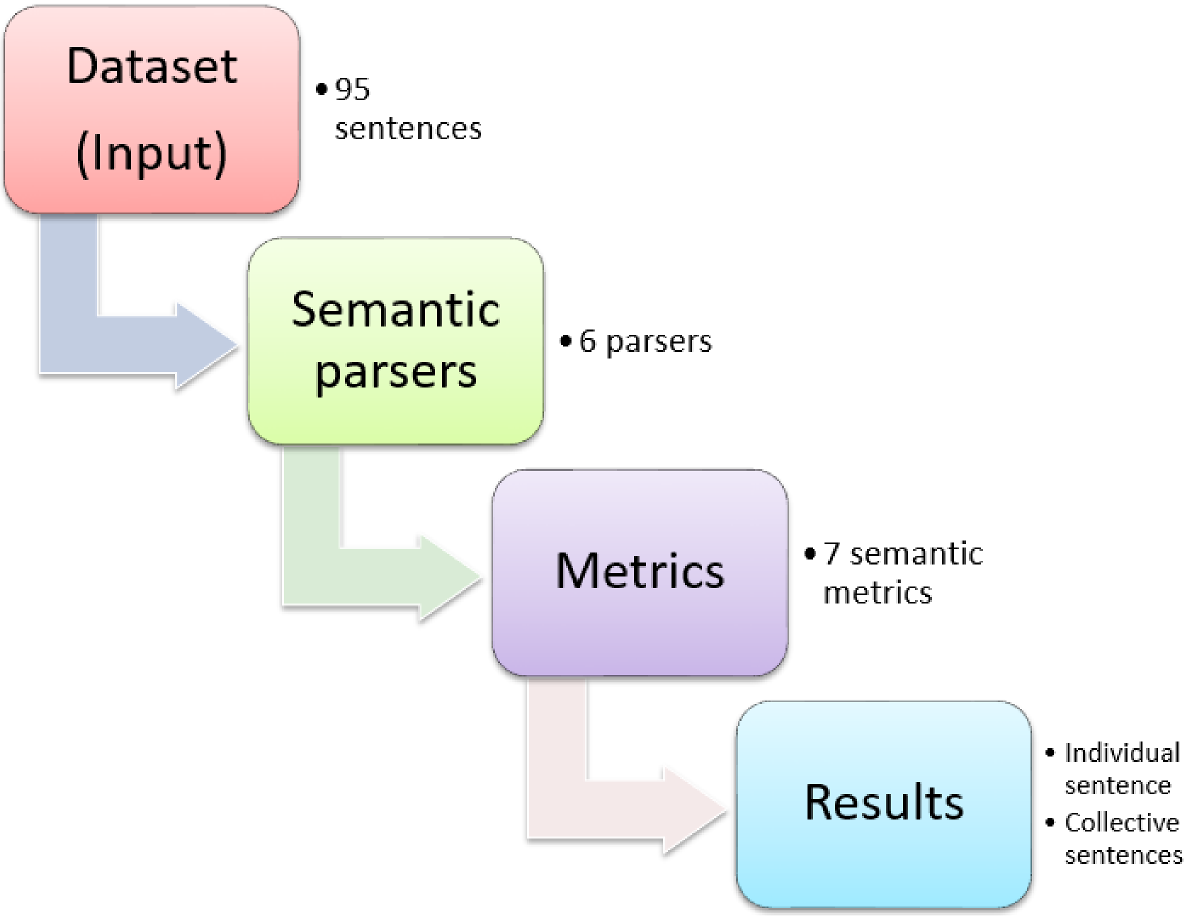

Figure 3.2.: A framework process

Table 3.1: Titles for the first experiment

Symbol	Titles
EI	Event Identification
NEI	Named Entity Identification
OI	Object Identification
PR1	Predicate Relationship - Identification
PR2	Predicate Relationship - Argument Count
WSD1	WSD - Identification
WSD2	WSD - Sense Count

3.4.1 Experiments

There are two experiment that were used to grade the accuracy and precision of these parsers. The first experiment focused on a single sentence. For each sentence, the resulting output is compared to its annotated version to determine if a metric has been correctly identified. This process will be repeated for each parser until all sentences has been evaluated on each parser. In order to streamline the efficiency and to provide fairness between all parsers, the evaluation of the comparison was handled strictly.

Each sentence has been manually annotated and compared with a resulting output from a parser, an example output is displayed below to illustrate how the output will appear in table 3.2. Due to spacing constraints, titles used in table 3.1 have been encoded with the mappings listed out in table 3.1.

The second experiment is similar to the first, the only difference is that in the first experiment, only single sentences were analyzed each time and in the second, the results from the experiment are collectively summed and averaged over the entire dataset used in this study. Additionally, contrary to the first experiment, where individual sentences were viewed in detail, the second experiment provides a performance view for each participating parser. Each table in the second

Table 3.2: First experiment example for individual sentences

	EI	NEI	OI	PR1	PR2	WSD1	WSD2
Parser 1	1	0	2	Pred1(1, 2)	Pred1(4)	Pred1(Y)	Pred1(3)
Parser 2	0	0	2				X
Parser 3	1	1	1	Pred1(1)	Pred1(2)	Pred1(N)	X
Parser n							
Manual annot.	1	0	2	Pred1(arg1, arg2)	2	NA	Pred1(3)

experiment contains the results for each parser with their respective ratio. The algorithm below represents the average for each parser. For an illustration, a sample result output for the second experiment is displayed (Table 3.4). Table 3.3 is a mapping key that lists out all title names in Table 3.4.

Table 3.3: Titles for the second experiment

Symbol	Title
EI	Events Identified
NEI	Named Entities Identified
OI	Objects Identified
PR1	Predicate Relationship - Total number of correct arguments
PR2	Predicate Relationship - Total number of arguments identified
WSD1	Word Sense Disambiguation - Total number of correctly labeled sense
WSD2	Word Sense Disambiguation - Total number of senses identified

for *each metric* **do**
    for *each parser* **do**
        Sum the total number of counts for a metric over all sentences (J)
    end for
    Sum the total number of manually annotated counts for a metric over all sentences (K)
    Calculate the ratio by averaging of each total metric (J/K)
end for

The second experiment algorithm

Table 3.4: Second experiment example for all sentences

	EI	NEI	OI	PR1	PR2	WSD1	WSD2
Parser 1 - Results	36/63%	77/72%	201/98%	74/74%	82/28%	201/70%	807/80%
Manual annot.	57	107	205	87	100	288	1005

3.5 Analysis Example

In this section, an example of a sentence is introduced. This sentence was taken from the dataset used in this study to help illustrate the thought process of manually annotating sentences and identifying metrics from parsers. This also provides a comparison to the results of the participating semantic parsers. The sentence below is a modified example (Màrquez, Carreras, Litkowski, & Stevenson, 2008):

”Hundreds of protesters snarled traffic in Auckland, New Zealand on Thursday to protest the signing of a controversial trade pact that was years in the making.”

The first step was to manually annotate all sentences prior to inputting the sentences into the parsers. This is to reduce any temptations to modify the experiment as it may be biased to change the grading system which all systems are being compared. The manual annotation process started by identifying all events within the sentence. From the example above, the events are: snarled, protest, signing and making. For each parser's output, the same events must be identified in order for the parser to be considered correct for this metric. Partial credit may be given if some events match the manually identified.

"Hundreds of protesters *snarled* traffic in Auckland, New Zealand on Thursday to *protest* the *signing* of a controversial trade pact that was years in the *making*."

After the events have been identified, named entities and objects are identified. Continuing from the example, the named entities are: Auckland [LOCATION], New Zealand [LOCATION], Thursday [TIME-DAY OF WEEK]. As for the objects: protesters(Hundreds), traffic, trade pact(controversial), years. As mentioned in the literature review, the name entities must be identified and tagged correctly to the appropriate NER tag. In this study, the extended NER hierarchy was used. The extended NER hierarchy contains 150 NER tags while most traditional NER tags are: counts, time, persons, locations, organization, etc. This posed a problem if a named entity, such as World War II, is displayed. If a parser identifies a phrase as an original NER tag (for example), the extended hierarchy can still accommodating the original tags by transversing to more abstract NER tags. For each parser's output, the same words or phrases must be identified with the correct NER tag. If a parser identifies the correct NER phrase but with the wrong NER tag, it would be marked as incorrect in that instance.

The next step is the identification of objects, which are non-named entities that act as arguments within predicate phrases. Verbs, nouns and adjectives can take form of objects. In addition, objects can have word modifications, which are included with each identified object. In the example above, protesters, traffic, trade pact and years are the objects. Protesters has the modification of hundreds and trade pact has the modification of controversial. For each parser's output, the same objects must be identified with the same words or phrases (with the exceptions of articles and punctuation). There are instances where a single object can have multiple words (as identified with trade pact).

Once the events, named entities and objects have been identified, the next stage is to associate named entities and objects to their respectful predicate. Continuing the example, the predicates arguments for snarled are: protesters, traffic, Auckland, New Zealand and Thursday. The arguments for protest are: protesters and signing. The argument for the predicate signing is trade pact. The arguments for the predicate making are trade pact and years. Notice that from the list of identified objects, protesters has the modifier of hundreds. For each parser's output, the same arguments associated to each predicate must be identified with the same predicate phrases and its arguments. For each parser's output, it would receive a score equal to the number of arguments that it correctly identified for each predicate.

In the event of a parser identifying arguments of a predicate but containing more tokens with the arguments, then it would be mark incorrect, even though the argument may contain the correct argument in the phrase. For example, in the above sentence, if a parser declared that the phrase, signing of a controversial trade pact that, as a single argument for the predicate making, that metric would be marked as incorrect. Also, in the event that a parser identifies one argument but contains tokens involved in both manually classified arguments, it would be considered incorrect as this would be unfair to other parsers that differentiated between the two arguments. As an example, in the predicate, making, the two arguments are trade pact and years - if a parser identified one argument for making as, trade pact that was years, as one single argument, it would be considered incorrect.

Finally, the word senses are the final two metrics. The Word Sense Identification (WSD) metric determines if a parser can identify that multiple word senses are associated with a predicate. The WSD count metric sums the total number of word senses for each identified predicate the sentence contains and Wordnet online was used to count the number of word senses. Because the number of word senses are revealed in the WSD count section in the manual annotation, it is unnecessary to determine if a word has multiple senses for each predicate. For each parser's output, a simple Y for yes and N for no suffices to declare that a predicate contains multiple word senses. Figure 3.5 below shows a simple example of the manual classification organization.

3.6 Summary

In this chapter, a extensive description of the experiment has been described. This provides the necessary background knowledge needed to understand the results that will follow in the next chapter.

- Hundreds of protesters snarled traffic in Auckland, New Zealand on Thursday to protest the signing of a controversial trade pact that was years in the making.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
Manual annotation	Snarled, protest, signing, making	Auckland(location), New Zealand(location), Thursday(time-day_of_week)	Protesters(hundreds), traffic, trade pact(controversial), years	snarled(protesters, traffic, Auckland, New Zealand, Thursday), protest(Hundreds of protesters, signing), signing(trade pact), making(trade pact, years)	Snarled(5), protest(2), signing(1), making(2)	NA	Snarled(5), protest(6), signing(9), making(52)

Figure 3.3.: Example output

CHAPTER 4. RESULTS AND DISCUSSIONS

4.1 Results

The results from experiment 1 are listed in Appendix A of this thesis. Experiment 1 contains the specific details of the individual sentences from the manual annotation process and the deciphered results from each parser. As a summary view of the information, the figures below represent the performance of parsers at each individual sentence. To understand the tables from experiment 1, in the manual annotation row, the events, named entities (and the associative NER tags), objects, predicates (with their associative arguments), predicate counts (number of arguments identified manually for each predicate) and WSD count (for each predicate) are listed.

Figure 4.1 illustrates the performance of all parsers by the ratio of correctly identified events by sentence. The ratio is the number of events identified by the parser divided by the number of manually identified events. In addition, the values of each parser are sorted from highest to lowest for easier readability. The LTH SRL system identified the most events. Followed by SEMAFOR, Boxer DRT, the Illinois Curator, SHALMANESER and Stanford NLP, respectfully.

Similar to Figure 4.1, Figure 4.2 illustrates the performance of all parsers by the ratio of correctly named entity tags by sorting, from highest to lowest, by sentence. The Stanford NLP tool suite had a high ranking of NER tags, followed by Boxer DRT, SEMAFOR, LTH SRL. Both the Illinois Curator and SHALMANESER systems performed the lowest in this category.

Figure 4.3 shows the performance of all parsers by the ratio of correctly identified objects, from highest to lowest, by sentence. Both the SEMAFOR and

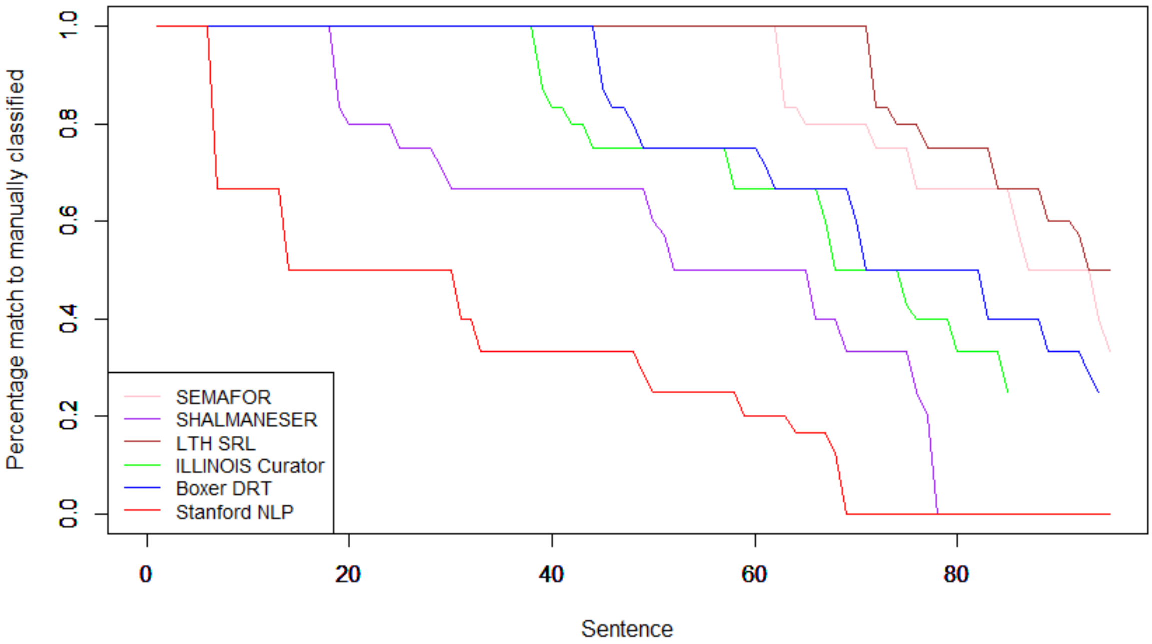

Figure 4.1.: Experiment 1 - Sorted ratios of event identification by parser

Illinois Curator systems obtained higher ratios over the dataset, followed by SHALMANESER, LTH SRL and Stanford in last.

Figure 4.4 shows the sorted ratios of all parsers for each sentence. The Box DRT system obtained a higher ratio, followed by the LTH SRL, the Illnois Curator, SEMAFOR, and both Stanford and SHALMANESER in last.

Figure 4.5 shows the total number of arguments used for each manually identified predicate, regardless if the arguments were correct.

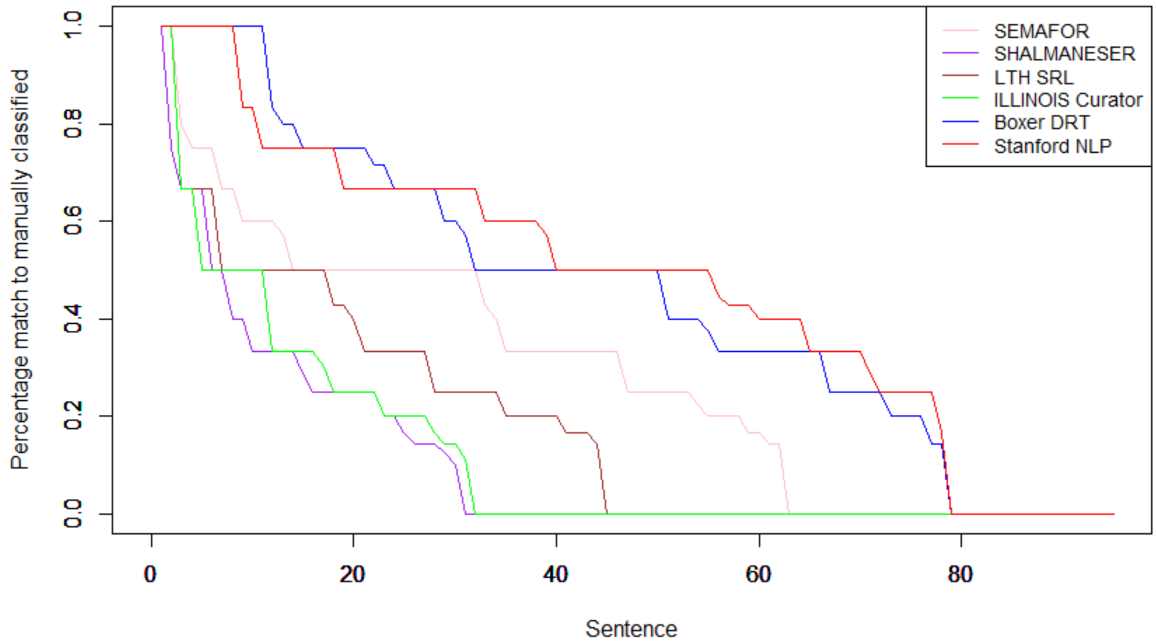

Figure 4.2.: Experiment 1 - Sorted ratios of named entities by parser

For each parser, the integer values in the Event Identification, Named Entity Recognition, and Object Identification represents the total count of correctly identified metrics according to the manual annotations. The Predicate Relationship Identification counts the number of correctly identified objects for each predicate (according to the manual annotation for each sentence). While the Predicate Relationship Count metric counts the number of arguments that each parser associates to each predicate - regardless if the number of arguments are correctly identified. The WSD identification is a simple binary response (Y for yes and N for no) to determine if the parser correctly identified the correct sense of non-named entities.

For the experiment 2, the summation of each parser and metric was summed. In addition, the percentage next to each integer represents the scaled ratio of correctly identified metrics. That is, some sentences were unable to be parsed and

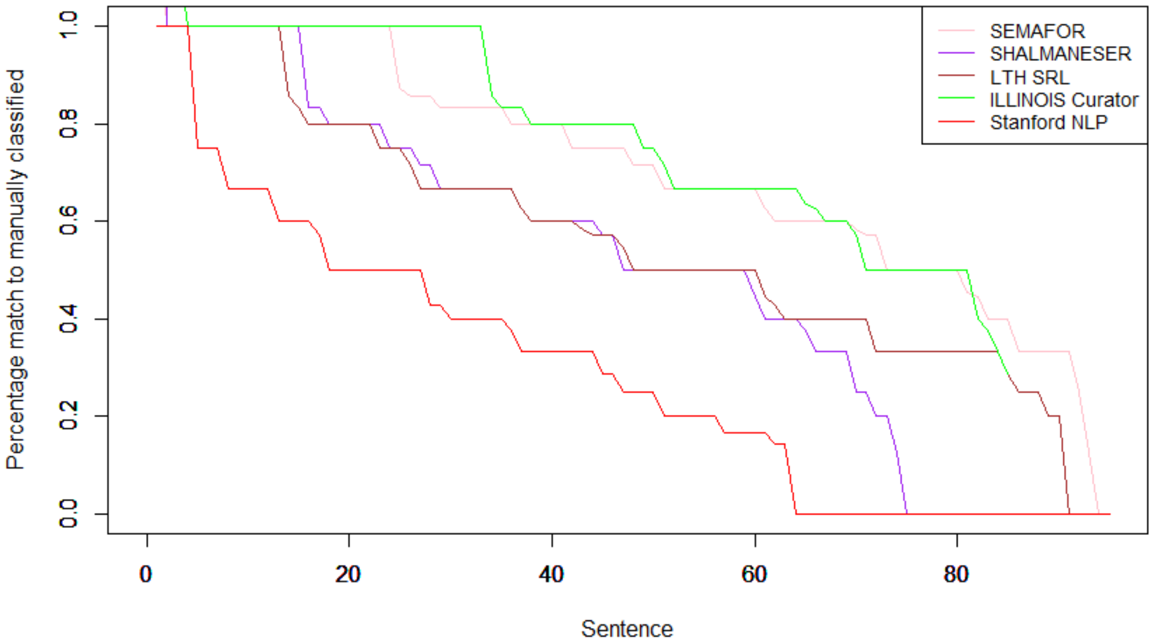

Figure 4.3.: Experiment 1 - Sorted ratios of objects identified by parser

were labeled as "Unavailable" and the total manual annotations for the parser was reduced. The results for experiment 2 is listed in Figure 4.6.

4.2 Discussion

4.2.1 Event Identification

In terms of ranking similarities with manually annotations, the LTH SRL and SEMAFOR systems performed the best in matching compared to the other systems investigated. This may be contributed these two systems sharing similar architectures. Both systems preprocess the data and follow the same data process in selecting the frame and its elements. The target selection in both systems prune its

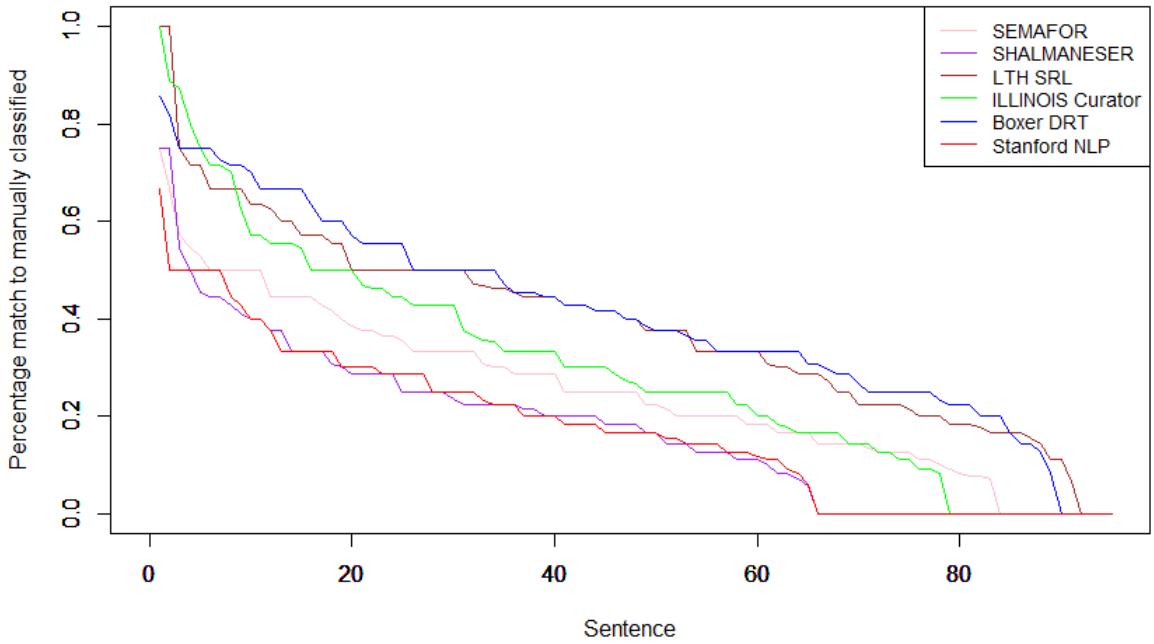

Figure 4.4.: Experiment 1 - Sorted ratios of predicate arguments by parser

target list by applying predefined rules, which are almost identical (which the exception of SEMAFOR containing a few more rules in the pruning stages). This is despite the fact that SEMAFOR uses FrameNet and LTH SRL using Propbank as their annotated frame resources. Some examples of the SEMAFOR system performing the identification of events better than other systems are Appendices A.11, A.16, A.25, A.30 and A.40. While prime examples of the LTH SRL system performing well are Appendix sentences: 9, 24, 30, 31 and 33.

Some of the lowest event identification systems are the SHALMANESER and Stanford systems. Surprisingly, as the SEMAFOR system performed was one of the best performance in the identification of events, the SHALMANESER system performed one of the lowest even though both system use FrameNet as its frame reference. Its performance may have been hindered due to having relatively older models for its supervised training set where SHALMANESER was using a

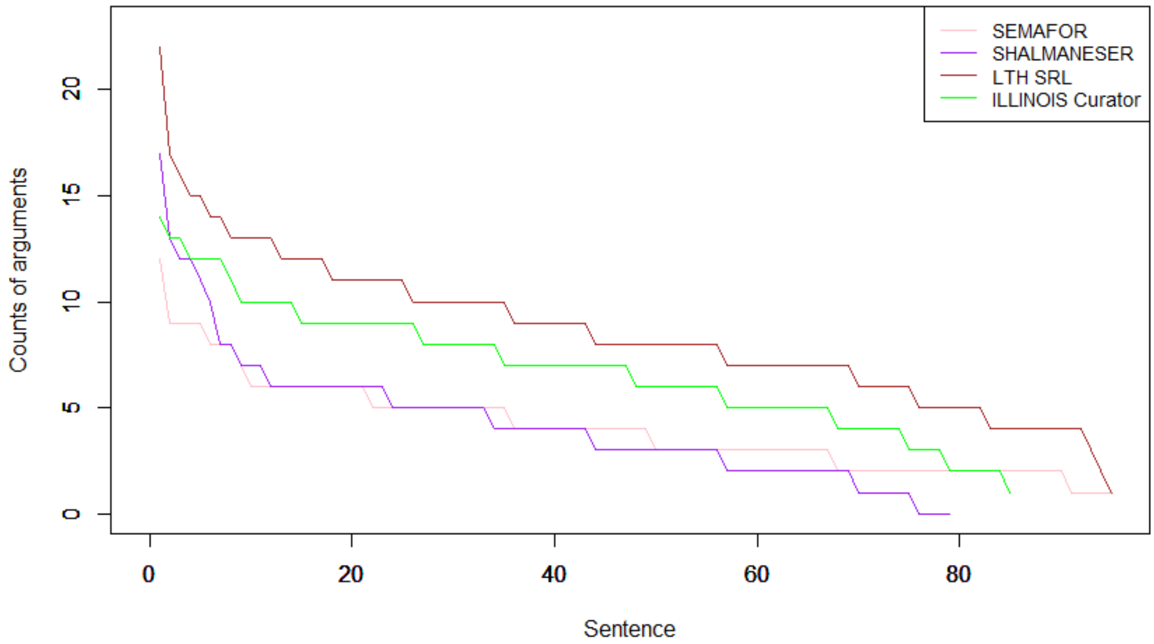

Figure 4.5.: Experiment 1 - Total number of arguments for each predicate by parser

FrameNet version prior to the latest version that SEMAFOR used. Its performance may also have been hindered by the use of older pre-trained classifiers (since default classifiers were supplied and used from the author's site). Some examples of the SHALMANESER tool performing is listed in the Appendix sentences: 36, 75, 66, 84 and 95.

The Stanford CoreNLP suite contains segmented tools for different NLP tasks. Therefore, multiple tools within this suite were used, which may have hindered the performance of this tool (note - only the Stanford NER tagger and OpenIE tools were used). Since the Stanford OpenIE tool only focuses the minimum number of predicate phrases in a given sentence and does not provide a entry for incomplete predicate phrases, this hinders the number of events identified as well as other metrics used in this comparison. Some examples where the Stanford OpenIE system performed poorly are on Appendix sentences: 41, 46, 48, 70 and 73.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	WSD - Ident.	WSD - Count
SEMAFOR	300/87%	109/31%	318/69%	203/24%	293/97%	NA
SHALMANESER	182/63%	41/14%	229/58%	156/22%	180/98%	NA
LTH SRL	311/90%	61/17%	253/55%	322/38%	311/100%	NA
Illinois Curator	238/74%	40/13%	310/73%	248/32%	238/100%	NA
Boxer (DRT)	253/73%	169/48%	NA	337/39%	NA	NA
Stanford NLP	89/25%	171/49%	124/27%	134/16%	NA	NA
Manual annotation	346	352	460	857	346	3210

Figure 4.6.: Experiment 2 - Results Summary Table

4.2.2 Named Entities

Despite the Stanford tool lacking in event identification in this comparison, the Stanford NER tagger was performed the best compared to other parsers in this study even though all participating parsers performed generally low with this metric. The task of named entities was generally difficult as parsers must identify the token and the correct NER tag. Some examples of the Stanford NER tagger performing well are Appendix sentences: 8, 22, 25, 80 and 84. Meanwhile, some of the lower name entity identification systems were the frame based parsers. This is attributed to the inherent nature of annotated frame databases where frames do not have NER tags. However, some frames, such as CALENDRIC UNIT, WHOLE ORIENTATION and others make it obvious that NER tags are with these frame based systems and are counted as being correct. Another contribution is to the strict analysis method used in this study where, with the exceptions of articles and punctuation, the frame elements for the correct predicate must match the manual annotated sentence. Some example of poor performance of frame based parsers are Appendix sentences: 52, 53, 64, 65, 84 and 93.

4.2.3 Object Identification

In terms of identifying objects, SEMAFOR and the Illinois Curator performed the best compared to other parsers. This may be attributed to core frame elements that help refine the relationship between words within a sentence. That is, the frame elements associated with each frame make it easy to identify objects for a predicate and in the sentence as a whole. Some examples of parsers performing well with object identification are Appendix sentences: 17, 21, 31, 42 and 44.

The Stanford Open IE parser performed the worse in this group. This is attributed to its inherent theory the developers designed and the strict analysis method used in this study. The Stanford Open IE system only provides two arguments for each predicate. Therefore, in the case where sentences have multiple events, it may be hard for the system to associate an argument to an event because arguments may overlap with other arguments in other predicates. Second, to maintain integrity of the analysis, only objects that are solely identified count (with the additions of punctuation and articles). Some examples of parsers performing poorly with object identification are Appendix sentences: 18, 27, 29, 40 and 44. As for the Boxer parser, the objects were not counted because all words in a sentence were split and were all declared as "objects", regardless if it was associated with a predicate or not.

4.2.4 Predicate Relationship Identification

The best performing system to identify the correct arguments for each predicate was the Boxer parser. Contrary to the object identification where all objects in the sentences are identified, correctly associating the objects and named entities to the correct predicate was the challenge for this parser. Thus, making it more effective compared to other parsers but only at a 337/857 correctness, which could be due to the strict analysis method used. Some examples where the predicate identification performed well with the Boxer parser are Appendix sentences: 16, 30, 49, 67 and 71. There were many instances during the experiments where frame based parsers would have the correct arguments identified to the appropriate predicate but with additional words, which contained words for another object. Therefore, these were marked as incorrect, which partially explained the decreased in the scoring for frame based parsers, especially.

The lowest performing parser for this metric was the Stanford OpenIE tool, which has a correlation to the number of objects identified from the previous metric. Again, this could be due to the nature of the system where it identifies the most basic core predicate phrases, which limits the system to identify two objects for each predicate. Some examples of the Stanford OpenIE system performing predicate relationship identification are Appendix sentences: 44, 83 and 94.

4.2.5 Word Sense Disambiguation

Word sense disambiguation identification includes the parser identifying if the correct sense for each predicate has been used. The tools in this study mostly used the correct senses of the identified predicates. However, there were instances where frame based parsers used the incorrect sense of the predicate. As an example, Appendix A.41 where hand is disambiguated as a body part rather than the movement of passing. Some other examples of parsers incorrectly identifying word senses are Appendix sentences: 20 and 22.

The Word Sense Disambiguation (WSD) count is unavailable for all of the participating parsers. Therefore, this metric was rendered uninsightful for this study. However, between all of the manually identified predicates, there were a total of 3210 word senses.

4.3 Similarities between Parsers

Despite the differences between parsers, many similarities are revealed from the same results. Figure 4.7 shows the similarities mean of identifying events between each parser and every other parser. For each sentence and between two parsers, the total number of events that appear in both parsers was divided over the total number of events to produce the mean. The mean is averaged over all 95 sentences, which is represented in Figure 4.7.

Surprisingly, the highest agreement in event identification was between the Illinois Curator and Boxer parsers. Compared to other parsers, the SHALMANESER system obtained the highest agreement with the SEMAFOR system, which is understood as both systems are based on FrameNet. However, the opposite is not true where the SEMAFOR system obtained the highest agreement with the LTH SRL system. Many of these relationships can be further depicted in Figure 4.1.

Stanford NLP	X	X	X	X	X	X
Boxer (DRT)	X	X	X	X	X	37.55%
Illinois Curator	X	X	X	X	92.89%	38.17%
LTH SRL	X	X	X	83.36%	83.41%	33.72%
SHALMANESER	X	X	64.29%	64.5%	63.24%	33.05%
SEMAFOR	X	73.65%	84.44%	75.7%	73.6%	32.28%
	SEMAFOR	SHALMANESER	LTH SRL	Illinois Curator	Boxer (DRT)	Stanford NLP

Figure 4.7.: Similarities between parsers - Event Identification

Some of the lowest agreements between parsers were systems paired with the Stanford NLP system. Again, this could be attributed the inherent nature of the system by identifying the minimum number of events within a sentence.

Alike Figure 4.7, Figure 4.8 was calculated from the same procedure. However, instead of calculating event similarities, named entities were determined instead. Some of the highest agreements of named entities are between frame based parsers. This is possibly attributed to the nature of frame based parsers not able to identify named entities and their tags explicitly.

The lowest agreement between parsers are system compared against the Stanford NLP parser. Recall that the Stanford NLP system scored among the highest in identifying the correct named entities. Perhaps the cause could be attributed to the high quantity of named entities identified compared to other systems.

Finally, Figure 4.9 followed the same procedure as Figure 4.7 and 4.8 except similarities in identifying objects was the objective. The Illinois Curator and SEMAFOR systems obtained the highest similarities in identifying objects, which closely resembles Figure 4.3 where the two systems were in tandem on a individual

Stanford NLP	X	X	X	X	X	X
Boxer (DRT)	X	X	X	X	X	45.68%
Illinois Curator	X	X	X	X	36.58%	13.66%
LTH SRL	X	X	X	75.68%	45.85%	18.26%
SHALMANESER	X	X	53.37%	59.4%	38.26%	13.25%
SEMAFOR	X	45.29%	44.54%	50.70%	43.4%	28.49%
	SEMAFOR	SHALMANESER	LTH SRL	Illinois Curator	Boxer (DRT)	Stanford NLP

Figure 4.8.: Similarities between parsers - Named Entity Identification

(and sorted) sentence basis. Interesting enough, Figure 4.3 also depicts the SHALMANESER and LTH SRL systems in tandem as well, however, it's not well represented in Figure 4.9 compared to other averages where these averages are higher. This could possibly infer that both systems were able to identify a similar number of objects but a different number of objects. In addition, similar to the Figures 4.7 and 4.8, the Stanford NLP system scored the lowest agreement between other parsers.

Stanford NLP	X	X	X	X	X
Illinois Curator	X	X	X	X	36.94%
LTH SRL	X	X	X	64.17%	42.39%
SHALMANESER	X	X	65.52%	67.35%	44.43%
SEMAFOR	X	69.7%	67.58%	73.21%	40.45%
	SEMAFOR	SHALMANESER	LTH SRL	Illinois Curator	Stanford NLP

Figure 4.9.: Similarities between parsers - Object Identification

4.4 Future Improvements

This study provides a glace at the performance of the participating parsers. While these parsers can be improved and refined further to provide more semantic information of natural language text, there can be some improvements made to this specific study to aid in obtaining this information. Below is a list of future improvements that could be used to modify this study.

- Expand the dataset - For two reasons: First by providing more sentences, which would provide a better representation of the parser's performances. Second, by using paragraphs as inputs instead of individual sentences. This can allow for more semantic metrics to be introduced (such as entailment).
- Modify the metrics - As it currently stands, the analysis was done to keep efficiency and fairness when parsing the sentences. Therefore, a binary count of arguments determined the scoring. Modifying this metric to allow for partial scoring could result in higher accuracy of arguments for predicates.
- Create an automated tool - The results of this study only show the calculations and scores of the performance for each parser. By creating an automated tool that takes raw text and outputs the results from each of the parsers, this allows the ability to reproduce the results of this study (and future studies) of the laborious work to setup and configure the parsers and parse the sentences.

4.5 Final Summary

In this thesis, the study explored a variety of semantic parsers that have been regarded as "golden standards" to some. The goal of this study was to provide a snapshot performance of the participating parsers across multiple types of parsers by grading the accuracy to a human annotator in hopes to highlight and stress the quality of performance to the industry.

From a computational linguistic standpoint and based on the results of these experiments, there is not a specific system, out of the tools used, that can be used for all circumstances. Rather, different tools are better suited for different tasks. For event identification, some of the best systems for this task were the LTH SRL and SEMAFOR systems. For the named entities identification, the Stanford NER tagger performed well. For identifying objects, the Illinois Curator and SEMAFOR systems performed better than others. The Boxer DRT and LTH SRL systems performed well associating correct arguments to each predicate. In general, all participating systems performed well when disambiguating word senses.

While these parsers primarily input formal and well-structured sentences to extract semantic information, it is possible to input unstructured, informal and colloquial text. However, the results would be limited in terms of understanding the given text. This may be attributed to inherent imperfections that these systems use to identify new and unknown words. In the cases parsers that rely on more statistical mechanics (such as Boxer or Stanford NLP), these could possibly extract better results. However, based on the results of this thesis, even these systems can be limited.

LIST OF REFERENCES

LIST OF REFERENCES

Allan, J., Papka, R., & Lavrenko, V. (1998). On-line new event detection and tracking. In *Proceedings of the 21st annual international acm sigir conference on research and development in information retrieval* (pp. 37–45).

Angeli, G., Premkumar, M. J., & Manning, C. D. (2015). Leveraging linguistic structure for open domain information extraction. *Linguistics*(1/24).

Baker, C. F., & Fellbaum, C. (2009). WordNet and FrameNet as complementary resources for annotation. *Proceedings of the Third Linguistic Annotation*(August), 125–129. Retrieved from `http://dl.acm.org/citation.cfm?id=1698402` doi: 10.3115/1698381.1698402

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet Project. *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 86–90. Retrieved from `http://dx.doi.org/10.3115/980845.980860` doi: 10.3115/980845.980860

Beltagy, I., Erk, K., & Mooney, R. (2014). Semantic Parsing using Distributional Semantics and Probabilistic Logic. *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, 7–11. Retrieved from `http://www.aclweb.org/anthology/W/W14/W14-2402`

Blackburn, P., & Bos, J. (2005). Representation and inference for natural language. *Semantics*.

Bos, J. (2008). Wide-coverage semantic analysis with boxer. In *Proceedings of the 2008 conference on semantics in text processing* (pp. 277–286).

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Pado, S., & Pinkal, M. (2006). Salto–a versatile multi-level annotation tool. In *Proceedings of lrec 2006* (pp. 517–520).

Carreras, X., & Marquez, L. (2005). Introduction to the CoNLL-2005 Shared Task : Semantic Role Labeling. In *Conll '05 proceedings of the ninth conference on computational natural language learning* (pp. 152–164).

Chinchor, N., Brown, E., Ferro, L., & Robinson, P. (1999). 1999 named entity recognition task definition. *MITRE and SAIC*.

Dipanjan, D., Schneider, N., Desai, C., & Smith, N. a. (2010). SEMAFOR 1.0: A probabilistic frame-semantic parser. *... Institute, School of ...*, 1–20.

Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S., & Weischedel, R. M. (2004). The automatic content extraction (ace) program-tasks, data, and evaluation. In *Lrec* (Vol. 2, p. 1).

Edmonds, P. (2002). SENSEVAL: The evaluation of word sense disambiguation systems. *ELRA newsletter*, *7*(3), 5–14.

Ellsworth, M., Erk, K., Kingsbury, P., & Padó, S. (2004). Propbank, salsa, and framenet: How design determines product. In *Proc. of lrec* (pp. 17–23).

Erk, K., & Pado, S. (2006). SHALMANESER A Toolchain For Shallow Semantic Parsing. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, *6*(2), 527–532. Retrieved from http://www.nlpado.de/ sebastian/pub/papers/lrec06_erk.pdf

Erk, K., & Padó, S. (2008). A structured vector space model for word meaning in context. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 897–906).

Fillmore, C. (1982). Frame semantics. *Linguistics in the morning calm*, 111–137.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics* (pp. 363–370).

Francez, N. (2014). A logic inspired by natural language: quantifiers as subnectors. *Journal of Philosophical Logic*, *43*(6), 1153–1172.

Gildea, D., & Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational linguistics*, *28*(3), 245–288.

Grishman, R., & Sundheim, B. (1996a). Message Understanding Conference-6: A Brief History. *Proceedings of the 16th conference on Computational linguistics*, *1*, 466–471. Retrieved from http://portal.acm.org/citation.cfm?doid=992628.992709 doi: 10.3115/992628.992709

Grishman, R., & Sundheim, B. (1996b). Message understanding conference-6: A brief history. In *Coling* (Vol. 96, pp. 466–471).

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 146–162. Retrieved from http://psycnet.apa.org/psycinfo/1956-02807-001

Hermann, K. M., Das, D., Weston, J., & Ganchev, K. (2014). Semantic Frame Identification with Distributed Word Representations. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1448–1458. Retrieved from http://www.aclweb.org/anthology/P14-1136

Hickey, R. (2003). *Motives for language change.* Cambridge: Cambridge University Press. Retrieved from http://books.google.be/books?id=GOx9R3MFIW4C doi: 10.1017/CBO9780511486937

Ide, N., & Véronis, J. (1998). Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics*, *24*(1), 1–40. doi: 10.1016/j.csl.2004.05.005

Johansson, R., & Nugues, P. (2008). Dependency-based semantic role labeling of PropBank. *Proceedings of the Conference on Empirical Methods in Natural Language Processing - EMNLP '08*(October), 69–78. Retrieved from http://portal.acm.org/citation.cfm?doid=1613715.1613726 doi: 10.3115/1613715.1613726

Jurgens, D., & Stevens, K. (2010). The S-Space Package: An Open Source Package for Word Space Models. *In Proceedings of the ACL 2010 System Demonstrations, Uppsala, Sweden, 13 July 2010*, 30–35. Retrieved from papers3://publication/uuid/E8CF42C3-9467-41E2-99BF-97BA6E329BB5

Kingsbury, P., Palmer, M., & Marcus, M. (2002). Adding predicate argument structure to the penn treebank. In *Proceedings of the second international conference on human language technology research* (pp. 252–256). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=1289189.1289207

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

Laorden, C., Santos, I., Sanz, B., Alvarez, G., & Bringas, P. G. (2012). Word sense disambiguation for spam filtering. *Electronic Commerce Research and Applications*, *11*(3), 290–298.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *From context to meaning: Distributional models of the lexicon in linguistics and cognitive science, special issue of the Italian Journal of Linguistics*, *20*(1), 1–31.

Lund, C. B., & Kevin. (1997). Modelling parsing constraints with high-dimensional context space. *Language and cognitive processes*, *12*(2-3), 177–210.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: System demonstrations* (pp. 55–60).

Màrquez, L., Carreras, X., Litkowski, K. C., & Stevenson, S. (2008). Semantic Role Labeling: An Introduction to the Special Issue. *Computational Linguistics*, *34*, 145–159. Retrieved from http://eprints.pascal-network.org/archive/00004514/ doi: 10.1162/coli.2008.34.2.145

Matsubayashi, Y., Okazaki, N., & Tsujii, J. (2014). Generalization of semantic roles in automatic semantic role labeling. *Information and Media Technologies*, *9*(4), 736–770.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., & Grishman, R. (2004, May 2 - May 7). The nombank project: An interim report. In A. Meyers (Ed.), *Hlt-naacl 2004 workshop: Frontiers in corpus annotation* (pp. 24–31). Boston, Massachusetts, USA: Association for Computational Linguistics.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, *30*(1), 3–26. doi: 10.1075/li.30.1.03nad

Palmer, A., & Sporleder, C. (2010). Evaluating FrameNet-style semantic parsing: the role of coverage gaps in FrameNet. *Computational Linguistics*(August), 928–936. Retrieved from `http://eprints.pascal-network.org/archive/00007109/`

Punyakanok, V., Roth, D., & Yih, W.-t. (2008). The Importance of Syntactic Parsing and Inference in Semantic Role Labeling. *Computational Linguistics*, *34*, 257–287. doi: 10.1162/coli.2008.34.2.257

Ritter, A., Etzioni, O., Clark, S., et al. (2012). Open domain event extraction from twitter. In *Proceedings of the 18th acm sigkdd international conference on knowledge discovery and data mining* (pp. 1104–1112).

Schtze, H. (1993). Word space. In *Advances in neural information processing systems 5* (pp. 895–902). Morgan Kaufmann.

Sekine, S., Sudo, K., & Nobata, C. (2002). Extended named entity hierarchy. In *Lrec.*

Sheth, A., Aleman-Meza, B., Arpinar, I. B., Bertram, C., Warke, Y., Ramakrishanan, C., … others (2005). Semantic association identification and knowledge discovery for national security applications. *Journal of Database Management (JDM)*, *16*(1), 33–53.

Tjong Kim Sang, E. F. (2002). Introduction to the conll-2002 shared task: Language-independent named entity recognition. In *Proceedings of conll-2002* (pp. 155–158). Taipei, Taiwan.

Tjong Kim Sang, E. F., & De Meulder, F. (2003). Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proceedings of the seventh conference on natural language learning at hlt-naacl 2003-volume 4* (pp. 142–147).

Van Emden, M. H., & Kowalski, R. A. (1976). The semantics of predicate logic as a programming language. *Journal of the ACM (JACM)*, *23*(4), 733–742.

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., & Pustejovsky, J. (2007). Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th international workshop on semantic evaluations* (pp. 75–80).

Yang, Y., Pierce, T., & Carbonell, J. (1998). A study of retrospective and on-line event detection. In *Proceedings of the 21st annual international acm sigir conference on research and development in information retrieval* (pp. 28–36).

Zhou, D., & He, Y. (2011). Semantic parsing for biomedical event extraction. , 395–399.

APPENDIX

APPENDIX A: Experiment 1 Results

- Their assault began at 9:20 p.m. Friday, when one terrorist detonated a suicide bomb outside the gates of the soccer stadium on the northern outskirts of Paris.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	3	3	Assault(1), suicide(0)	Assault(1), suicide(0)	Assault(Y), suicide(Y)	NA
SHALMANESER	2	2	4	Assault(1), suicide(1)	Assault(3), suicide(3)	Assault(Y), suicide(Y)	NA
LTH SRL	2	2	1	Assault(1), detonated(3),	Assault(1), detonated(3)	Assault(Y), detonated(Y),	NA
Illinois Curator	1	1	4	detonated(2)	detonated(2)	detonated(Y)	NA
Boxer (DRT)	1	2	NA	detonated(2)	NA	NA	NA
Stanford NLP	1	3	1	Assault(1)	NA	NA	NA
Manual annotation	Assault, detonated, suicide	9:20 p.m. (time), Friday (day), one terrorist (N_count), soccer (sports), Paris (City)	Their, Bomb, gates, stadium, outskirts	Assault(their, 9:20p.m., Friday, detonated), detonated(one terrorist ,bomb, outside the gates), suicide(bomb)	Assault(4), detonated(3), suicide(1)	NA	Assault(7), detonated(2), suicide(2),

Table A.1: Experiment 1 - Sentence 1

- In Germany, the police were exploring whether a man they arrested last week with weapons in his car and his GPS navigator set for Paris was linked to the attacks.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	4	Exploring(2), arrested(2), attacks(0)	Exploring(2), arrested(3), attacks(0)	Exploring(Y), arrested(Y), attacks(Y)	NA
SHALMANESER	2	0	4	arrested(0), attacks(0)	arrested(2), attacks(0)	arrested(Y), attacks(Y)	NA
LTH SRL	2	2	3	Exploring(2), arrested(2)	Exploring(3), arrested(4)	Exploring(Y), arrested(Y)	NA
Illinois Curator	2	2	4	Exploring(2), arrested(3)	Exploring(3), arrested(4)	Exploring(Y), arrested(Y)	NA
Boxer (DRT)	2	2	NA	Exploring(0), arrested(0)	NA	NA	NA
Stanford NLP	2	2	3	Exploring(2), arrested(1)	NA	NA	NA
Manual annotation	Exploring, arrested, attacks	Germany (country), last week(date), Paris (city)	Police, man, weapons, car, GPS navigator	Exploring(police, man, Germany), arrested(police, man, last week), attacks(Paris)	Exploring(3), arrested(3), attacks(1)	NA	Exploring(4), arrested(4), attacks(15)

Table A.2: Experiment 1 - Sentence 2

- The possibility that one of the attackers was a migrant or had posed as one is sure to further complicate the already vexing problem for Europe of how to handle the unceasing flow of people from Syria, Iraq, Afghanistan and elsewhere.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	2	handle(0), flow(0)	handle(1), flow(1)	handle(Y), flow(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	4	0	2	Posed(0), Complicate(1), handle(0), flow(1)	Posed(2), Complicate(4), handle(1), flow(3)	Posed(Y), Complicate(Y), handle(Y), flow(Y)	NA
Illinois Curator	3	1	3	Posed(1), Complicate(0), handle(0)	Posed(2), Complicate(1), handle(1)	Posed(Y), Complicate(Y), handle(Y)	NA
Boxer (DRT)	3	3	NA	Posed(1), Complicate(0), handle(0)	NA	NA	NA
Stanford NLP	1	4	1	Posed(0)	NA	NA	NA
Manual annotation	Posed, Complicate, handle, flow	Europe (location-continent), Syria(location-country), Iraq(location-country), Afghanistan(location-country)	Possibility(attacker was a migrant, posed as one), attackers(migrant, posed as one), migrant, problem(vexing), flow(unceasing), people(Syria, Iraq, Afghanistan, elsewhere)	Posed(attackers, migrant), Complicate(possibility, problem, Europe, handle), handle(flow), flow(people, Syria, Iraq, Afghanistan, elsewhere)	Posed(2), Complicate(4), handle(1), flow(5)	NA	Posed(7), Complicate(2), handle(7), flow(14)

Table A.3: Experiment 1 - Sentence 3

- The attacks, and the threat of the Islamic State, are likely to dominate a summit meeting of leaders of the Group of 20 nations that starts on Sunday in Turkey.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	0	Attacks(0), dominate(0), meeting(0)	Attacks(0), dominate(1), meeting(1)	Attacks(Y), dominate(Y), meeting(Y)	NA
SHALMANESER	2	1	2	Attacks(0), meeting(0)	Attacks(0), meeting(1)	Attacks(Y), meeting(Y)	NA
LTH SRL	2	2	1	dominate(1), meeting(0)	Attacks(NA), dominate(2), meeting(2)	dominate(Y), meeting(Y)	NA
Illinois Curator	1	4	2	dominate(1)	dominate(2)	dominate(Y)	NA
Boxer (DRT)	1	3	NA	dominate(1)	NA	NA	NA
Stanford NLP	1	2	1	dominate(1)	NA	NA	NA
Manual annotation	Attacks, dominate, meeting	Islamic State (organization-political org.), Group of 20 nations (organization - government), Sunday (timex - day of week), Turkey(location-country)	Threat(Islamic State), leaders(Group of 20 nations)	Attacks(Islamic State), dominate(attacks, meeting), meeting(leaders, Sunday, Turkey)	Attacks(1), dominate(2), meeting(3)	NA	Attacks(15), dominate(5), meeting(17)

Table A.4: Experiment 1 - Sentence 4

- The death toll far surpassed that of a massacre at the satirical newspaper Charlie Hebdo and related attacks by Islamic extremists around the French capital in January.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	2	2	Surpassed(0), massacre(1), attacks(1)	Surpassed(1), massacre(1), attacks(3)	Surpassed(Y), massacre(Y), attacks(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	3	0	0	Surpassed(0), massacre(0), attacks(1)	Surpassed(2), massacre(1), attacks(1)	Surpassed(Y), massacre(Y), attacks(Y)	NA
Illinois Curator	1	0	1	Surpassed(0)	Surpassed(2)	Surpassed(Y)	NA
Boxer (DRT)	1	1	NA	Surpassed(2)	NA	NA	NA
Stanford NLP	1	0	1	Surpassed(1)	NA	NA	NA
Manual annotation	Surpassed, massacre, attacks	Charlie Hebdo(org - company), Islamic extremists (org), French capital (loc – city), January(date)	Toll(death), newspaper(satirical , Charlie Hebdo)	Surpassed(toll, massacre), massacre(Charlie Hebdo), attacks(Islamic extremists, French capital, January)	Surpassed(2), massacre(1), attacks(3)	NA	Surpassed(4), massacre(2), attacks(15)

Table A.5: Experiment 1 - Sentence 5

- About 1,500 people are packed into the small concert hall in the 11th arrondissement in eastern Paris for a sold-out concert by the Californian rock group Eagles of Death Metal when rounds of machine gunfire ring out at 9.40pm, 45 minutes into the performance.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	4	3	Packed(2), ring out(1)	Packed(2), ring out(1)	Packed(Y),ring out(Y)	NA
SHALMANESER	2	1	3	Packed(1), ring out(0)	Packed(3), ring out(3)	Packed(Y), ring out(Y)	NA
LTH SRL	2	1	1	Packed(1), concert(1)	Packed(2), concert(2)	Packed(Y), concert(Y)	NA
Illinois Curator	1	0	2	Packed(2)	Packed(3)	Packed(Y)	NA
Boxer (DRT)	2	1	NA	Packed(2), ring out(3)	NA	NA	NA
Stanford NLP	0	2	1		NA	NA	NA
Manual annotation	Packed, concert, gunfire, ring out	1,500 people(numex - N_person), 11th arrondissement (facility – public_institution), Paris(location - city), Californian (org – ethnic group), Eagles of Death Metal (org), 9.40pm(timex), 45 minutes(periodx)	Concert hall(small), rock group, rounds (machine gunfire), performance	Packed(1,500 people, 11th arrondissement, concert hall, eastern Paris), concert(sold-out, Californian rock group, Eagles of Death Metal), gunfire(rounds, machine), ring out(gunfire, 9:40 pm, 45 minutes)	Packed(4), concert(4), gunfire(2), ring out(4)	NA	Packed(15), concert(3), gunfire(1), ring out(1)

Table A.6: Experiment 1 - Sentence 6

- Calmly he begins to fire at people enjoying an evening at La Belle Equipe, which had been fully booked on a typically busy night in this popular part of Paris.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	3	Fire(2), enjoying(2)	Fire(2), enjoying(2)	Fire(Y), enjoying(Y)	NA
SHALMANESER	2	0	3	Fire(1), enjoying(1)	Fire(2), enjoying(2)	Fire(Y), enjoying(Y)	NA
LTH SRL	2	1	3	Fire(1), enjoying(2)	Fire(2), enjoying(2)	Fire(Y), enjoying(Y)	NA
Illinois Curator	2	0	4	Fire(2), enjoying(3)	Fire(2), enjoying(3)	Fire(Y), enjoying(Y)	NA
Boxer (DRT)	2	1	NA	Fire(0), enjoying(3)	NA	NA	NA
Stanford NLP	1	2	2	Fire(2)	NA	NA	NA
Manual annotation	Fire, enjoying	La Belle Equipe(facility), evening(timex), Paris(location - city)	He, People, night (typically busy), part (popular)	Fire(he, people (enjoying), evening, La Belle Equipe(fully booked, night)), enjoying(people, evening, La Belle Equipe(fully booked, night), Paris)	Fire(4), enjoying(4)	NA	Fire(22), enjoying(5)

Table A.7: Experiment 1 - Sentence 7

- The coordinated nature of Friday's deadly attacks across Paris points to an organized terrorist group, security experts say, giving credence to ISIS' claim of responsibility.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	3	Attacks(2), say(1), giving(2), claim(0)	Attacks(2), say(2), giving(3), claim(1)	Attacks(Y), say(Y), giving(Y), claim(Y)	NA
SHALMANESER	4	0	4	Attacks(2), say(1), giving(1), claim(0)	Attacks(3), say(1), giving(1), claim(1)	Attacks(Y), say(Y), giving(Y), claim(Y)	NA
LTH SRL	4	2	2	Attacks(2), say(1), giving(2), claim(1)	Attacks(2), say(3), giving(2), claim(1)	Attacks(Y), say(Y), giving(Y), claim(Y)	NA
Illinois Curator	2	0	4	say(0), giving(2)	say(2), giving (3)	say(Y), giving(Y)	NA
Boxer (DRT)	2	1	NA	say(2), giving(2)	NA	NA	NA
Stanford NLP	0	3	0		NA	NA	NA
Manual annotation	Attacks, points, say, giving, claim	Friday(timex – day_of_week), Paris (location – city), ISIS(org)	Nature(coordinated), group (organized terrorist), experts(security), credence(ISIS, claim), claim(responsibility)	Attacks(deadly, nature, Friday, Paris, group), points(attacks, group, experts), say(experts, attacks), giving(experts, credence, ISIS), claim(responsibility)	Attacks(5), points(3), say(2), giving(3), claim(1)	NA	Attacks(15), points(39), say(12), giving(48), claim(11)

Table A.8: Experiment 1 - Sentence 8

- ISIS has sought to extend its reach globally, with an ISIS affiliate claiming credit for the recent downing of a Russian passenger plane over Egypt and a bombing in Lebanon that killed more than 40 people this week.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	**4**	1	3	Sought(1), claiming(1), bombing(0), killed(1)	Sought(1), claiming(2), bombing(0), killed(1)	Sought(Y), claiming(Y),bombing(Y), killed(Y)	NA
SHALMANESER	4	0	4	Sought(1), claiming(2), downing(0), killed(0)	Sought(1),claiming(3), downing(2), killed(1)	Sought(Y),claiming(Y),downing(Y), killed(Y)	NA
LTH SRL	6	3	1	Sought(2), extend(1), claiming(0), downing(0), bombing(1), killed(3)	Sought(3), extend(3), claiming(2),downing(2), bombing(1),killed(4)	Sought(Y),extend(Y),claiming(Y),downing(Y), bombing(Y), killed(Y)	NA
Illinois Curator	4	1	3	Sought(1), extend(2), claiming(2), killed(2)	Sought(1), extend(4), claiming(1), killed(2)	Sought(Y),extend(Y),claiming(Y), killed(Y)	NA
Boxer (DRT)	4	4	NA	Sought(0), extend(1), claiming(2), killed(2)	NA	NA	NA
Stanford NLP	0	4			NA	NA	NA
Manual annotation	Sought, Extend, claiming, downing, bombing, killed	ISIS(org - nationality), ISIS(org - nationality), Russian (org – nationality), Egypt(location - country), Lebanon(location – city), 40 people (countx), this week(date)	reach(globally), affiliate(ISIS), credit, plane(passenger)	Sought(ISIS, extend), Extend(ISIS, reach), claiming(ISIS affiliate, credit, downing), downing(Russian, plane, Egypt), bombing(Lebanon, bombing), killed(Lebanon, 40 people, week)	Sought(2), extend(2), claiming(3), downing(3), bombing(2), killed(3)	NA	Sought(7), extend(17), claiming(5), downing(7), bombing(4), killed(15)

Table A.9: Experiment 1 - Sentence 9

- On a night when thousands of Paris residents and tourists were reveling and fans were enjoying a soccer match between France and world champion Germany, horror struck in an unprecedented manner.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship – Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	5	enjoying(2), match(2), struck(2)	enjoying(2), match(1), struck(2)	enjoying(Y), match(Y), struck(Y)	NA
SHALMANESER	2	0	3	enjoying(2), struck(1)	enjoying(2), struck(1)	enjoying(Y), match(Y)	NA
LTH SRL	4	0	5	Reveling(3), enjoying(1), match(1), struck(1)	Reveling(2), enjoying(4), match(2), struck(3)	Reveling(Y), enjoying(Y), match(Y), struck(Y)	NA
Illinois Curator	3	0	3	Reveling(1), enjoying(1), struck(2)	Reveling(3), enjoying(3), struck(3)	Reveling(Y), enjoying(Y), struck(Y)	NA
Boxer (DRT)	3	3	NA	Reveling(2), enjoying(2), struck(2)	NA	NA	NA
Stanford NLP	2	1	3	Reveling(1), enjoying(2)	NA	NA	NA
Manual annotation	Reveling, enjoying, match, struck	Paris(location-city),soccer(product-sport), France(org), Germany(org)	Night, residents (Paris, thousands), tourists(Paris, thousands), fans, champion(world), horror, manner, match	Reveling(night, residents, tourists), enjoying(fans, match), match(soccer, France, Germany(champion)), struck(horror, manner(unprecedented))	Reveling(3), enjoying(2), match(2), struck(2)	NA	Reveling(2), enjoying(5), match(17), struck(23)

Table A.10: Experiment 1 - Sentence 10

- U.S. President Barack Obama spoke with French President Francois Hollande to offer condolences and assistance in the investigation, the White House said.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	0	1	Spoke(1), offer(0), investigation(0), said(1)	Spoke(1), offer(1), investigation(0), said(2)	Spoke(), offer(), investigation(), said()	NA
SHALMANESER	3	0	2	Spoke(2), investigation(0), said(1)	Spoke(2), investigation(2), said(2)	Spoke(), investigation(), said()	NA
LTH SRL	3	0	2	Spoke(2), offer(0), said(1)	Spoke(), offer(), said()	Spoke(), offer(), said()	NA
Illinois Curator	3	0	2	Spoke(2), offer(0), said(1)	Spoke(4), offer(3), said(2)	Spoke(), offer(), said()	NA
Boxer (DRT)	3	5	NA	Spoke(2), offer(2), said(0)	NA	NA	NA
Stanford NLP	0	3	0				
Manual annotation	Spoke, offer, investigation, said	U.S.(org), Barack Obama(person), French(org), Francois Hollande(person), White House(org)	President(U.S., Barack Obama), President(French, Francois Hollande), condolences, assistance(investigation)	Spoke(Barack Obama, Francois Hollande, offer), offer(condolences, assistance), investigation, said(White House, spoke(offer()))	Spoke(3), offer(2), investigation(0), said(2)	NA	Spoke(7), offer(16), investigation(2), said(12)

Table A.11: Experiment 1 - Sentence 11

- The American rock band was performing at the Bataclan, a theater located in eastern Paris near the trendy Oberkampf area.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	2	3	Performing(1)	Performing(2)	Performing(Y)	NA
SHALMANESER	0	1	2				NA
LTH SRL	1	2	1	Performing(1)	Performing(2)	Performing(Y)	NA
Illinois Curator	1	2	1	Performing(1)	Performing(2)	Performing(Y)	NA
Boxer (DRT)	1	2	NA	Performing(2)	NA	NA	NA
Stanford NLP	1	1	2	Performing(2)	NA	NA	NA
Manual annotation	Performing	American(org-nationality), Bataclan(facility), Paris(location-city), Oberkampf(location)	Band(rock), theater(Paris(eastern)), area	Performing(American band, Bataclan, Oberkampf, Paris)	Performing(4)	NA	Performing(5)

Table A.12: Experiment 1 - Sentence 12

- The Islamic State of Iraq and Syria has claimed responsibility for the attacks, considered the deadliest on France since World War II.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	1	Claimed(2), attacks(0)	Claimed(2), attacks(2)	Claimed(Y), attacks(Y)	NA
SHALMANESER	2	0	1	Claimed(1), attacks(0)	Claimed(2), attacks(1)	Claimed(Y), attacks(Y)	NA
LTH SRL	1	2	0	Claimed(2)	Claimed(4)	Claimed(Y)	NA
Illinois Curator	1	1	1	Claimed(1)	Claimed(2)	Claimed(Y)	NA
Boxer (DRT)	1	3	NA	Claimed(1)	NA	NA	NA
Stanford NLP	1	0	1	Claimed(1)	NA	NA	NA
Manual annotation	Claimed, attacks	Islamic State of Iraq and Syria(org), France(org), World War II(event-war)	Responsibility	Claimed(Islamic State of Iraq and Syria, responsibility(attacks)), attacks(Islamic State of Iraq and Syria, France)	Claimed(2), attacks(2)	NA	Claimed(5), attacks(15)

Table A.13: Experiment 1 - Sentence 13

- CNN reported a total of 112 people were killed at the concert hall, along with 14 at a Cambodian restaurant nearby, 19 outside a bar near the Stade de France outside the city, four at the stadium, and four elsewhere.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	5	Reported(1), killed(1)	Reported(1), killed(1)	Reported(Y), killed(Y)	NA
SHALMANESER	2	0	2	Reported(1), killed(1)	Reported(2), killed(1)	Reported(Y), killed(Y)	NA
LTH SRL	2	0	2	Reported(1), killed(3)	Reported(2), killed(5)	Reported(Y), killed(Y)	NA
Illinois Curator	2	0	3	Reported(1), killed(1)	Reported(2), killed(2)	Reported(Y), killed(Y)	NA
Boxer (DRT)	2	2	NA	Reported(2), killed(1)	NA	NA	NA
Stanford NLP	1	0	0	Reported(1)	NA	NA	NA
Manual annotation	Reported, killed	CNN(org), 112 people(countx-n_person), Cambodian(org-nationality), Stade de France(facility)	hall(concert), restaurant(Cambodian), bar(near, Stade de France), city, stadium, elsewhere	Reported(CNN, killed), killed(112 people, hall, restaurant, bar, stadium, elsewhere)	Reported(2), killed(6)	NA	Reported(7), killed(15)

Table A.14: Experiment 1 - Sentence 14

- Gunmen and bombers carried out a wave of attacks on restaurants, a concert hall and near a sports stadium across Paris on Friday in a deadly rampage claimed by Islamic State that killed 129 people and wounded 352, of which 99 remain in a critical condition.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	2	6	Carried out(2), attacks(1),killed(0)	Carried out(2), attacks(1),killed(0)	Carried out(Y), attacks(Y),kille d(Y)	NA
SHALMANESER	4	0	4	Carried out(0), attacks(0), claimed(1),killed(0)	Carried out(2), attacks(1),claimed(2) ,killed(0)	Carried out(Y), attacks(Y),clai med(Y),killed(Y)	NA
LTH SRL	5	0	0	Carried out(1), attacks(0), claimed(0),killed(2), wounded(1)	Carried out(2),attacks(1), claimed(3),killed(4), wounded(2)	Carried out(Y), attacks(Y), claimed(),kille d(Y), wounded(Y)	NA
Illinois Curator	4	0	2	Carried out(0), claimed(1),killed(1), wounded(1)	Carried out(2), claimed(2),killed(2), wounded(3)	Carried out(Y), claimed(Y),kill ed(Y), wounded(Y)	NA
Boxer (DRT)	5	4	NA	Carried out(0), attacks(0), claimed(0),killed(0), wounded(0)	NA	NA	NA
Stanford NLP	0	3	0		NA	NA	NA
Manual annotation	Carried out, attacks, rampage, claimed, killed, wounded	Paris(location-city), Friday(timex-day_of_week), Islamic State (org), 129 people(countx-N-person)	Gunman, bombers, wave, restaurants, hall(concert), stadium(sports), condition(critical)	Carried out (gunman, bombers, attacks), attacks(wave, restaurants, hall, stadium(near), paris(across), Friday, rampage), rampage(deadly), claimed(rampage, Islamic State), killed(Islamic State, 129 people), wounded(352 people, 99 condition)	Carried out(3), attacks(7), rampage(1), claimed(3),killed(2), wounded(2)	NA	Carried out(2), attacks(15), rampage(2), claimed(5),kille d(15), wounded(4)

Table A.15: Experiment 1 - Sentence 15

- A suicide bomber activates an explosive belt near a gate of the sports stadium Stade de France in the northern suburb of Saint Denis, where President Francois Hollande and the German foreign minister were watching a friendly soccer international.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	5	Suicide(0), activates(1), watching(2)	Suicide(0), activates(1), watching(2)	Suicide(Y), activates(Y), watching(Y)	NA
SHALMANESER	2	0	5	Suicide(1), watching(0)	Suicide(1), watching(1)	Suicide(Y), watching(Y)	NA
LTH SRL	2	0	4	Activates(1), watching(1)	Activates(2), watching(3)	Activates(Y), watching(Y)	NA
Illinois Curator	2	0	4	Activates(1), watching(0)	Activates(3), watching(4)	Activates(Y), watching(Y)	NA
Boxer (DRT)	2	1	NA	activates(3), watching(3)	NA	NA	NA
Stanford NLP	1	2	4	activates(2)	NA	NA	NA
Manual annotation	Suicide, activates, watching	Stade de France(facility), Saint Denis(location), Francois Hollande (person), German (org-nationality), soccer(product-sport)	Bomber(suicide), belt(explosive), gate, stadium(sports), suburb(northern), President, minister(foreign), international	Suicide(bomber), Activates(bomber, belt, gate(stadium)), watching(Francois Hollande, minister(German), international(soccer))	Suicide(1), activates(3), watching(3)	NA	Suicide(2), activates(5), watching(8)

Table A.16: Experiment 1 - Sentence 16

- In Auckland and in the US, among other countries, protestors have voiced their anger about the trade pact over the past several months.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	0	3	Voiced(1)	Voiced(2)	Voiced(Y)	NA
SHALMANESER	1	2	4	Voiced(0)	Voiced(2)	Voiced(Y)	NA
LTH SRL	1	0	3	Voiced(3)	Voiced(5)	Voiced(Y)	NA
Illinois Curator	1	0	5	Voiced(4)	Voiced(3)	Voiced(Y)	NA
Boxer (DRT)	1	2	NA	Voiced(0)	NA	NA	NA
Stanford NLP	1	2	1	Voiced(1)	NA	NA	NA
Manual annotation	Voiced	Auckland(location-city), US(location-country)	Countries(other), protestors, anger, pact(trade), months(past, several)	Voiced(protestors, Auckland, US, countries, anger, pact, months)	Voiced(7)	NA	Voiced(3)

Table A.17: Experiment 1 - Sentence 17

- Japanese Prime Minister Shinzo Abe, facing one of his biggest tests since his second turn at the helm began in 2012, is set to fare better than when scandals helped bring down his first administration.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	4	Facing(0), helped(2)	Facing(0), helped(2)	Facing(Y), helped(Y)	NA
SHALMANESER	1	2	3	bring down(1)	bring down(3)	bring down(Y)	NA
LTH SRL	3	1	3	Facing(2), helped(2), bring down(1)	Facing(3), helped(3), bring down(2)	Facing(Y), helped(Y), bring down(Y)	NA
Illinois Curator	3	1	3	Facing(1), helped(2), bring down(1)	Facing(2), helped(2), bring down(2)	Facing(Y), helped(Y), bring down(Y)	NA
Boxer (DRT)	3	1	NA	Facing(0), helped(0), bring down(1)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Facing, began, scandals, helped, bring down	Japanese(org-ethnic group), Shinzo Abe(person), 2012(timex-date)	Minister(prime), tests(biggest), turn(second), helm, administration(first)	Facing(Shinzo Abe(Japanese Minister),tests(helm),turn), scandals(Shinzo Abe), helped(scandals, administration), bring down(administration)	Facing(3), scandals(1), helped(2), bring down(1)	NA	Facing(13), scandals(2), helped(8), bring down(6)

Table A.18: Experiment 1 - Sentence 18

- That suggests he has learned lessons from a tumultuous six year period in Japanese politics before the 2012 election, when the country saw a revolving door of six prime ministers.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	3	6	Learned(2), election(1), saw(1)	Learned(2), election(1), saw(2)	Learned(Y), election(Y), saw(Y)	NA
SHALMANESER	2	1	4	Learned(1), saw(1)	Learned(2), saw(2)	Learned(Y), saw(Y)	NA
LTH SRL	2	1	3	Learned(1), saw(1)	Learned(2), saw(4)	Learned(Y), saw(Y)	NA
Illinois Curator	2	1	4	Learned(2), saw(1)	Learned(3), saw(1)	Learned(Y), saw(Y)	NA
Boxer (DRT)	2	0	NA	Learned(1), saw(2)	NA	NA	NA
Stanford NLP	2	1	4	Learned(2), saw(2)	NA	NA	NA
Manual annotation	Learned, election, saw, revolving	Six year (periodx), Japanese (org-nationality), 2012(year), six prime ministers(countx-N_person)	He, lessons, period (six year), politics(Japanese), country, door	Learned(He, lessons, Japanese politics), election(2012), saw(country, revolving, six prime ministers), revolving(door, six prime ministers)	Learned(3), election(1), saw(3), revolving(2)	NA	Learned(9), election(4), saw(28), revolving(3)

Table A.19: Experiment 1 - Sentence 19

- Continued control of parliament means Abe's administration would retain the ability to pursue a reform program that includes joining the TransPacific Partnership, a U.S. led freetrade agreement, opening up the agriculture sector, and establishing zones with less business regulation.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	7	0	5	Control(0), retain(0), pursue(1), joining(1), opening(0), establishing(1), regulation(1)	Control(1), retain(2), pursue(2), joining(1), opening(1), establishing(1), regulation(1)	Control(Y), retain(N),pursue(Y),joining(Y),opening(Y), establishing(Y), regulation(Y)	NA
SHALMANESER	5	0	5	retain(0), pursue(0), joining(0), opening(0), establishing(0)	retain(1), pursue(1), joining(1), opening(0), establishing(0)	Control(Y), retain(Y),pursue(),joining(Y), opening(N), establishing(Y), regulation(Y)	NA
LTH SRL	7	1	5	Control(1), retain(0), pursue(0), joining(0), opening(1), establishing(1), regulation(1)	Control(2), retain(3), pursue(2), joining(2), opening(2), establishing(2), regulation(1)	Control(Y), retain(Y), pursue(Y), joining(Y), opening(Y), establishing(Y), regulation(Y)	NA
Illinois Curator	Unavailable						
Boxer (DRT)	5	1	NA	retain(1), pursue(1), joining(1), opening(1), establishing(0)	NA	NA	NA
Stanford NLP	2	1	2	retain(1), establishing(1)	NA	NA	NA
Manual annotation	Control, retain, pursue, joining, opening, establishing, regulation	Abe(person), TransPacific Partnership(product-other), U.S.(org)	Parliament, administration(Abe), ability, program(reform), agreement(freetrade), sector(agriculture), zones	Control(parliament, administration), retain(ability, pursue), pursue(program, joining, opening, establishing), joining(TransPacific Partnership), opening(sector), establishing(zones), regulation(business, less)	Control(2), retain(2), pursue(4), joining(1), opening(1), establishing(1), regulation(2)	NA	Control(20), retain(4), pursue(4), joining(6), opening(25), establishing(8), regulation(7)

Table A.20: Experiment 1 - Sentence 20

- With each of their parties engrossed in the heat of presidential primary season, President Barack Obama and House Speaker Paul Ryan were set to meet Tuesday, even as prospects for new bipartisan agreement appear to diminish by the day.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	2	3	Engrossed(2), meet(0), diminish(1)	Engrossed(2), meet(0), diminish(1)	Engrossed(Y), meet(Y), diminish(Y)	NA
SHALMANESER	2	0	5	meet(0), diminish(1)	meet(3), diminish(2)	meet(Y), diminish(Y)	NA
LTH SRL	3	1	2	Engrossed(2), meet(1), diminish(1)	Engrossed(2), meet(2), diminish(2)	Engrossed(Y), meet(Y), diminish(Y)	NA
Illinois Curator	3	1	6	Engrossed(1), meet(1), diminish(0)	Engrossed(3), meet(4), diminish(2)	Engrossed(Y), meet(Y), diminish(Y)	NA
Boxer (DRT)	3	3	NA	Engrossed(2), meet(1), diminish(0)	NA	NA	NA
Stanford NLP	2	2	1	meet(3), diminish(0)	NA	NA	NA
Manual annotation	Engrossed, meet, diminish	President(product-position_title), House Speaker (product-position_title, Barack Obama(person), Paul Ryan(person), Tuesday(timex-day_of_week)	Parties, heat, season(presidential primary), prospects, agreement(new, bipartisan), day	Engrossed(parties, season), meet(Barack Obama, Paul Ryan, Tuesday), diminish(prospects(agreement))	Engrossed(2), meet(3), diminish(1)	NA	Engrossed(4), meet(13), diminish(2)

Table A.21: Experiment 1 - Sentence 21

- They sparred over fiscal matters when Ryan headed the House Budget Committee, and competed on the campaign trail during Ryan's stint as the 2012 GOP vice presidential nominee.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	4	Headed(0), competed(0)	Headed(1), competed(1)	Headed(N), competed(N)	NA
SHALMANESER	1	0	4	headed(2)	headed(2)	headed(Y)	NA
LTH SRL	3	1	2	Sparred(1), headed(2), competed(2)	Sparred(2), headed(3), competed(3)	Sparred(Y), headed(Y), competed(Y)	NA
Illinois Curator	3	0	4	Sparred(2), headed(2), competed(1)	Sparred(3), headed(3), competed(3)	Sparred(Y), headed(Y), competed(Y)	NA
Boxer (DRT)	3	3	NA	Sparred(1), headed(2), competed(2)	NA	NA	NA
Stanford NLP	3	5	4	Sparred(2), headed(2), competed(2)	NA	NA	NA
Manual annotation	Sparred, headed, competed	Ryan(person), House Budget Committee(org), Ryan(person), 2012(year), GOP(org), vice presidential nominee(product-position_title)	They, Matters(fiscal), trail(campaign), stint(Ryan)	Sparred(They, matters, 2012), headed(Ryan, House Budget Committee), competed(they, trail, Ryan, 2012)	Sparred(3), headed(2), competed(4)	NA	Sparred(4), headed(13), competed(8)

Table A.22: Experiment 1 - Sentence 22

- Of the items that remain on Obama's governing agenda, most appear dead on Capitol Hill, including passing comprehensive immigration reform, advancing new gun control laws and closing the military prison at Guantanamo Bay.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	5	2	5	Remain(2), appear(0), passing(1), advancing(0), closing(1)	Remain(2),appear(1), passing(1),advancing (0),closing(2)	Remain(Y),appear(Y),passing (N),advancing(Y), closing(Y)	NA
SHALMANESER	5	2	3	Remain(0), appear(1), passing(1), advancing(1), closing(1)	Remain(1),appear(2), passing(2),advancing (5),closing(1)	Remain(Y), appear(Y), passing(Y), advancing(Y), closing(Y)	NA
LTH SRL	5	1	4	Remain(2), appear(2), passing(1), advancing(1), closing(0)	Remain(3),appear(5), passing(1),advancing (1), closing(1)	Remain(Y),appear(Y),passing (Y),advancing(Y), closing(Y)	NA
Illinois Curator	4	0	5	Remain(2), appear(2), passing(1), advancing(1), closing(2)	Remain(3),appear(4), passing(2),advancing (), closing(3)	Remain(Y),appear(Y),passing (Y),advancing(Y), closing(Y)	NA
Boxer (DRT)	5	1	NA	Remain(2), appear(0), passing(1), advancing(1), closing(0)	NA	NA	NA
Stanford NLP	2	2	2	appear(2), advancing(1)	NA	NA	NA
Manual annotation	Remain, appear, passing, advancing, closing	Obama(person), Capitol Hill(org), Guantanami Bay(facility)	Items(agenda, dead), agenda(Obama, governing), reform(comprehensive, immigration), laws(new, gun control), prison(military)	Remain(items, Obama's, agenda), appear(items(agenda), dead, Capitol Hill), passing(reform), advancing(laws), closing(prison, Guantanamo Bay)	Remain(2),appear(3), passing(1), advancing(1), closing(2)	NA	Remain(4), appear(7), passing(38), advancing(13), closing(23)

Table A.23: Experiment 1 - Sentence 23

- Hundreds of protesters snarled traffic in Auckland, New Zealand on Thursday to protest the signing of a controversial trade pact that was years in the making.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	2	Snarled(1), making(0)	Snarled(1),making(0)	Snarled(Y),making(Y)	NA
SHALMANESER	1	0	2	Snarled(2)	Snarled(2)	Snarled(Y)	NA
LTH SRL	4	2	2	Snarled(4), protest(1), signing(0), making(0)	Snarled(5), protest(2), signing(1), making(0)	Snarled(Y), protest(Y), signing(Y), making(Y)	NA
Illinois Curator	2	2	4	Snarled(3), protest(0)	Snarled(5), protest(2)	Snarled(Y), protest(Y)	NA
Boxer (DRT)	2	3	NA	Snarled(0), protest(0),	NA	NA	NA
Stanford NLP	1	3	2	Snarled(3)	NA	NA	NA
Manual annotation	Snarled, protest, signing, making	Auckland(location), New Zealand(location), Thursday(time-day_of_week)	Protesters(hundreds), traffic, trade pact(controversial), years	snarled(protesters, traffic, Auckland, New Zealand, Thursday), protest(Hundreds of protesters, signing), signing(trade pact), making(trade pact, years)	Snarled(5), protest(2), signing(1), making(2)	NA	Snarled(5), protest(6), signing(9), making(52)

Table A.24: Experiment 1 - Sentence 24

- The Obama administration quickly endorsed the analysis from the Washington based Peterson Institute for International Economics to buttress its uphill fight for Congress's approval of the TransPacific Partnership, completed last October after years of negotiations.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	5	1	2	Endorsed(1), analysis(0), buttress(0), fight(0), negotiations(0)	Endorsed(1), analysis(0), buttress(1), fight(1), negotiations(0)	Endorsed(Y), analysis(Y), buttress(Y), fight(Y), negotiations(Y)	NA
SHALMANESER	2	0	3	analysis(0), fight(0)	analysis(0), fight(3)	analysis(Y), fight(Y)	NA
LTH SRL	3	1	1	Endorsed(2), buttress(0), fight(0)	Endorsed(4), buttress(2), fight(3)	Endorsed(Y), buttress(Y), fight(Y)	NA
Illinois Curator	Unavailable						
Boxer (DRT)	2	3	NA	Endorsed(2), buttress(1)	NA	NA	NA
Stanford NLP	1	4	1	Endorsed(2)	NA	NA	NA
Manual annotation	Endorsed, analysis, buttress, fight, negotiations	Obama administration(org), Washington(location), Peterson Institute for International Economics(org), Congress(org), TransPacific Partnership(product-other), October(date)	Analysis, fight(uphill), approval(Congress's)	Endorsed(Obama administration, analysis), analysis(Peterson Institute for International Economics, Washington), buttress(Obama administration, fight), fight(Obama administration, approval), negotiations(Obama administration, Congress, TransPacific Partnership, last October)	Endorsed(2), analysis(2), buttress(3), fight(2), negotiations(4)	NA	Endorsed(4), analysis(6), buttress(3), fight(9), negotiations(2)

Table A.25: Experiment 1 - Sentence 25

- He also played down the risk of the Australian government being sued under controversial investor-state dispute settlement (ISDS) clauses, pointing to carve-outs for health and environmental measures.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	3	Played(0), pointing(0)	Played(0), pointing(2)	Played(Y), pointing(Y)	NA
SHALMANESER	1	0	2	Played(1)	Played(1)	Played(Y)	NA
LTH SRL	3	0	1	Played(1), sued(2), pointing(0)	Played(4), sued(2), pointing(2)	Played(Y), sued(Y), pointing(Y)	NA
Illinois Curator	3	0	3	Played(1), sued(2), pointing(0)	Played(2), sued(2), pointing(1)	Played(Y), sued(Y), pointing(Y)	NA
Boxer (DRT)	3	0	NA	Played(1), sued(0), pointing(3)	NA	NA	NA
Stanford NLP	2	0	2	sued(1), pointing(1)	NA	NA	NA
Manual annotation	Played, sued, pointing	Australian government(org), investor-state dispute settlement (ISDS) clauses (product-rule)	He, risk, carve-outs, health, measures(environ mental)	Played(he, risk, sued), sued(Australian government, investor-state dispute settlement (ISDS) clauses, pointing), pointing(carve-outs, health, measures)	Played(3), sued(2), pointing(3)	NA	Played(37), sued(1), pointing(14)

Table A.26: Experiment 1 - Sentence 26

- U.S. Trade Representative Michael Froman has said the current administration is doing everything in its power to move the deal and on Thursday told reporters he was confident the deal would get the necessary support in Congress.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship – Count	WSD - Ident.	WSD - Count
SEMAFOR	5	2	5	Said(1), doing(1), move(0), told(2), support(0)	Said(2), doing(2), move(0), told(2), support(0)	Said(Y),doing(Y), move(Y), told(Y),support(Y)	NA
SHALMANESER	3	0	6	Said(1), move(0), told(2)	Said(1), move(0), told(2)	Said(Y),move(y), told(Y)	NA
LTH SRL	5	1	3	Said(1), doing(2), move(1), told(2), support(1)	Said(2), doing(2), move(1), told(4), support(2)	Said(Y),doing(Y), move(Y), told(), support(Y)	NA
Illinois Curator	4	0	4	Said(1), doing(2), move(1), told(2)	Said(2), doing(2), move(1), told(3)	Said(Y),doing(Y), move(Y), told(Y)	NA
Boxer (DRT)	4	3	NA	Said(1), doing(0), move(1), told(3)	NA	NA	NA
Stanford NLP	0	3	0		NA	NA	NA
Manual annotation	Said, doing, move, told, support	U.S. Trade Representative (org-position_title), Michael Froman(person), Thursday(date), Congress(org)	Administration(current), power(administration), deal, reporters, deal, support	Said(Michael Froman(U.S. Trade representative), administration(power), doing), doing(administration, move), move(administration, deal), told(Thursday, Michael Froman(U.S. Trade Representative), reporters, deal, support(necessary)), support(deal, Congress)	Said(3), doing(2), move(2), told(5), support(2)	NA	Said(4), doing(16), move(21), told(8), support(20)

Table A.27: Experiment 1 - Sentence 27

- The TransPacific Partnership, one of the world's biggest multinational trade deals, was signed by 12 member nations on Thursday in New Zealand, but the massive trade pact will still require years of tough negotiations before it becomes a reality.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	2	2	Signed(1), negotiations(0)	Signed(1), negotiations(0)	Signed(Y), negotiations(Y)	Signed(), negotiations()
SHALMANESER	1	1	4	Signed(2)	Signed(3)	Signed(Y)	Signed()
LTH SRL	2	0	2	Signed(1), negotiations(0)	Signed(3), negotiations(1)	Signed(Y), negotiations(Y)	Signed(), negotiations()
Illinois Curator	Unavailable						
Boxer (DRT)	1	3	NA	Signed(3)	NA	NA	Signed()
Stanford NLP	1	3	2	Signed(2)	NA	NA	Signed(), negotiations()
Manual annotation	Signed, negotiations	TransPacific Partnership(product-other), 12 member nations(countx-N_org), Thursday(date), New Zealand(location)	Deals(worlds biggest, multinational, trade), pact(massive, trade), years, reality	Signed(deals(TransPacific Partnership),12 member nations, Thursday, New Zealand), negotiations(pact, tough, reality)	Signed(4), negotiations(3)	NA	Signed(10), negotiations(2)

Table A.28: Experiment 1 - Sentence 28

- The TPP will now undergo a two year ratification period in which at least six countries - that account for 85 percent of the combined gross domestic production of the 12 TPP nations - must approve the final text for the deal to be implemented.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	3	3	Undergo(1), approve(1), implemented(0)	Undergo(2), approve(2), implemented(0)	Undergo(Y), approve(Y), implemented(Y)	NA
SHALMANESER	1	2	3	Undergo(1)	Undergo(3)	Undergo(Y)	NA
LTH SRL	3	1	2	Undergo(1), approve(0), implemented(1)	Undergo(5), approve(4), implemented(1)	Undergo(Y), approve(Y), implemented(Y)	NA
Illinois Curator	Too long						
Boxer (DRT)	3	1	NA	Undergo(1), approve(2), implemented(0)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Undergo, approve, implemented	TPP(product-other), two year(date measurement), six countries(countx-N_org), 85 , 12 TPP nations(countx-N_org)	Period(ratification), gross domestic production(combined), text(final), deal	Undergo(TPP, period(two year), six countries), approve(text, implemented), implemented(deal)	Undergo(3), approve(2), implemented(1)	NA	Undergo(1), approve(2), implemented(4)

Table A.29: Experiment 1 - Sentence 29

- U.S. Trade Representative Michael Froman has said the current administration is doing everything in its power to move the deal and on Thursday told reporters he was confident the deal would get the necessary support in Congress.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship – Count	WSD - Ident.	WSD - Count
SEMAFOR	5	2	5	Said(), doing(), move(), told(), support()	Said(2), doing(2), move(0), told(2), support(0)	Said(Y),doing(Y), move(Y), told(Y),support(Y)	NA
SHALMANESER	3	0	6	Said(1), move(0), told(2)	Said(1), move(0), told(2)	Said(Y),move(y), told(Y)	NA
LTH SRL	5	1	3	Said(1), doing(2), move(1), told(2), support(1)	Said(2), doing(2), move(1), told(4), support(2)	Said(Y),doing(Y), move(Y), told(), support(Y)	NA
Illinois Curator	4	0	4	Said(1), doing(2), move()1, told(2)	Said(2), doing(2), move(1), told(3)	Said(Y),doing(Y), move(Y), told(Y)	NA
Boxer (DRT)	4	3	NA	Said(1), doing(0), move(1), told(3)	NA	NA	NA
Stanford NLP	0	3	0		NA	NA	NA
Manual annotation	Said, doing, move, told, support	U.S. Trade Representative (org-position_title), Michael Froman(person), Thursday(date), Congress(org)	Administration(current), power(administration), deal, reporters, deal, support	Said(Michael Froman(U.S. Trade representative), administration(power), doing), doing(administration, move), move(administration, deal), told(Thursday, Michael Froman(U.S. Trade Representative), reporters, deal, support(necessary)), support(deal, Congress)	Said(3), doing(2), move(2), told(5), support(2)	NA	Said(4), doing(16), move(21), told(8), support(20)

Table A.30: Experiment 1 - Sentence 30

- Opponents have criticized the secrecy surrounding TPP talks, raised concerns about reduced access to things like affordable medicines, and a clause which allows foreign investors the right to sue if they feel their profits have been impacted by a law or policy in the host country.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	7	Criticized(2), talks(0), raised(0)	Criticized(2), talks(0), raised(1)	Criticized(Y), talks(Y), raised(Y)	NA
SHALMANESER	3	0	3	Criticized(1), talks(1), raised(1)	Criticized(2), talks(1), raised(2)	Criticized(Y), talks(Y), raised(Y)	NA
LTH SRL	4	0	7	Criticized(1), talks(1), raised(0), sue(1)	Criticized(2), talks(1), raised(1), sue(3)	Criticized(Y), talks(Y), raised(Y), sue(Y)	NA
Illinois Curator	3	0	10	Criticized(1), raised(1), sue(1)	Criticized(2), raised(2), sue(2)	Criticized(Y), raised(Y), sue(Y)	NA
Boxer (DRT)	2	0	NA	raised(2), sue(0)	NA	NA	NA
Stanford NLP	0	0	3		NA	NA	NA
Manual annotation	Criticized, talks, raised, sue	TPP (product-other)	Opponents, secrecy, concerns, access, medicines(affordable), clause, investors(foreign), right, profits(investors), law, policy, country(host)	Criticized(Opponents, secrecy(talks(TPP))), talks(TPP), raised(Opponents, concerns, access, medicines, clause(investors, sue(profits))), sue(investors, country)	Criticized(2), talks(1), raised(5), sue(2)	NA	Criticized(2), talks(12), raised(30), sue(2)

Table A.31: Experiment 1 - Sentence 31

- U.S. government and industry officials said on Monday they expected the Trans-Pacific Partnership (TPP) trade deal signed by 12 countries earlier this month to bolster trade in aviation products once it was enacted.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	3	4	Said(1), signed(2), bolster(1), trade(1)	Said(3), signed(3), bolster(1), trade(1)	Said(Y), signed(Y), bolster(Y), trade(Y)	NA
SHALMANESER	2	1	1	Said(0), signed(0)	Said(2), signed(3)	Said(Y), signed(Y)	NA
LTH SRL	4	1	3	Said(1), signed(2), bolster(0), trade(1)	Said(3), signed(4), bolster(2), trade(1)	Said(Y), signed(Y), bolster(Y), trade(Y)	NA
Illinois Curator	3	1	4	Said(1), signed(2), bolster(0)	Said(3), signed(3), bolster(3)	Said(Y), signed(Y), bolster(Y)	NA
Boxer (DRT)	3	1	NA	Said(2), signed(2), bolster(1)	NA	NA	NA
Stanford NLP	1	1	1	Said(2)	NA	NA	NA
Manual annotation	Said, signed, bolster, trade	U.S. government(org), Monday(date), Trans-Pacific Partnership(TPP)(product-other), 12 countries(countx-N_org)	Officials(industry), deal(trade), month, trade, products(aviation)	Said(U.S. government, officials, Monday, Trans-Pacific Partnership(TPP), signed), signed(deal, 12 countries, month(earlier)), bolster(deal, trade), trade(products)	Said(5), signed(4), bolster(2), trade(1)	NA	Said(12), signed(10), bolster(4), trade(12)

Table A.32: Experiment 1 - Sentence 32

- While Trump has had the nomination locked down for weeks, he has now reached the threshold of 1,237 delegates with the help of previously uncommitted delegates who now support his candidacy.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	3	Reached(1), support(0)	Reached(3), support(1)	Reached(Y), support(Y)	NA
SHALMANESER	1	0	1	reached(0)	reached(2)	Nomination(Y), reached(Y), support(Y)	NA
LTH SRL	3	0	4	Nomination(1), reached(1), support(2)	Nomination(1), reached(5), support(4)	Nomination(Y), reached(Y), support(Y)	NA
Illinois Curator	2	0	4	Reached(1), support(2)	Reached(3), support(5)	Nomination(Y), reached(Y), support(Y)	NA
Boxer (DRT)	2	0	NA	reached(2), support(1)	NA	NA	NA
Stanford NLP	1	0	1	reached(1)	NA	NA	NA
Manual annotation	Nomination, reached, support	Trump(person), 1,237 delegates(countx-other)	Weeks, threshold, help, delegates(previously uncommitted), candidacy	Nomination(Trump, locked down, for weeks), reached(Trump, threshold(1,237 delegates)), support(delegates, candidacy)	Nomination(3), reached(2), support(2)	NA	Nomination(3), reached(9), support(22)

Table A.33: Experiment 1 - Sentence 33

- For months, drama and tumult have rocked the Republican Party, as a fervent anti-Trump movement launched a full-on onslaught to derail his candidacy.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	4	Rocked(1), launched(1), derail(1)	Rocked(1), launched(2), derail(1)	Rocked(Y), launched(Y), derail(Y)	NA
SHALMANESER	2	0	3	launched(2), derail(1)	launched(3), derail(2)	launched(Y), derail(Y)	NA
LTH SRL	3	0	3	Rocked(2), launched(2), derail(1)	Rocked(4), launched(3), derail(2)	Rocked(Y), launched(Y), derail(Y)	NA
Illinois Curator	3	0	4	Rocked(2), launched(1), derail(1)	Rocked(3), launched(3), derail(2)	Rocked(Y), launched(Y), derail(Y)	NA
Boxer (DRT)	3	1	NA	Rocked(4), launched(1), derail(1)	NA	NA	NA
Stanford NLP	2	1	2	launched(2), derail(1)	NA	NA	NA
Manual annotation	Rocked, launched, derail	Republican Party (org-political_party), anti-Trump movement(org)	Months, drama, tumult, onslaught(full-on), candidacy	Rocked(drama, tumult, Republican Party, for months), launched(anti_Trump movement, onslaught(derail)), derail(candidacy(Trump))	Rocked(4), launched(3), derail(1)	NA	Rocked(2), launched(6), derail(2)

Table A.34: Experiment 1 - Sentence 34

- Hillary Clinton claimed the Democratic presidential nomination on Tuesday night after decisive victories in the California, New Jersey and New Mexico primaries, and she quickly appealed to supporters of Senator Bernie Sanders of Vermont to unite with her against Donald J. Trump.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	2	3	Claimed(1), appealed(1), unite(0)	Claimed(2), appealed(1), unite(0)	Claimed(Y), appealed(Y), unite(Y)	NA
SHALMANESER	3	0	3	Claimed(1), appealed(0), unite(2)	Claimed(2), appealed(1), unite(3)	Claimed(Y), appealed(Y), unite(Y)	NA
LTH SRL	3	0	4	Claimed(1), appealed(3), unite(2)	Claimed(2), appealed(4), unite(3)	Claimed(Y), appealed(Y), unite(Y)	NA
Illinois Curator	3	1	3	Claimed(3), appealed(1), unite(3)	Claimed(4), appealed(2), unite(3)	Claimed(Y), appealed(Y), unite(Y)	NA
Boxer (DRT)	3	3	NA	Claimed(3), appealed(1), unite(2)	NA	NA	NA
Stanford NLP	1	4	3	appealed(2)	NA	NA	NA
Manual annotation	Claimed, appealed, unite	Hillary Clinton(person), Democratic presidential nomination(org-position_title), Tuesday night(date), California primaries(event), New Jersey primaries(event), New Mexico primaires(event), Senator(org-position_title), Bernie Sanders of Vermont(person), Donald J. Trump(person)	Victories(decisi ve), she, supporters, her	Claimed(Hillary Clinton, Democratic presidential nomination, Tuesday night), appealed(she, supporters(Senator Bernie Sanders of Vermont)), unite(supporters(Senator Bernie Sanders of Vermont), her, against Donlad J. Trump)	Claimed(3), appealed(2), unite(3)	NA	Claimed(5), appealed(5), unite(6)

Table A.35: Experiment 1 - Sentence 35

- With the 14 month Democratic race nearing a close, Mrs. Clinton savored the biggest night of her extraordinary journey from lawyer, wife and first lady to senator, secretary of state and, now, the first woman to win a major party's nomination.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	3	4	Savored(2), win(1)	Savored(2), win(1)	Savored(Y), win(Y)	NA
SHALMANESER	0	2	4				NA
LTH SRL	2	1	2	Savored(1), win(1)	Savored(3), win(1)	Savored(Y), win(Y)	NA
Illinois Curator	Too long						
Boxer (DRT)	2	3	NA	Savored(1), win(1)	NA	NA	NA
Stanford NLP	1	1	1	Savored(2)	NA	NA	NA
Manual annotation	Savored, win	14 month Democratic race(periodx-period_month),Democratic race(event), Mrs.Clinton(person), first lady(product-title), senator(product-title), secretary of state(product-title)	Night(biggest), journey(extradinary), lawyer, wife, woman(first), nomination(major party)	Savored(Mrs. Clinton, night(journey)), win(woman, nomination)	Savored(2), win(2)	NA	Savored(4), win(7)

Table A.36: Experiment 1 - Sentence 36

- He was left hoping for a longshot victory in the California primary, to justify staying in the race and lobbying Democratic officials to support him in a contested convention next month.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	0	4	Victory(1), staying(1), lobbying(2), support(2)	Victory(2), staying(2), lobbying(2), support(3)	Victory(Y), staying(Y), lobbying(Y), support(Y)	NA
SHALMANESER	4	0	2	Victory(1), staying(1), race(0), lobbying(0)	Victory(1), staying(1), race(0), lobbying(3)	Victory(Y), staying(), race(Y), lobbying(Y)	NA
LTH SRL	3	0	4	Victory(1), lobbying(1), support(3)	Victory(2), lobbying(2), support(4)	Victory(Y), lobbying(Y), support(Y)	NA
Illinois Curator	Unavailable						
Boxer (DRT)	2	0	NA	lobbying(1), support(3)	NA	NA	NA
Stanford NLP	1	0	2	support(1)	NA	NA	NA
Manual annotation	Victory, staying, race, lobbying, support	California primary(event), Democratic officials(person)	He, him, convention(contested), month(next)	Victory(He(hoping), California primary), staying(he, race), race(he), lobbying(He, Democratic officials, support), support(him, convention, month)	Victory(2), staying(2), race(1), lobbying(3), support(3)	NA	Victory(1), staying(11),race (10), lobbying(1), support(23)

Table A.37: Experiment 1 - Sentence 37

- The unexpected news on Monday set off conversations within the two campaigns, with Clinton representatives preparing to make overtures to the Sanders camp as early as Wednesday.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	4	2	Set off(0), conversations(0), preparing(1), make(1)	Set off(1), conversations(0), preparing(2), make(3)	Set off(Y), conversations(Y), preparing(Y), make(Y)	NA
SHALMANESER	4	1	1	Set off(0), conversations(0), preparing(0), make(1)	Set off(2), conversations(0), preparing(1), make(3)	Set off(Y), conversations(Y), preparing(Y), make(Y)	NA
LTH SRL	4	2	1	Set off(1), preparing(1), make(1)	Set off(4), preparing(4), make(4)	Set off(Y), preparing(Y), make(Y)	NA
Illinois Curator	3	1	1	Set off(0), preparing(0), make(0)	Set off(3), preparing(1), make(2)	Set off(Y), preparing(Y), make(Y)	NA
Boxer (DRT)	3	2	NA	Set off(1), preparing(0), make(1)	NA	NA	NA
Stanford NLP	1	2	0	make(1)	NA	NA	NA
Manual annotation	Set off, conversations, preparing, make	Monday(date), Clinton representative(org), Sanders camp(org), Wednesday(date), two campaigns(countx)	News(unexpected), overtures	Set off(conversations, news, Monday), conversations(two campaigns), preparing(Clinton representatives, Sanders camp, make), make(overtures)	Set off(3), conversations(1), preparing(3), make(1)	NA	Set off(7), conversations(1), preparing(8), make(51)

Table A.38: Experiment 1 - Sentence 38

- They said he had been wholly focused on winning the California primary in hopes that a victory in such a politically important state would help him lobby superdelegates to shift their support.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	6	0	5	Said(1), focused(0), winning(0), Help(1), lobby(1), shift(1)	Said(2), focused(1), winning(1), Help(2), lobby(2), shift(1)	Said(Y), focused(Y), winning(Y), Help(Y), lobby(Y), shift(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	6	0	4	Said(1), focused(0), winning(1), Help(1), lobby(1), shift(1)	Said(2), focused(3), winning(2), Help(4), lobby(3), shift(2)	Said(Y), focused(Y), winning(Y), Help(Y), lobby(Y), shift(Y)	NA
Illinois Curator	5	0	2	Said(1), focused(0), winning(1), Help(0), shift(2)	Said(2), focused(1),winning(2), Help(3), shift(4)	Said(Y), focused(Y), winning(Y), Help(Y), shift(Y)	NA
Boxer (DRT)	6	0	NA	Said(1), focused(1), winning(0), Help(0), lobby(0), shift(2)	NA	NA	NA
Stanford NLP	0	0	0		NA	NA	NA
Manual annotation	Said, focused, winning, help, lobby, shift	California primary(event)	They, State(important(politically)), victory, superdelegates, support	Said(they, focused), focused(they, winning), winning(California primary), help(state, him, lobby, shift), lobby(superdelegates), shift(superdelegates, support)	Said(2), focused(2), winning(1), Help(4), lobby(1), shift(2)	NA	Said(12), focused(8), winning(8), Help(13), lobby(4), shift(23)

Table A.39: Experiment 1 - Sentence 39

- The two had a tense but productive meeting that Thursday night, and two days later she delivered a speech thanking her supporters and urging them to unite around the party's presumptive nominee.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	6	2	2	Meeting(0),delivered(1), speech(0), thanking(0), urging(1), unite(0)	Meeting(0),delivered (3), speech(0), thanking(0), urging(2), unite(0)	Meeting(Y),delivered(Y),speech(N),thanking(Y),urging(Y), unite(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	5	2	2	delivered(1), speech(1), thanking(1), urging(1), unite(0)	delivered(4),speech(1), thanking(2),urging(3) , unite(3)	Meeting(Y),delivered(Y),speech(Y),thanking(Y), urging(Y), unite(Y)	NA
Illinois Curator	4	2	3	delivered(1), thanking(1), urging(0), unite(1)	delivered(3), thanking(2), urging(3), unite(2)	delivered(Y), thanking(Y), urging(Y), unite(Y)	NA
Boxer (DRT)	3	1	NA	thanking(1), urging(1), unite(0)	NA	NA	NA
Stanford NLP	3	0	0	delivered(1), urging(1), unite(0)	NA	NA	NA
Manual annotation	Meeting, delivered, speech, thanking, urging, unite	The two(countx-N_person), Thursday night(date), two days(timex)	Supporters, them, nominee(party presumptive)	Meeting(the two, Thursday night), delivered(two days(Thursday night), speech), speech(she, thanking), thanking(supporters), urging(she, them), unite(nominee)	Meeting(2), delivered(2), speech(2), thanking(1), urging(2), unite(1)	NA	Meeting(17), delivered(12), speech(8), thanking(1), urging(6), unite(6)

Table A.40: Experiment 1 - Sentence 40

- Embrace his brand of racial politics, one that could taint the party's image well beyond this election, or abandon their presumptive nominee and hand the White House to the Democrats for another four years.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	2	2	Embrace(0), election(0), abandon(0), hand(0)	Embrace(0), election(0),abandon(2), hand(0)	Embrace(Y), election(Y), abandon(Y), hand(N)	NA
SHALMANESER	2	0	4	abandon(1), hand(3)	abandon(1), hand(5)	abandon(Y), hand(Y)	NA
LTH SRL	4	1	2	Embrace(0), taint(0), abandon(1), hand(3)	Embrace(1), taint(5), abandon(1), hand(3)	Embrace(Y), taint(Y), abandon(Y), hand(Y)	NA
Illinois Curator	2	1	4	abandon(1), hand(3)	abandon(3), hand(4)	abandon(Y), hand(Y)	NA
Boxer (DRT)	2	0	NA	Embrace(2), taint(1)	NA	NA	NA
Stanford NLP	0	1	3		NA	NA	NA
Manual annotation	Embrace, taint, election, abandon, hand	White House (facility), Democrats(org), four years(timex)	Brand, his, politics(racial), one, image(party's), nominee(presumptive)	Embrace(brand(politics), his), taint(image), election(), abandon(nominee), hand(White House, Democrats, four years)	Embrace(2), taint(1), election(0), abandon(1), hand(3)	NA	Embrace(6), taint(3), election(4), abandon(7), hand(16)

Table A.41: Experiment 1 - Sentence 41

- He took questions from reporters for about 30 minutes, speaking with a backdrop of more than a dozen unpledged North Dakota delegates who are now supporting him.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	2	Took(3), speaking(1), supporting(0)	Took(3), speaking(2), supporting(1)	Took(Y), speaking(Y), supporting(Y)	NA
SHALMANESER	2	1	4	Took(1), speaking(1)	Took(2), speaking(2)	Took(Y), speaking(Y)	NA
LTH SRL	3	1	4	Took(2), speaking(1), supporting(1)	Took(4), speaking(2), supporting(4)	Took(Y), speaking(Y), supporting(Y)	NA
Illinois Curator	3	1	5	Took(3), speaking(0), supporting(2)	Took(3), speaking(1), supporting(4)	Took(Y), speaking(Y), supporting(Y)	NA
Boxer (DRT)	3	1	NA	Took(2), speaking(2), supporting(2)	NA	NA	NA
Stanford NLP	1	1	2	Took(3)	NA	NA	NA
Manual annotation	Took, speaking, supporting	30 minutes(timex), North Dakota(location)	He, questions, reporters, backdrop, delegates, him	Took(He, questions, reporters, 30 minutes), speaking(he, backdrop, delegates(North Dakota)), supporting(delegates(North Dakota, him)	Took(4), speaking(3), supporting(2)	NA	Took(42), speaking(8), supporting(14)

Table A.42: Experiment 1 - Sentence 42

- If elected on Nov. 8, the 68 year old former U.S. senator from New York would return the Clinton family to the White House 16 years after her husband, Bill Clinton, completed two terms as president.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	4	3	Elected(0), return(1), completed(2)	Elected(0), return(2), completed(3)	Elected(Y), return(Y), completed(Y)	NA
SHALMANESER	2	1	3	Elected(0), return(1)	Elected(0), return(2)	Elected(Y), return(Y)	NA
LTH SRL	3	2	2	Elected(1), return(2), completed(1)	Elected(1), return(8), completed(2)	Elected(Y), return(Y), completed(Y)	NA
Illinois Curator	Unavailable						
Boxer (DRT)	3	3	NA	Elected(0), return(1), completed(2)	NA	NA	NA
Stanford NLP	2	4	0	Elected(1), return(1)	NA	NA	NA
Manual annotation	Elected, return, completed	Nov. 8(date), 68 year old (), U.S. senator from New York (person), Clinton family(org), White House (facility), 16 years(timex), Bill Clinton (person), two terms(countx)	Her, husband, president	Elected(Nov. 8, U.S. senator from New York, return), return(Clinton family, White House, 16 years), completed(Bill Clinton, president,two terms)	Elected(3), return(3), completed(3)	NA	Elected(3), return(29), completed(8)

Table A.43: Experiment 1 - Sentence 43

- Trump said money given to the Clinton Foundation charity from foreign donors had earned the Clintons millions of dollars and had a corrupting influence when Clinton was secretary of state.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	2	Said(1), given(2), earned(0)	Said(2), given(2), earned(2)	Said(Y), given(Y), earned(Y)	NA
SHALMANESER	2	1		Said(1), given(3)	Said(2), given(4)	Said(Y), given(Y)	NA
LTH SRL	3	0	2	Said(1), given(1), earned(0)	Said(2), given(2), earned(2)	Said(Y), given(Y), earned(Y)	NA
Illinois Curator	3	0	3	Said(1), given(2), earned(1)	Said(2), given(2), earned(2)	Said(Y), given(Y), earned(Y)	NA
Boxer (DRT)	3	3	NA	Said(1), given(2), earned(2)	NA	NA	NA
Stanford NLP	1	2	0	Said(1)	NA	NA	NA
Manual annotation	Said, given, earned	Trump(person), Clinton Foundation charity(org), Clintons(org), Clinton(person), millions of dollars(money), secretary of state(org-position_title)	Money, donors(foreign), influence(corrupting),	Said(Trump, given, influence), given(money, Clinton Foundation charity, donors, earned), earned(influence, Clinton)	Said(3), given(4), earned(2)	NA	Said(13), given(46), earned(3)

Table A.44: Experiment 1 - Sentence 44

- Hopes flickered among some anti-Trump Republicans that there would be a revolt against him when delegates convene to nominate him formally in Cleveland from July 18 to 21.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	2	Flickered(0), revolt(1), convene(1)	Flickered(0), revolt(1), convene(1)	Flickered(Y), revolt(Y), convene(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	3	2	3	Flickered(2), convene(1), nominate(3)	Flickered(3), convene(3), nominate(5)	Flickered(Y), convene(Y), nominate(Y)	NA
Illinois Curator	Unavailable						
Boxer (DRT)	3	2	NA	Flickered(2), convene(1), nominate(3)	NA	NA	NA
Stanford NLP	2	2	3	Flickered(2), nominate(1)	NA	NA	NA
Manual annotation	Flickered, revolt, convene, nominate	Anti-Trump Republicans(org), Cleveland(location), July 18 to 21(date)	Hopes, him, delegates, him	Flickered(hopes, anti-Trump Republicans), revolt(against him, delegates), convene(delegates, Cleveland, July 18 to 21), nominate(him, Cleveland, July 12 to 21)	Flickered(2), revolt(2), convene(3), nominate(3)	NA	Flickered(3), revolt(4), convene(2), nominate(4)

Table A.45: Experiment 1 - Sentence 45

- Although the text of the 2017 intelligence authorization bill is not yet available to the public, two members of the Senate intelligence committee have said the bill could expand the remit of a nonjudicial subpoena called a National Security Letter (NSLs) to acquire Americans' email records, chat or messaging accounts, account login records, browser histories and social-media service usage.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	3	5	Said(0), expand(2), subpoena(0), acquire(0)	Said(2), expand(2), subpoena(0), acquire(1)	Said(Y), expand(Y), subpoena(Y), acquire(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	3	0	6	Said(0), expand(1), acquire(0)	Said(3), expand(3), acquire(1)	Said(Y), expand(Y), acquire(Y)	NA
Illinois Curator	3	0	7	Said(0), expand(2), acquire(2)	Said(2), expand(3), acquire(2)	Said(Y), expand(Y), acquire(Y)	NA
Boxer (DRT)	3	2	NA	Said(0), expand(2), acquire(2)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Said, expand, subpoena, acquire	2017(year), two members(countx-N-persons), Senate intelligence committee(org), National Security Letter(NSLs)(product-other), Americans(org-nationality)	Text, authorization bill(not available), public, bill, remit, email records, chat, accounts(chat, messaging), records(account login), histories(browser), usage(social-media service)	Said(Senate intelligence committee, bill, expand), expand(bill, remit, subpoena), subpoena(noniuudicial), acquire(NSLs, American, records, accounts, records, histories, usage)	Said(4), expand(3), subpoena(1), acquire(7)	NA	Said(12), expand(7), subpoena(2), acquire(7)

Table A.46: Experiment 1 - Sentence 46

- The senator from Vermont, his voice hoarse, struggled to be heard above screaming supporters in Santa Monica.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	1	Struggled(0), heard(0), screaming(0)	Struggled(0), heard(2), screaming(0)	Struggled(Y), heard(Y), screaming(Y)	NA
SHALMANESER	2	0	1	heard(0), screaming(0)	heard(1), screaming(0)	heard(Y), screaming(Y)	NA
LTH SRL	3	1	2	Struggled(0), heard(0), screaming(0)	Struggled(2), heard(2), screaming(1)	Struggled(Y), heard(Y), screaming(Y)	NA
Illinois Curator	3	1	1	Struggled(0), heard(0), screaming(0)	Struggled(2), heard(2), screaming(2)	Struggled(Y), heard(Y), screaming(Y)	NA
Boxer (DRT)	3	1	NA	Struggled(0), heard(0), screaming(2)	NA	NA	NA
Stanford NLP	1	1	0	Struggled(1)	NA	NA	NA
Manual annotation	Struggled, heard, screaming	Senator from Vermont(person), Santa Monica(location)	Voice(his), supporters	Struggled(senator from Vermont, be heard), heard(senator from Vermont), screaming(supporters, Santa Monica)	Struggled(2), heard(1), screaming(2)	NA	Struggled(4), heard(6), screaming(8)

Table A.47: Experiment 1 - Sentence 47

- Many see her as emblematic of a political system corrupted by corporate influence and the beneficiary of a Democratic establishment that has conspired to ensure she becomes the nominee.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	0	4	conspired(1)	conspired(2)	conspired(Y)	NA
SHALMANESER	1	0	4	conspired(1)	conspired(1)	conspired(Y)	NA
LTH SRL	3	0	4	Corrupted(1), conspired(1), ensure(1)	Corrupted(2), conspired(3), ensure(3)	Corrupted(Y), conspired(Y), ensure(Y)	NA
Illinois Curator	3	0	7	Corrupted(1), conspired(1), ensure(0)	Corrupted(2), conspired(2), ensure(3)	Corrupted(Y), conspired(Y), ensure(Y)	NA
Boxer (DRT)	3	0	NA	Corrupted(2), conspired(1), ensure(0)	NA	NA	NA
Stanford NLP	0	0	0		NA	NA	NA
Manual annotation	Corrupted, conspired, ensure	Democratic establishment(org)	Many, her, system(political), influence(corporate), beneficiary, she, nominee	Corrupted(system, influence, Democratic establishment), conspired(Democratic establishment, ensure), ensure(Democratic establishment, she, becomes nominee)	Corrupted(3), conspired(2), ensure(3)	NA	Corrupted(6), conspired(2), ensure(2)

Table A.48: Experiment 1 - Sentence 48

- Sanders supporters were furious on Monday when the Associated Press and other media outlets declared Clinton the presumptive nominee after she quietly acquired a slew of new superdelegate supporters.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	2	Declared(0), acquired(1)	Declared(3), acquired(2)	Declared(Y), acquired(Y)	NA
SHALMANESER	2	0	3	Declared(0), acquired(2)	Declared(2), acquired(4)	Declared(Y), acquired(Y)	NA
LTH SRL	2	0	2	Declared(2), acquired(2)	Declared(4), acquired(3)	Declared(Y), acquired(Y)	NA
Illinois Curator	2	0	4	Declared(1), acquired(2)	Declared(3), acquired(3)	Declared(Y), acquired(Y)	NA
Boxer (DRT)	2	2	NA	Declared(2), acquired(2)	NA	NA	NA
Stanford NLP	2	3	4	Declared(1), acquired(2)	NA	NA	NA
Manual annotation	Declared, acquired	Sanders supports(org), Monday(date-day_of_week), Associated Press(org), Clinton(person)	Outlets(other media), nominee(presumptive), she, supporters(superdelegates, slew of new)	Declared(outlets, Clinton, nominee), acquired(she, nominee, supporters)	Declared(3), acquired(3)	NA	Declared(10), acquired(8)

Table A.49: Experiment 1 - Sentence 49

- New information also emerged Monday on Mr. Mateen, including frightening statements he made three years ago to co-workers at a security firm about being tied to terrorism, and that the resulting F.B.I. investigation was quite extensive, lasting 10 months.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	3	6	Emerged(1), made(3), investigation(0), lasting(1)	Emerged(1), made(4), investigation(0), lasting(1)	Emerged(Y), made(Y), investigation(Y), lasting(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	4	0	4	Emerged(1), made(3), investigation(1), lasting(1)	Emerged(3), made(4), investigation(2), lasting(1)	Emerged(Y), made(Y), investigation(Y), lasting(Y)	NA
Illinois Curator	3	1	7	Emerged(3), made(2), lasting(1)	Emerged(4), made(4), lasting(2)	Emerged(Y), made(Y), lasting(Y)	NA
Boxer (DRT)	2	1	NA	Emerged(2), made(3)	NA	NA	NA
Stanford NLP	1	3	1	Emerged(3)	NA	NA	NA
Manual annotation	Emerged, made, investigation, lasting	Monday(date-day_of_week), Mr. Mateen(person), three years(periodx-years), F.B.I.(org), 10 months(periodx-months)	Information(new), statements(frightening), he, co-workers, firm(security), terrorism, extensive(investigation)	Emerged(information, Mr. Mateen, Monday), made(he, co-workers, statements, three years ago, firm), investigation(F.B.I., extensive, lasting), lasting(investigation, 10 months)	Emerged(3), made(5), investigation(3), lasting(2)	NA	Emerged(5), including(4), made(51), investigation(2), lasting(8)

Table A.50: Experiment 1 - Sentence 50

- Additional officers rushed to the scene, he said, and entered the nightclub, where they engaged in a gun battle with Mr. Mateen, forcing him to retreat to a bathroom where officers believed he had four to five hostages.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	8	1	7	Rushed(2), said(1), entered(0), engaged(1), gun battle(0), forcing(1), retreat(1), believed(2)	Rushed(2), said(2), entered(2), engaged(2), gun battle(0), forcing(1), retreat(1), believed(2)	Rushed(Y),said(Y),entered(Y),engaged(Y), gun battle(Y),forcing(Y),retreat(Y), believed(Y)	NA
SHALMANESER	6	0	5	Rushed(2), said(1), entered(0), gun battle(2), retreat(1), believed(1)	Rushed(4), said(1), entered(2), gun battle(4), retreat(4), believed(2)	Rushed(Y), said(Y),entered(Y), gun battle(Y),retreat(Y), believed(Y)	NA
LTH SRL	8	0	8	Rushed(2), said(1), entered(0), engaged(1), gun battle(1), forcing(1), retreat(1), believed(1)	Rushed(2), said(2), entered(2), engaged(5), gun battle(2), forcing(3), retreat(2), believed(4)	Rushed(Y),said(Y),entered(Y),engaged(Y), gun battle(Y),forcing(Y),retreat(Y),believed(Y)	NA
Illinois Curator	7	0	5	Rushed(2), said(1), entered(0), engaged(1), forcing(0), retreat(1), believed(1)	Rushed(2), said(2), entered(1), engaged(2), forcing(1), retreat(2), believed(2)	Rushed(Y),said(Y) entered(Y),engaged(Y), forcing(Y), retreat(Y),believed(Y)	NA
Boxer (DRT)	7	0	NA	Rushed(2), said(2), entered(2), engaged(0), forcing(0), retreat(1), believed(1)	NA	NA	NA
Stanford NLP	1	1	0	engaged(2)	NA	NA	NA
Manual annotation	Rushed, said, entered, engaged, gun battle, forcing, retreat, believed	Mr. Mateen(person), four to five hostages(countx-N-persons)	Officers(additional), scene, he, they, nightclub, bathroom, officers, he	Rushed(officers, scene), said(he, rushed), entered(officers, nightclub), engaged(they, gun battle, Mr. Mateen), gun battle(they, Mr. Mateen), forcing(him, retreat), retreat(him, bathroom), believed(officers, he, had four to five hostages)	Rushed(2), said(2), entered(2), engaged(3), gun battle(2), forcing(2), retreat(2), believed(2)	NA	Rushed(8), said(12), entered(9), engaged(17), gun battle(0), forcing(8), retreat(11), believed(5)

Table A.51: Experiment 1 - Sentence 51

- The victims of the Pulse nightclub massacre shared much in common: they were members of the LGBT community, for the most part, and Orlando being Orlando, many worked in tourism or travel.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	5	Massacre(1), worked(0)	Massacre(1), worked(1)	Massacre(Y), worked(Y)	NA
SHALMANESER	2	0	0	Massacre(1), worked(0)	Massacre(2), worked(2)	Massacre(Y), worked(Y)	NA
LTH SRL	3	0	2	Massacre(2), shared(0), worked(0)	Massacre(3), shared(2), worked(3)	Massacre(Y), shared(Y), worked(Y)	NA
Illinois Curator	3	0	3	shared(0), being(2), worked(0)	shared(3), being(3), worked(3)	Massacre(Y), shared(Y), being(Y), worked(Y)	NA
Boxer (DRT)	2	2	NA	shared(1), worked(2)	NA	NA	NA
Stanford NLP	1	2	1	worked(1)	NA	NA	NA
Manual annotation	Massacre, shared, being, worked	Pulse nightclub(facility), Orlando(location), Orlando(location)	Victims, members, LGBT community, tourism, travel	Massacre(victims, Pulse nightclub), shared(victims), being(Orlando, Orlando), worked(members(LGBT community), tourism, travel)	Massacre(2), shared(1), being(1), worked(3)	NA	Massacre(2), shared(7), being(15), worked(27)

Table A.52: Experiment 1 - Sentence 52

- Vigils are held in front of the White House in Washington, on the streets of San Francisco, in the Boystown neighbourhood of Chicago and in the parking lot of a famous LGBT bistro in Atlanta.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	1	3	Held(1)	Held(2)	Held(Y)	NA
SHALMANESER	1	0	1	Held(1)	Held(4)	Held(Y)	NA
LTH SRL	1	0	2	Held(1)	Held(4)	Held(Y)	NA
Illinois Curator	1	1	4	Held(2)	Held(4)	Held(Y)	NA
Boxer (DRT)	1	5	NA	Held(5)	NA	NA	NA
Stanford NLP	1	5	3	Held(2)	NA	NA	NA
Manual annotation	Held	White House(facility), Washington(location), San Francisco(location), Boystown neighborhood(location), Chicago(location), Atlanta(location)	Vigils, streets, lot(parking), bistro(famous, LGBT)	Held(vigils, White House, Washington, streets, San Francisco, Boystown neighbourhood of Chicago, lot, bistro, Atlanta)	Held(9)	NA	Held(37)

Table A.53: Experiment 1 - Sentence 53

- With the sudden bang of a gavel Saturday night, representatives of 195 nations reached a landmark accord that will, for the first time, commit nearly every country to lowering planet-warming greenhouse gas emissions to help stave off the most drastic effects of climate change.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	2	4	Bang(0), reached(0), commit(0), lowering(1)	Bang(1), reached(2), commit(2), lowering(1)	Bang(Y), reached(Y), commit(Y), lowering(Y)	NA
SHALMANESER	4	0	4	reached(0), commit(1), lowering(0), help stave(0)	reached(2), commit(1), lowering(3), help stave(0)	reached(Y), commit(Y), lowering(Y), help stave(Y)	NA
LTH SRL	4	1	4	reached(0), commit(1), lowering(1), emissions(1)	reached(3), commit(5), lowering(2), emissions(1)	reached(Y), commit(Y), lowering(Y), emissions(Y)	NA
Illinois Curator	3	1	6	reached(0), commit(2), lowering(0)	reached(4), commit(4), lowering(1)	reached(Y), commit(Y), lowering(Y)	NA
Boxer (DRT)	2	1	NA	commit(0), lowering(2)	NA	NA	NA
Stanford NLP	0	0	0		NA	NA	NA
Manual annotation	Bang, reached, commit, lowering, emissions, help stave	Saturday night(date-day_of_week), 195 nations(countx)	Gavel, representative, accord(landmark), time(first), country(every), gas(planet-warming, greenhouse), effects(drastic), change(climate)	Bang(sudden, gavel, Saturday night), reached(195 nations, representative, accord), commit(accord, country, lowering), lowering(gas, emissions), emissions(gas), help stave(according, effects)	Bang(3), reached(3), commit3(), lowering(2), emissions(1), help stave(2)	NA	Bang(12), reached(9), commit(6), lowering(8), emissions(5), help stave(12)

Table A.54: Experiment 1 - Sentence 54

- Traditionally, such pacts have required developed economies like the United States to take action to lower greenhouse gas emissions, but they have exempted developing countries like China and India from such obligations.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	5	Required(0), take(0), lower(0), emissions(0)	Required(2), take(2), lower(1), emissions(0)	Required(Y), take(Y), lower(Y), emissions(Y)	NA
SHALMANESER	2	0	4	take(1), lower(0)	take(2), lower(6)	take(Y), lower(Y)	NA
LTH SRL	4	0	5	Required(1), take(1), emissions(1), exempted(1)	Required(4), take(4), emissions(2), exempted(3)	Required(Y), take(Y), emissions(Y), exempted(Y)	NA
Illinois Curator	2	0	6	Required(1), take(1)	Required(3), take(2)	Required(Y), take(Y)	NA
Boxer (DRT)	2	3	NA	Required(2), take(3)	NA	NA	NA
Stanford NLP	1	0	2	exempted(1)	NA	NA	NA
Manual annotation	Required, take, lower, emissions, exempted	United States(org), China(org), India(org)	Pacts, economies(developed), action, gas(greenhouse, emissions), countries(developing), obligations	Required(pacts, economies, take), take(economies, action, lower), lower(gas, emissions), emissions(gas), exempted(pacts, countries, obligations)	Required(3), take(3), lower(2), emissions(1), exempted(3)	NA	Required(6), take(44), lower(17), emissions(5), exempted(2)

Table A.55: Experiment 1 - Sentence 55

- The agreement, adopted after 13 days of intense bargaining in a Paris suburb, puts the world's nations on a course that could fundamentally change the way energy is produced and consumed, gradually reducing reliance on fossil fuels in favor of cleaner forms of energy.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD – Ident.	WSD - Count
SEMAFOR	5	1	5	Adopted(0), change(0), produced(0), consumed(0), reducing(0)	Adopted(1), change(2), produced(1),consumed(0), reducing(1)	Adopted(Y),change(Y),produced(Y),consumed(Y), reducing(Y)	NA
SHALMANESER	4	0	4	Adopted(2), produced(1), consumed(1), reducing()	Adopted(3),produced(1),consumed(1),reducing(7)	Adopted(Y),produced(Y),consumed(Y),reducing(Y)	NA
LTH SRL	6	0	4	Adopted(1), bargaining(0),change(0), produced(1), consumed(1), reducing(0)	Adopted(3), bargaining(1),change(5), produced(2),consumed(1),reducing(3)	Adopted(Y),bargaining(Y),change(Y),produced(Y),consumed(Y), reducing(Y)	NA
Illinois Curator	5	0	5	Adopted(1), change(0),produced(1), consumed(1), reducing(0)	Adopted(2),change(4), produced(3),consumed(3), reducing(2)	Adopted(Y),change(Y),produced(Y),consumed(Y),reducing(Y)	NA
Boxer (DRT)	5	1	NA	Adopted(3), change(0), produced(1), consumed(1), reducing(0)	NA	NA	NA
Stanford NLP	1	1	3	reducing(1)	NA	NA	NA
Manual annotation	Adopted, bargaining, change, produced, consumed, reducing	13 days(periodx-days), Paris suburb(location)	Agreement, nations(world's), energy, reliance, fuels(fossil), energy(cleaner forms)	Adopted(agreement, bargaining, 13 days, Paris suburb), bargaining(Paris suburb, intense), change(agreement, produced, consumed), produced(nations, energy), consumed(nations, energy), reducing(agreement, reliance, fuels)	Adopted(4), bargaining(3), change(3), produced(2), consumed(2), reducing(3)	NA	Adopted(8), bargaining(3), change(20), produced(7), consumed(6), reducing(22)

Table A.56: Experiment 1 - Sentence 56

- Not long after governments at the United Nations climate summit in Paris finally reached an agreement, friends and colleagues concerned about the rights and welfare of animals began to bombard me with messages.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	4	Reached(0), concerned(1)	Reached(1), concerned(2)	Reached(), concerned()	NA
SHALMANESER	2	0	3	Reached(1), concerned(1)	Reached(2), concerned(1)	Reached(), concerned()	NA
LTH SRL	3	1	2	Reached(1), concerned(0), bombard(2)	Reached(3), concerned(2), bombard(3)	Reached(), concerned(), bombard()	NA
Illinois Curator	3	1	4	Reached(0), concerned(0), bombard(1)	Reached(3), concerned(2), bombard(2)	Reached(), concerned(), bombard()	NA
Boxer (DRT)	3	1	NA	Reached(2), concerned(4), bombard(2)	NA	NA	NA
Stanford NLP	0	1	0		NA	NA	NA
Manual annotation	Reached, concerned, bombard	United Nations climate summit(event), Paris(location)	Governments, agreement, friends, colleagues, rights(animals), welfare(animals)	Reached(United Nations climate summit, Paris, agreement), concerned(friends, colleagues, rights, welfare), bombard(friends, colleagues, me, messages)	Reached(3), concerned(4), bombard(4)	NA	Reached(9), concerned(5), bombard(5)

Table A.57: Experiment 1 - Sentence 57

- It may seem strange, but such absences are a fact of the unwieldy process that leads to international agreements like the one hashed out in Paris last Saturday.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	5	Seem(0), leads(2)	Seem(0), leads(2)	Seem(Y), leads(Y)	NA
SHALMANESER	2	0	7	Seem(2), leads(0)	Seem(2), leads(1)	Seem(Y), leads(Y)	NA
LTH SRL	3	2	4	Seem(2), leads(1), hashed out(2)	Seem(4), leads(3), hashed out(3)	Seem(Y), leads(Y), hashed out(Y)	NA
Illinois Curator	3	2	6	Seem(2), leads(1), hashed out(2)	Seem(3), leads(4), hashed out(3)	Seem(Y), leads(Y), hashed out(Y)	NA
Boxer (DRT)	3	2	NA	Seem(0), leads(2), hashed out(3)	NA	NA	NA
Stanford NLP	1	2	3	Seem(2)	NA	NA	NA
Manual annotation	Seem, leads, hashed out	Paris(location), Saturday(date-day_of_week)	It, Strange, Absences, fact, process(unwieldy), agreements(international), one	Seem(it, strange), leads(process, agreements), hashed out(agreements, Paris, Saturday)	Seem(2), leads(2), hashed out(3)	NA	Seem(4), leads(31), hashed out(1)

Table A.58: Experiment 1 - Sentence 58

- More than 190 countries recently signed the Paris climate agreement, which urged constructive steps toward keeping global temperature increases below 2 degrees Celsius, but it's unclear whether the framework will lead to more aggressive pollution controls.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	6	1	4	Signed(0), urged(0), keeping(2), increases(0), lead(1), controls(1)	Signed(3), urged(0), keeping(1), increases(0), lead(2), controls(2)	Signed(Y), urged(Y), keeping(Y), increases(Y), lead(Y), controls(Y)	NA
SHALMANESER	4	1	2	Signed(1), urged(2), increases(0), lead(0)	Signed(2), urged(4), increases(1), lead(1)	Signed(Y),urged(Y),increases(Y), lead(Y)	NA
LTH SRL	6	0	3	Signed(1), urged(0), keeping(0), increases(1), lead(2), controls(1)	Signed(3), urged(4), keeping(2), increases(3), lead(3), controls(2)	Signed(Y),urged(Y),keeping(Y), increases(Y), lead(Y), controls(Y)	NA
Illinois Curator	4	0	4	Signed(1), urged(0), keeping(1), lead(1)	Signed(3), urged(3), keeping(1), lead(3)	Signed(Y), urged(Y), keeping(Y), lead(Y)	NA
Boxer (DRT)	4	1	NA	Signed(1), urged(0), keeping(2), lead(1)	NA	NA	NA
Stanford NLP	1	0	2	lead(2)	NA	NA	NA
Manual annotation	Signed, urged, keeping, increases, lead, controls	190 countries(countx), Paris climate agreement(product-other), 2 degrees Celsius(numex-measurement)	Steps(constructive), temperature(global), framework, pollution(aggressive)	Signed(190 countries, Paris climate agreement, urged), urged(190 countries, steps, keeping), keeping(temperature, increases), increases(below(2 degrees Celsius)), lead(framework, controls), controls(pollution)	Signed(3), urged(3), keeping(2), increases(1), lead(2), controls(1)	NA	Signed(10), urged(3), keeping(25), increases(7), lead(31), controls(20)

Table A.59: Experiment 1 - Sentence 59

- The agreement sets a goal of keeping warming well below 2 degrees Celsius, or 3.6 degrees Fahrenheit, and, for the first time, agrees to pursue efforts to limit the increase in temperatures to 1.5 degrees Celsius, 2.7 degrees Fahrenheit.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	5	Sets(1), keeping(1), pursue(1), increase (1)	Sets(1), keeping(1), pursue(2), increase (1)	Sets(Y),keeping(Y),warming(Y), pursue(Y), limit(Y),increase (Y)	NA
SHALMANESER	3	3	4	Sets(1), pursue(0), increase (1)	Sets(2), pursue(1), increase (1)	Sets(Y),pursue (Y), increase (Y)	NA
LTH SRL	6	0	4	Sets(1), keeping(0), warming(0), pursue(0), limit(0), increase (1)	Sets(2), keeping(2), warming(1), pursue(1), limit(1), increase (3)	Sets(Y),keeping(Y),warming(Y), pursue(Y), limit(Y), increase (Y)	NA
Illinois Curator	4	0	5	Sets(1), keeping(0), pursue(0), limit(0)	Sets(2), keeping(1), pursue(2), limit(2)	Sets(Y),keeping(Y),warming(Y), pursue(Y), limit(Y),increase (Y)	NA
Boxer (DRT)	4	3	NA	Sets(3), keeping(0), pursue(1), limit(1)	NA	NA	NA
Stanford NLP	1	0	2	Sets(2)	NA	NA	NA
Manual annotation	Sets, keeping, warming, pursue, limit, increase	2 degrees Celsius(numex-measurement), 3.6 degrees Fahrenheit(numex-measurement), 1.5 degrees Celsius(numex-measurement), 2.7 degrees Fahrenheit(numex-measurement)	Agreement, goal, time(first), efforts, temperatures	Sets(agreement, goal, keeping), keeping(warming), warming(below(2 degrees Celsius), below(3.6 degrees Fahrenheit), pursue(efforts, limit), limit(increase), increase (temperatures)	Sets(3), keeping(1), warming(2), pursue(3), limit(1), increase (1)	NA	Sets(38), keeping(25), warming(6), pursue(4), limit(9), increase (7)

Table A.60: Experiment 1 - Sentence 60

- The uproar over the sentence, fueled in part by the victim's harrowing letter in which she detailed the assault in graphic terms, is part of growing outrage over sexual assault on U.S. college campuses.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship – Count	WSD - Ident.	WSD - Count
SEMAFOR	5	0	5	fueled(0), detailed(1), assault(1), growing(0), assault(1)	fueled(2), detailed(2), assault(1), growing(1), assault(1)	fueled(Y), detailed(Y), assault(Y), growing(Y), assault(Y)	NA
SHALMANESER	3	0	4	assault(1), growing(1), assault(1)	assault(3), growing(2), assault(1)	assault(Y), growing(Y), assault(Y)	NA
LTH SRL	6	1	4	Uproar(0), fueled(0), detailed(2), assault(1), growing(0), assault(1)	Uproar(1), fueled(4), detailed(4), assault(1), growing(1), assault(2)	Uproar(Y), fueled(Y), detailed(Y), assault(Y), growing(Y), assault(Y)	NA
Illinois Curator	2	0	6	fueled(0), growing(1)	fueled(4), growing(1)	fueled(Y), growing(Y)	NA
Boxer (DRT)	3	1	NA	fueled(0), detailed(2), growing(2)	NA	NA	NA
Stanford NLP	2	1	2	detailed(1), assault(2)	NA	NA	NA
Manual annotation	Uproar, fueled, detailed, assault, growing, assault	U.S. (location)	Sentence, letter(victims, harrowing),she, terms(graphic), outrage, campuses(college)	Uproar(sentence), fueled(detailed(letter)), detailed(assault, terms), assault(she, terms), growing(outrage, assault), assault(sexual,U.S. campuses)	Uproar(1), fueled(1), detailed(2), assault(1), growing(2), assault(2)	NA	Uproar(2), fueled(5), detailed(3), assault(7), growing(13), assault(7)

Table A.61: Experiment 1 - Sentence 61

- The judge handed down a six-month sentence and three years probation to Mr. Turner, a champion swimmer convicted in March of attacking the 23-year-old woman behind a dumpster on campus in 2015.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	3	4	Handed(1), convicted(0), attacking(1)	Handed(1), convicted(0), attacking(1)	Handed(Y), convicted(Y), attacking(Y)	NA
SHALMANESER	3	2	7	Handed(1), convicted(1), attacking(3)	Handed(3), convicted(4), attacking(5)	Handed(Y), convicted(Y), attacking(Y)	NA
LTH SRL	3	3	6	Handed(1), convicted(2), attacking(4)	Handed(3), convicted(3), attacking(5)	Handed(Y), convicted(Y), attacking(Y)	NA
Illinois Curator	3	1	7	Handed(1), convicted(2), attacking(1)	Handed(3), convicted(4), attacking(2)	Handed(Y), convicted(Y), attacking(Y)	NA
Boxer (DRT)	3	5	NA	Handed(2), convicted(3), attacking(4)	NA	NA	NA
Stanford NLP	1	3	3	convicted(2)	NA	NA	NA
Manual annotation	Handed, convicted, attacking	Six-month(periodx-month), three years (periodx-years), Mr. Turner(person), March (date-month), 23-year-old(person), 2015(year)	Judge, sentence(six-month), probation(three years), swimmer(champion), woman(23-year-old), dumpster(behind), campus	Handed(judge, six-month sentence, probation, Mr. Turner), convicted(swimmer, March, attacking), attacking(swimmer, woman, dumpster, 2015)	Handed(4), convicted(3), attacking(4)	NA	Handed(3), convicted(1), attacking(7)

Table A.62: Experiment 1 - Sentence 62

- Recall proponents need to file a notice of intent to recall, then collect valid signatures from voters equaling at least 20% of the ballots cast in Persky's last election.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	5	File(0), recall(0), collect(0), election(1)	File(1), recall(0), collect(0), election(1)	File(Y), recall(Y), collect(Y), election(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	4	0	4	File(1), recall(0), collect(1), election(2)	File(2), recall(0), collect(3), election(2)	File(Y), recall(Y), collect(Y), election(Y)	NA
Illinois Curator	3	0	5	File(1), recall(0), collect(1)	File(1), recall(3), collect(2)	File(Y), recall(Y), collect(Y)	NA
Boxer (DRT)	3	0	NA	File(1), recall(1), collect(0)	NA	NA	NA
Stanford NLP	1	1	2	File(2)	NA	NA	NA
Manual annotation	File, recall, collect, election	20%(numex-percent), Persky(person)	Proponents(recall), notice(intent), signatures(valid), voters, ballots(cast)	File(proponents, notice(recall)), recall(notice), collect(signatures, voters, 20%(ballots, election(last))), election(Persky, last)	File(2), recall(1), collect(3), election(2)	NA	File(9), recall(12), collect(8), election(4)

Table A.63: Experiment 1 - Sentence 63

- The California Assembly could impeach him, after which he could be removed from office on a two-thirds vote in the state Senate.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	0	3	removed(2)	removed(2)	removed(Y)	NA
SHALMANESER	1	0	2	removed(2)	removed(2)	removed(Y)	NA
LTH SRL	2	0	4	Impeach(2), removed(2)	Impeach(5), removed(4)	Impeach(Y), removed(Y)	NA
Illinois Curator	2	0	5	Impeach(2), removed(2)	Impeach(4), removed(5)	Impeach(Y), removed(Y)	NA
Boxer (DRT)	2	2	NA	Impeach(1), removed(1)	NA	NA	NA
Stanford NLP	1	1	3	Impeach(2)	NA	NA	NA
Manual annotation	Impeach, removed	California Assembly(org), two-thirds (numex-measurement)	Him, he, office, vote(two-thirds), Senate(state)	Impeach(California Assembly, him), removed(he, office)	Impeach(2), removed(2)	NA	Impeach(3), removed(10)

Table A.64: Experiment 1 - Sentence 64

- Widespread outrage has erupted over a California judge's decision to give a former Stanford University swimmer a six-month jail sentence for sexually assaulting an unconscious woman.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	5	Erupted(1), decision(1), give(1), assaulting(1)	Erupted(1), decision(2), give(1), assaulting(2)	Erupted(Y),decision(Y), give(Y),assaulting(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	4	0	2	Erupted(0), decision(0), give(1), assaulting(1)	Erupted(2),decision(2), give(3), assaulting(2)	Erupted(Y),decision(Y),give(Y),assaulting(Y)	NA
Illinois Curator	3	1	5	Erupted(1), give(0), assaulting(1)	Erupted(2), give(3), assaulting(2)	Erupted(Y), give(Y), assaulting(Y)	NA
Boxer (DRT)	3	3	NA	Erupted(1), give(3), assaulting(1)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Erupted, decision, give, assaulting	California(location), Stanford University (org), six-month (periodx)	Outrage(widespread), judge(California), swimmer(Stanford University), sentence(jail, six-month), woman(unconscious)	Erupted(outrage, decision), decision(judge, give), give(swimmer, six-month, jail sentence), assaulting(swimmer, woman)	Erupted(2), decision(2), give(3), assaulting(2)	NA	Erupted(8), decision(5), give(45), assaulting(3)

Table A.65: Experiment 1 - Sentence 65

- Muhammad Ali, the three-time world heavyweight boxing champion who helped define his turbulent times as the most charismatic and controversial sports figure of the 20th century, died on Friday in a Phoenix-area hospital.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	2	2	Boxing(1), defined(1), died(1)	Boxing(2), defined(2), died(1)	Boxing(Y), defined(Y), died(Y)	NA
SHALMANESER	2	1	2	defined(0), died(1)	defined(2), died(2)	defined(Y), died(Y)	NA
LTH SRL	3	1	3	Boxing(0), defined(1), died(2)	Boxing(3), defined(2), died(3)	Boxing(Y), defined(Y), died(Y)	NA
Illinois Curator	3	1	3	Boxing(2), defined(1), died(2)	Boxing(2), defined(3), died(2)	Boxing(Y), defined(Y), died(Y)	NA
Boxer (DRT)	3	2	NA	Boxing(1), defined(2), died(3)	NA	NA	NA
Stanford NLP	1	2	1	died(2)	NA	NA	NA
Manual annotation	Boxing, defined, died	Muhammad Ali(person), 20th century(timex), Friday(timex-day_of_week), Phoenix-area(location)	Champion(boxing, heavyweight, three-time world), times(turbulent), figure(sports), charismatic, controversial), hospital	Boxing(Muhammad Ali, champion), defined(champion, times, figure, 20th century), died(Muhammad Ali, Friday, hospital(Phoenix-area))	Boxing(2), defined(4), died(3)	NA	Boxing(5), defined(7), died(11)

Table A.66: Experiment 1 - Sentence 66

- Earlier in the day, his coffin traveled through nearly 20 miles of Louisville, cheered and saluted by tens of thousands of people who tossed flowers onto the hearse and chanted his name.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	5	Traveled(1), cheered(0), tossed(2), chanted(1)	Traveled(2),cheered(2), tossed(2), chanted(1)	Traveled(Y),cheered(Y),saluted(Y),tossed(Y), chanted(Y)	NA
SHALMANESER	4	2	5	Traveled(3), cheered(0), tossed(2), chanted(1)	Traveled(2),cheered(0), tossed(2),chanted(1)	Traveled(Y),cheered(Y),tossed(Y),chanted(Y)	NA
LTH SRL	5	1	5	Traveled(2), cheered(0), saluted(0), tossed(3), chanted(2)	Traveled(3),cheered(1),saluted(2), tossed(4), chanted(3)	Traveled(Y), cheered(Y), saluted(Y), tossed(Y), chanted(Y)	NA
Illinois Curator	5	1	5	Traveled(2), cheered(0), saluted(0), tossed(3), chanted(1)	Traveled(3), cheered(2),saluted(1), tossed(5), chanted(2)	Traveled(Y),cheered(Y),saluted(Y),tossed(Y), chanted(Y)	NA
Boxer (DRT)	5	1	NA	Traveled(3), cheered(0), saluted(0), tossed(2), chanted(2)	NA	NA	NA
Stanford NLP	1	1	1	Traveled(2)	NA	NA	NA
Manual annotation	Traveled, cheered, saluted, tossed, chanted	Earlier in the day(timex), 20 miles(numex-measurement), Louisville(location)	coffin, people, flowers, hearse, name	Traveled(coffin, 20 miles, Louisville, earlier in the day), cheered(people(tens of thousands)), saluted(people(tens of thousands)), tossed(people(tens of thousands), flowers, hearse), chanted(people(tens of thousands), name)	Traveled(4), cheered(1), saluted(1), tossed(3), chanted(2)	NA	Traveled(8), cheered(5), saluted(6), tossed(6), chanted(3)

Table A.67: Experiment 1 - Sentence 67

- While touching tributes to Ali were pouring in from world leaders, fellow athletes and just regular folk, the boxing great had already addressed how he wanted the world to think about him after his death.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	0	4	Pouring(0), addressed(0), think(2), death(0)	Pouring(0), addressed(2), think(3), death(1)	Pouring(Y), addressed(Y), think(Y), death(Y)	NA
SHALMANESER	4	0	2	Pouring(0), addressed(0), think(2), death(0)	Pouring(1), addressed(0), think(2), death(1)	Pouring(Y), addressed(Y), think(Y), death(Y)	NA
LTH SRL	4	0	2	Pouring(0), addressed(1), think(1), death(0)	Pouring(3), addressed(4), think(4), death(1)	Pouring(Y), addressed(Y), think(Y), death(Y)	NA
Illinois Curator	3	0	4	Pouring(0), addressed(0), think(2)	Pouring(2), addressed(3), think(3)	Pouring(Y), addressed(Y), think(Y)	NA
Boxer (DRT)	3	1	NA	Pouring(3), addressed(1), think(1)	NA	NA	NA
Stanford NLP	1	0	2	think(2)	NA	NA	NA
Manual annotation	Pouring, addressed, think, death	Ali(person), boxing great(person)	Tributes(touching), leaders(world), athletes(fellow, Ali), folk(regular), world, him	Pouring(tributes(Ali), leaders, athletes, folk), addressed(boxing great, world, think), think(world,him), death(boxing great)	Pouring(4), addressed(3), think(2), death(2)	NA	Pouring(7), addressed(11), think(14), death(8)

Table A.68: Experiment 1 - Sentence 68

- Chuckling and declaring allegiance to the Islamic State, he opened fire at a gay and Latino nightclub here, leaving 49 people dead and wounding 53 others before he was killed by the police to end a protracted standoff.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	6	0	3	Chuckling(0), declaring(0), opened(2),leaving(0), wounding(0), killed(1)	Chuckling(0),declaring(1),opened(2),leaving(1), wounding(0), killed(2)	Chuckling(Y),declaring(Y),opened(Y),leaving(Y),wounding(Y), killed(Y)	NA
SHALMANESER	5	1	3	Chuckling(1), declaring(0), opened(1),leaving(0), killed(1)	Chuckling(2), declaring(1), opened(1),leaving(1), killed(2)	Chuckling(Y),declaring(Y),opened(Y),leaving(Y), killed(Y)	NA
LTH SRL	6	0	3	Chuckling(1), declaring(1), opened(2),leaving(1), wounding(1), killed(1)	Chuckling(2), declaring(1), opened(5),leaving(3), wounding(3), killed(1)	Chuckling(Y), declaring(Y), opened(),leaving(Y), wounding(Y), killed(Y)	NA
Illinois Curator	6	0	6	Chuckling(1), declaring(1), opened(2),leaving(0), wounding(0), killed(1)	Chuckling(2), declaring(1), opened(5),leaving(2), wounding(1), killed(2)	Chuckling(Y), declaring(Y), opened(),leaving(Y),wounding(Y), killed(Y)	NA
Boxer (DRT)	6	1	NA	Chuckling(0), declaring(1), opened(2),leaving(1), wounding(), killed(1)	NA	NA	NA
Stanford NLP	1	1	3	opened(2)	NA	NA	NA
Manual annotation	Chuckling, declaring, opened, leaving, wounding, killed	Islamic State(org), Latino(org-ethnic_group), 49 people(countx-N_persons), 53 others(countx-N_persons)	Allegiance, fire, nightclub(Latino), he, police, standoff	Chuckling(he), declaring(he, allegiance, Islamic State), opened(he, fire, nightclub(Latino)), leaving(49 people, wounding), wounding(he, 53 others), killed(he, police)	Chuckling(1), declaring(3), opened(3),leaving(2), wounding(3), killed(2)	NA	Chuckling(1), declaring(8), opened(14),leaving(15), wounding(4), killed(15)

Table A.69: Experiment 1 - Sentence 69

- But his professed embrace of the Islamic State and its call for disaffected Muslims to attack the West seem to have come suddenly, as if something snapped.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	3	Embrace(0), call(0), attack(1), snapped(1)	Embrace(0), call(2), attack(1), snapped(1)	Embrace(Y), call(Y), attack(Y), snapped(Y)	NA
SHALMANESER	3	0	2	call(0), attack(2), snapped(0)	call(2), attack(2), snapped(0)	call(Y), attack(Y), snapped(Y)	NA
LTH SRL	4	0	2	Embrace(1), call(0), attack(2), snapped(1)	Embrace(2), call(2), attack(2), snapped(1)	Embrace(Y), call(Y), attack(Y), snapped(Y)	NA
Illinois Curator	3	0	2	Embrace(1), attack(1), snapped(1)	Embrace(2), attack(2), snapped(1)	Embrace(Y), attack(Y), snapped(Y)	NA
Boxer (DRT)	1	1	NA	Attack(2)	NA	NA	NA
Stanford NLP	0	1	0		NA	NA	NA
Manual annotation	Embrace, call, attack, snapped	Islamic State(org), Muslims(org-ethnic_group), the West(org)	His, embrace, call, something	Embrace(his, Islamic State, call), call(Islamic State, Muslims(disaffected), attack), attack(Muslims(disaffected), the West), snapped(something)	Embrace(3), call(3), attack(2), snapped(1)	NA	Embrace(6), call(41), attack(15), snapped(13)

Table A.70: Experiment 1 - Sentence 70

- A first-generation American, he was born in New York City's melting-pot borough of Queens in 1986, and moved about four years later with his Afghan parents to Port St. Lucie in Florida, where he was quickly enrolled in an English for Speakers of Other Languages program.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	5	3	Born(1), moved(1), enrolled(0)	Born(2), moved(1), enrolled(2)	Born(Y), moved(Y), enrolled(Y)	NA
SHALMANESER	3	1	3	Born(1), moved(0), enrolled(3)	Born(2), moved(0), enrolled(5)	Born(Y), moved(Y), enrolled(Y)	NA
LTH SRL	3	2	2	Born(2), moved(2), enrolled(1)	Born(4), moved(4), enrolled(5)	Born(Y), moved(Y), enrolled(Y)	NA
Illinois Curator	3	3	2	Born(1), moved(2), enrolled(0)	Born(3), moved(4), enrolled(4)	Born(Y), moved(Y), enrolled(Y)	NA
Boxer (DRT)	3	4	NA	Born(2), moved(2), enrolled(1)	NA	NA	NA
Stanford NLP	1	5	3	Born(3)	NA	NA	NA
Manual annotation	Born, moved, enrolled	American(org-ethnic_group), New York City(location), Queens(location), 1986(year), four years(periodx), Afghan(org-ethnic_group), Port St. Lucie(location), Florida(location), English for Speakers(product-academic), Other Languages program(product-academic)	First-generation, borough(melting -pot), parents(Afghan)	Born(American(first-generation), Queens, 1986), moved(American(first-generation), four years(born, later), parents(Afghan), Port St. Lucie), enrolled(American(first-generation, English for Speakers, Port St. Lucie)	Born(3), moved(4), enrolled(3)	NA	Born(16), moved(17), enrolled(2)

Table A.71: Experiment 1 - Sentence 71

- There has been no claim of responsibility for the attack on jihadi forums, but ISIS sympathizers have reacted by praising the attack on pro-Islamic State forums.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	5	0	4	Claim(0), attack(1), reacted(2), praising(0), attack(1)	Claim(0), attack(1), reacted(2), praising(0), attack(1)	Claim(Y), attack(Y), reacted(Y), praising(Y), attack(Y)	NA
SHALMANESER	4	0	2	Claim(0), attack(1), praising(0), attack(1)	Claim(0), attack(1), praising(1), attack(1)	Claim(Y), attack(Y), praising(Y), attack(Y)	NA
LTH SRL	5	0	4	Claim(2), attack(1), reacted(1), praising(0), attack(1)	Claim(3), attack(1), reacted(2), praising(2), attack(1)	Claim(Y), attack(Y), reacted(Y), praising(Y), attack(Y)	NA
Illinois Curator	2	0	4	reacted(1), praising(0)	reacted(3), praising(2)	reacted(Y), praising(Y)	NA
Boxer (DRT)	2	1	NA	reacted(1), praising(1)	NA	NA	NA
Stanford NLP	1	1	1	praising(1)	NA	NA	NA
Manual annotation	Claim, attack, reacted, praising, attack	Jihadi(org-ethnic_group), ISIS(org), pro-Islamic State(org)	Responsibility, forums(jihadi) , sympathizers(ISIS), forums(pro_Islamic State)	Claim(no, responsibility, attack), attack(forums(jihadi)), reacted(sympathizers, praising, forums(pro-Islamic State)), praising(attack), attack(pro-Islamic State forums)	Claim(3), attack(1), reacted(3), praising(1), attack(1)	NA	Claim(11), attack(15), reacted(3), praising(2), attack(15)

Table A.72: Experiment 1 - Sentence 72

- She said the language is inconsistent with previous ISIS announcements and that the Arabic word for gay was used rather than an epithet normally used by ISIS.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	3	Said(1), used(1), used(1)	Said(2), used(1), used(1)	Said(Y), used(Y), used(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	3	0	2	Said(1), used(1), used(1)	Said(2), used(2), used(3)	Said(Y), used(Y), used(Y)	NA
Illinois Curator	3	0	4	Said(1), used(1), used(1)	Said(2), used(1), used(2)	Said(Y), used(Y), used(Y)	NA
Boxer (DRT)	3	0	NA	Said(1), used(1), used(2)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Said, used, used	ISIS(org), Arabic(product-language), ISIS(org)	She, language, announcements, word, epithet	Said(she, announcements(ISIS, previous), language, inconsistent), used(word(Arabic, gay)),used(normally, ISIS)	Said(3), used(1), used(2)	NA	Said(12), used(9), used(9)

Table A.73: Experiment 1 - Sentence 73

- The software giant said Monday morning that it would acquire LinkedIn in a $26.2 billion cash deal.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	2	Said(1), acquire(1)	Said(2), acquire(2)	Said(Y), acquire(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	2	0	1	Said(1), acquire(1)	Said(3), acquire(4)	Said(), acquire(Y)	NA
Illinois Curator	2	0	2	Said(1), acquire(0)	Said(2), acquire(4)	Said(), acquire(Y)	NA
Boxer (DRT)	2	4	NA	Said(2), acquire(2)	NA	NA	NA
Stanford NLP	0	3	0		NA	NA	NA
Manual annotation	Said, acquire	Monday(date-day_of_week),morning(time), LinkedIn(org), $26.2 billion cash(money)	Giant(software), deal	Said(software giant, acquire, Monday morning), acquire(software giant, LinkedIn, $26.2 billion cash, deal)	Said(4), acquire(4)	NA	Said(12), acquire(7)

Table A.74: Experiment 1 - Sentence 74

- In 2014, it paid nearly $9.4 billion for the smartphone operations of Nokia and some years earlier spent more than $6 billion for aQuantive, an internet advertising company, but ended up writing off most of the value of those deals after they performed poorly.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	3	4	Paid(2), spent(0), writing(0), performed(1)	Paid(3), spent(0), writing(1), performed(2)	Paid(Y), spent(Y), writing(Y), performed(Y)	NA
SHALMANESER	2	2	3	Paid(1), writing(0)	Paid(2), writing(3)	Paid(Y), spent(Y), writing(Y), performed(Y)	NA
LTH SRL	4	1	4	Paid(1), spent(2), writing(0), performed(1)	Paid(3), spent(4), writing(2), performed(2)	Paid(Y), spent(Y), writing(Y), performed(Y)	NA
Illinois Curator	4	1	4	Paid(3), spent(2), writing(1), performed(1)	Paid(4), spent(2), writing(2), performed(2)	Paid(Y), spent(Y), writing(Y), performed(Y)	NA
Boxer (DRT)	4	4	NA	Paid(3), spent(2), writing(1), performed(1)	NA	NA	NA
Stanford NLP	2	5	0	Paid(1), spent(1)	NA	NA	NA
Manual annotation	Paid, spent, writing, performed	2014(year), $9.4 billion(money), Nokia(org), $6 billion(money), aQuantive(org)	Operations(smartp hone), years(some, earlier), company(internet advertising), value(most of), deals	Paid(2014, $9.4 billion, Nokia(operations)), spent(years, $6 billion, aQuantive(company)), writing(aQuantive(company), deals, performed), performed(poorly)	Paid(3), spent(3), writing(3), performed(1)	NA	Paid(14), spent(5), writing(13), performed(4)

Table A.75: Experiment 1 - Sentence 75

- LinkedIn CEO Jeff Weiner will remain CEO of the social network for professionals, reporting directly to Microsoft CEO Satya Nadella.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	1	Remain(2), reporting(0)	Remain(2), reporting(0)	Remain(Y), reporting(Y)	NA
SHALMANESER	2	0	0	Remain(1), reporting(2)	Remain(3), reporting(2)	Remain(Y), reporting(Y)	NA
LTH SRL	2	0	0	Remain(2), reporting(2)	Remain(6), reporting(3)	Remain(Y), reporting(Y)	NA
Illinois Curator	2	0	1	Remain(1), reporting(1)	Remain(3), reporting(3)	Remain(Y), reporting(Y)	NA
Boxer (DRT)	2	4	NA	Remain(2), reporting(1)	NA	NA	NA
Stanford NLP	1	3	1	Remain(2)	NA	NA	NA
Manual annotation	Remain, reporting	LinkedIn(org), CEO(product-position_title), Jeff Weiner(person), Microsoft(org), Satya Nadella(person)	Network(social,professionals)	Remain(Jeff Weiner, CEO(network)), reporting(Jeff Weiner, Satya Nadella)	Remain(2), reporting(2)	NA	Remain(4), reporting(7)

Table A.76: Experiment 1 - Sentence 76

- Mr. Nadella said today's work is split between tools workers use to get their jobs done, such as Microsoft's Office programs, and professional networks that connect workers.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	5	Said(1), use(1), connect(1)	Said(2), use(2), connect(1)	Said(Y), use(Y), connect(Y)	NA
SHALMANESER	2	0	3	Said(1), use(0)	Said(1), use(0)	Said(Y), use(Y)	NA
LTH SRL	3	1	5	Said(1), use(1), connect(2)	Said(2), use(2), connect(3)	Said(Y), use(Y), connect(Y)	NA
Illinois Curator	3	0	4	Said(1), use(1), connect(2)	Said(2), use(3), connect(3)	Said(Y), use(Y), connect(Y)	NA
Boxer (DRT)	2	0	NA	Said(0), use(1)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Said, use, connect	Mr. Nadella(person), today(date), Microsoft's Office programs(product)	Work(today's), workers(tools), jobs, networks(professional), workers	Said(Mr. Nadella, work, use, connect), use(workers, jobs), connect(networks, workers)	Said(4), use(2), connect(2)	NA	Said(12), use(13), connect(11)

Table A.77: Experiment 1 - Sentence 77

- The deal is Chief Executive Satya Nadella's latest effort to revitalize Microsoft, which was viewed not long ago as left behind by shifts in technology.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	0	1	Deal(0), revitalize(1), viewed(0), shifts(0)	Deal(0), revitalize(1), viewed(3), shifts(0)	Deal(Y), revitalize(Y), viewed(Y), shifts(Y)	NA
SHALMANESER	2	1	1	viewed(0), shifts(0)	viewed(0), shifts(0)	viewed(Y), shifts(Y)	NA
LTH SRL	3	0	1	revitalize(0), viewed(0), shifts(1)	revitalize(1), viewed(3), shifts(1)	revitalize(Y), viewed(Y), shifts(Y)	NA
Illinois Curator	2	0	2	revitalize(0), viewed(0)	revitalize(1), viewed(3)	revitalize(Y), viewed(Y)	NA
Boxer (DRT)	2	2	NA	revitalize(1), viewed(1)	NA	NA	NA
Stanford NLP	0	2	1		NA	NA	NA
Manual annotation	Deal, revitalize, viewed, shifts	Chief Executive(product-position_title), Satya Nadella(person), Microsoft(org)	Deal, Effort(latest), technology	Deal(Satya Nadella, effort, revitalize), revitalize(Microsoft), viewed(revitalize, shifts), shifts(left behind, technology)	Deal(3), revitalize(1), viewed(2), shifts(3)	NA	Deal(22), revitalize(2), viewed(3), shifts(23)

Table A.78: Experiment 1 - Sentence 78

- Mr. Nadella hopes the deal will open new horizons for Microsoft's Office suite as well as LinkedIn, both of which have saturated their markets, and generally bolster Microsoft's revenue and competitive position.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	0	4	open(2), saturated(0), bolster(0)	open(2), saturated(0), bolster(2)	open(Y), saturated(Y), bolster(Y)	NA
SHALMANESER	2	0	2	open(1), saturated(1)	open(3), saturated(3)	open(Y), saturated(Y)	NA
LTH SRL	3	0	2	open(1), saturated(1), bolster(0)	open(3), saturated(2), bolster(4)	open(Y), saturated(Y), bolster(Y)	NA
Illinois Curator	3	0	5	open(2), saturated(1), bolster(0)	open(4), saturated(2), bolster(1)	open(Y), saturated(Y), bolster(Y)	NA
Boxer (DRT)	3	2	NA	open(2), saturated(3), bolster(0)	NA	NA	NA
Stanford NLP	2	2	3	open(2), bolster(1)	NA	NA	NA
Manual annotation	open, saturated, bolster	Mr. Nadella(person), Microsoft office suite(product), LinkedIn(org), Microsoft(org)	Deal, Horizons(new), markets, revenue, position(competitiv e)	open(deal, horizons, Microsoft Office suite, LinkedIn), saturated(Microsoft, LinkedIn, markets), bolster(revenue(Microsoft), position)	open(4), saturated(3), bolster(2)	NA	open(36), saturated(5), bolster(4)

Table A.79: Experiment 1 - Sentence 79

- Microsoft will pay $196 per LinkedIn share, a 50% premium to the social network's closing price on Friday.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	0	1	Pay(1)	Pay(2)	Pay(Y)	NA
SHALMANESER	1	0	1	Pay(1)	Pay(2)	Pay(Y)	NA
LTH SRL	1	1	1	Pay(1)	Pay(6)	Pay(Y)	NA
Illinois Curator	1	0	2	Pay(1)	Pay(3)	Pay(Y)	NA
Boxer (DRT)	1	3	NA	Pay(3)	NA	NA	NA
Stanford NLP	1	3	1	Pay(2)	NA	NA	NA
Manual annotation	Pay	Microsoft(org), $196(money), LinkedIn(org), 50% premium(numex-measurement), Friday(date-day_of_week)	Share(linkedIn), price(social network, closing)	Pay(Microsoft, LinkedIn,$196, share, Friday)	Pay(5)	NA	Pay(12)

Table A.80: Experiment 1 - Sentence 80

- The 50 largest U.S. corporations currently stash about $1.4 trillion in offshore tax havens, according to the analysis by anti-poverty group Oxfam America.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	1	Stash(1), analysis(0)	Stash(3), analysis(0)	Stash(Y), analysis(Y)	NA
SHALMANESER	2	0	1	Stash(2), analysis(1)	Stash(3), analysis(1)	Stash(Y), analysis(Y)	NA
LTH SRL	2	0	1	Stash(1), analysis(1)	Stash(4), analysis(1)	Stash(Y), analysis(Y)	NA
Illinois Curator	1	0	2	Stash(3)	Stash(4)	Stash(Y)	NA
Boxer (DRT)	1	1	NA	Stash(1)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Stash, analysis	50 largest U.S. corporations(org), $1.4 trillion(money), Oxfam America(org)	Havens(offshore, tax), group(anti-poverty)	Stash(50 largest U.S. corporations, $1.4 trillion, havens), analysis(group(Oxfam America))	Stash(3), analysis(1)	NA	Stash(2), analysis(6)

Table A.81: Experiment 1 - Sentence 81

- A March study from the liberal Citizens for Tax Justice found that the 500 largest corporations are holding $2.4 trillion overseas, allowing them to avoid paying $695 billion in taxes.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	3	0	Found(0), holding(2), allowing(1), paying(0)	Found(2), holding(2), allowing(3), paying(2)	Found(Y), holding(Y), allowing(Y), paying(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	4	0	2	Found(0), holding(3), allowing(0), paying(0)	Found(2), holding(4), allowing(3), paying(2)	Found(Y), holding(Y), allowing(Y), paying(Y)	NA
Illinois Curator	4	0	4	Found(0), holding(3), allowing(0), paying(0)	Found(2), holding(3), allowing(3), paying(2)	Found(Y), holding(Y), allowing(Y), paying(Y)	NA
Boxer (DRT)	4	2	NA	Found(0), holding(3), allowing(1), paying(1)	NA	NA	NA
Stanford NLP	2	3	0	holding(2), allowing(1)	NA	NA	NA
Manual annotation	Found, holding, allowing, paying	March study(timex-other), Citizens for Tax Justice(org), 500 largest corporations(countx-N_org), $2.4 trillion(money), $695 billion(money)	(liberal), taxes($695 billion), study(Citizens for Tax Justice)	Found(March study, 500 largest corporations, holding), holding(500 largest corporations, $2.4 trillion, overseas), allowing(them, paying), paying(avoid, taxes)	Found(3), holding(3), allowing(2), paying(2)	NA	Found(21), holding(38), allowing(10), paying(13)

Table A.82: Experiment 1 - Sentence 82

- The release of the Panama Papers, a leaked cache of documents from a Panamanian law firm, made waves earlier this month by exposing the elaborate tax-dodging schemes of the super-rich.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	1	Release(0), waves(0), exposing(0)	Release(1), waves(0), exposing(2)	Release(Y), waves(N), exposing(Y)	NA
SHALMANESER	1	0	1	waves(0)	waves(0)	waves(Y)	NA
LTH SRL	2	0	2	Release(0), exposing(0)	Release(1), exposing(2)	Release(Y), exposing(Y)	NA
Illinois Curator	1	0	4	exposing(0)	exposing(1)	exposing(Y)	NA
Boxer (DRT)	1	0	NA	exposing(2)	NA	NA	NA
Stanford NLP	1	1	2	exposing(0)	NA	NA	NA
Manual annotation	Release, waves, exposing	Panama Papers(product-other), Panamanian law firm(org)	Cache(leaked), document, waves, month, law firm, schemes(tax-dodging, elaborate), super-rich	Release(Panama Papers, waves), waves(month(earlier), exposing), exposing(super-rich, schemes(elaborate, tax-dodging)	Release(2), waves(2), exposing(2)	NA	Release(22), waves(14), exposing(9)

Table A.83: Experiment 1 - Sentence 83

- President Barack Obama cited the Panama Papers while touting a new Treasury Department rule aimed at curbing corporate inversions.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	1	Cited(2), touting(0)	Cited(2), touting(1)	Cited(Y), touting(Y)	NA
SHALMANESER	1	1	0	Cited(2)	Cited(2)	Cited(Y)	NA
LTH SRL	3	0	2	Cited(2), touting(0), curbing(1)	Cited(3), touting(2), curbing(2)	Cited(Y), touting(Y), curbing(Y)	NA
Illinois Curator	3	0	2	Cited(2), touting(1), curbing(1)	Cited(2), touting(1), curbing(2)	Cited(Y), touting(Y), curbing(Y)	NA
Boxer (DRT)	3	2	NA	Cited(1), touting(0), curbing(0)	NA	NA	NA
Stanford NLP	2	3	1	Cited(2), touting(0)	NA	NA	NA
Manual annotation	Cited, touting, curbing	Barack Obama(person), Panama Papers(product-other), Treasury Department(org)	Rule(Treasury Department, new), inversions(corporat e)	Cited(Barack Obama, Panama Papers), touting(rule, curbing), curbing(inversions)	Cited(2), touting(2), curbing(1)	NA	Cited(7), touting(2), curbing(4)

Table A.84: Experiment 1 - Sentence 84

- The story began back in February 2015, with an article in Süddeutsche Zeitung that revealed the German newspaper had a slug of secret files about offshore companies on the books of the Panamanian law firm Mossack Fonseca.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	5	Began(2), revealed(0)	Began(2), revealed(1)	Began(Y), revealed(Y)	NA
SHALMANESER	2	1	5	Began(2), revealed(0)	Began(2), revealed(1)	Began(Y), revealed(Y)	NA
LTH SRL	2	1	2	Began(2), revealed(0)	Began(4), revealed(3)	Began(Y), revealed(Y)	NA
Illinois Curator	2	1	4	Began(2), revealed(1)	Began(3), revealed(4)	Began(Y), revealed(Y)	NA
Boxer (DRT)	2	1	NA	Began(2), revealed(0)	NA	NA	NA
Stanford NLP	1	2	2	Began(2)	NA	NA	NA
Manual annotation	Began, revealed	February 2015 (date), Suddeutsche Zeitung(org), German newspaper(org), Panamanian law firm(org), Mossack Fonseca(org)	Story, article, files(slug, secret, companies), companies(offshore), books, Mossack Fonseca(Panamanian law firm, files)	Began(story, February 2015, article, revealed), revealed(Suddeutsche Zeitung, files)	Began(4), revealed(3)	NA	Began(10), revealed(3)

Table A.85: Experiment 1 - Sentence 85

- More than 140 high-ranking politicians and heads of state were uncovered in the data, leaving partners in less democratic countries at risk of government retaliation.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	4	leaving(1), retaliation(0)	leaving(1), retaliation(0)	leaving(Y), retaliation(Y)	NA
SHALMANESER	2	0	5	leaving(0), retaliation(0)	leaving(2), retaliation(0)	leaving(Y), retaliation(Y)	NA
LTH SRL	3	0	3	Uncovered(1), leaving(1), retaliation(1)	Uncovered(3), leaving(3), retaliation(1)	Uncovered(Y), leaving(Y), retaliation(Y)	NA
Illinois Curator	2	0	4	Uncovered(1), leaving(0)	Uncovered(2), leaving(2)	Uncovered(Y), leaving(Y)	NA
Boxer (DRT)	2	0	NA	Uncovered(1), leaving(1)	NA	NA	NA
Stanford NLP	0	0	0		NA	NA	NA
Manual annotation	Uncovered, leaving, retaliation	140 high-ranking politicians(countx-N_count)	Heads(state), data, partners, countries(less democratic), risk	Uncovered(140 high ranking politicians, heads, data, leaving), leaving(partners(countries), risk, retaliation), retaliation(government)	Uncovered(4), leaving(3), retaliation(1)	NA	Uncovered(4), leaving(15), retaliation(1)

Table A.86: Experiment 1 - Sentence 86

- Since the Panama Papers were published Ecuador's president, Rafael Correa, has used Twitter to name and admonish local journalists for not handing over all the data.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	1	2	used(1), name(0), admonish(0), handing(0)	used(2), name(0), admonish(0), handing(1)	used(Y),name(Y),admonish(Y), handing(Y)	NA
SHALMANESER	4	1	3	used(2), name(0), admonish(0), handing(2)	used(3), name(3), admonish(3), handing(4)	Published(Y), used(Y),name(Y),admonish(Y), handing(Y)	NA
LTH SRL	5	0	3	Published(1), used(2), name(0), admonish(2), handing(1)	Published(1), used(4), name(3), admonish(3), handing(1)	Published(Y), used(Y),name(Y),admonish(Y), handing(Y)	NA
Illinois Curator	5	0	2	Published(1), used(1), name(1), admonish(1), handing(2)	Published(2), used(2), name(3), admonish(3), handing(2)	Published(Y), used(Y), name(Y), admonish(Y), handing(Y)	NA
Boxer (DRT)	5	1	NA	Published(0), used(1), name(1), admonish(1), handing(1)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	Published, used, name, admonish, handing	Panama Papers(product-other), Rafael Correa(person), Twitter(org)	President(Ecuador), journalists(local), data(all)	Published(Panama Papers, Rafael Correa, used), used(Rafael Correa, Twitter, name, admonish), name(journalists, handing), admonish(journalists, handing), handing(not, data)	Published(3), used(4), name(2), admonish(2), handing(2)	NA	Published(5), used(9), name(15), admonish(3), handing(2)

Table A.87: Experiment 1 - Sentence 87

- In places far away from Britain, the impact of the Panama Papers was beginning to be felt in local and apparently isolated political scandals.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	1	2	Beginning(0)	Beginning(2)	Beginning(Y)	NA
SHALMANESER	1	0	0	Beginning(1)	Beginning(2)	Beginning(Y)	NA
LTH SRL	2	1	1	Beginning(0), scandals(1)	Beginning(3), scandals(2)	Beginning(Y), scandals(Y)	NA
Illinois Curator	1	1	2	Beginning(1)	Beginning(3)	Beginning(Y)	NA
Boxer (DRT)	Unavailable						
Stanford NLP	1	1	2	Beginning(1)	NA	NA	NA
Manual annotation	Beginning, scandals	Places(far away from Britain(location), Panama Papers(product-other)	Impact, local, isolated political scandals	Beginning(impact, Panama Papers, places(far away from Britain)), scandals(isolated, political)	Beginning(3), scandals(2)	NA	Beginning(16), scandals(2)

Table A.88: Experiment 1 - Sentence 88

- Forty journalists from a dozen news organisations including Le Monde, the Guardian and the BBC are invited to the International Consortium of Investigative Journalists in Washington to discuss collaboration.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	1	1	invited(0), discuss(1)	invited(1), discuss(1)	invited(Y), discuss(Y)	NA
SHALMANESER	1	1	1	discuss(1)	discuss(2)	discuss(Y)	NA
LTH SRL	2	0	1	invited(0), discuss(1)	invited(2), discuss(2)	invited(Y), discuss(Y)	NA
Illinois Curator	2	0	1	invited(0), discuss(0)	invited(3), discuss(2)	invited(Y), discuss(Y)	NA
Boxer (DRT)	2	5	NA	invited(2), discuss(1)	NA	NA	NA
Stanford NLP	0	3	0		NA	NA	NA
Manual annotation	Invited, discuss	Fort journalists(countx-N_count), dozen news organisations(countx-N_count), Le Monde(org), Guardian(org), BBC(org), International Consortium of Investigative Journalists(org), Washington(location)	Collaboration	invited(forty journalists, International Consortium of Investigative Journalists, Washington), discuss(Washington, collaboration)	invited(3), discuss(2)	NA	invited(8), discuss(2)

Table A.89: Experiment 1 - Sentence 89

- The sheer size and complexity of the $2bn Russian money-laundering operation surrounding Vladimir Putin's close friends has begun to emerge.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	1	1	2	emerge(0)	emerge(1)	emerge(Y)	NA
SHALMANESER	1	0	3	emerge(0)	emerge(1)	emerge(Y)	NA
LTH SRL	1	0	2	emerge(0)	emerge(1)	emerge(Y)	NA
Illinois Curator	Unavailable						
Boxer (DRT)	1	0	NA	emerge(0)	NA	NA	NA
Stanford NLP	0	2	0		NA	NA	NA
Manual annotation	emerge	$2bn (money), Russian(org-nationality), Vladimir Putin(person)	Size, complexity, money-laundering operation, friends(close)	emerge($2bn, Russian money laundering operation, friends(Vladimir Putin))	emerge(3)	NA	emerge(5)

Table A.90: Experiment 1 - Sentence 90

- The Panama Papers show that billions of illicit dollars secretly flow through some of the largest banks licensed in America.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	0	2	Show(1), flow(0)	Show(2), flow(1)	Show(Y), flow(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	2	1	2	Show(1), flow(1)	Show(2), flow(3)	Show(Y), flow(Y)	NA
Illinois Curator	1	0	1	flow(0)	flow(2)	flow(Y)	NA
Boxer (DRT)	2	1	NA	Show(0), flow(2)	NA	NA	NA
Stanford NLP	1	2	1	flow(2)	NA	NA	NA
Manual annotation	Show, flow	Panama Papers(product-other), America(location)	Dollars(billions, illicit), banks(largest, America)	Show(Panama Papers, dollars, flow), flow(dollars, banks, America)	Show(3), flow(3)	NA	Show(16), flow(14)

Table A.91: Experiment 1 - Sentence 91

- The answer is that Congress has mastered the political art of erecting Potemkin villages that create the appearance of regulation, especially when it comes to large global banks and other international financial institutions.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	3	1	5	Mastered(0), erecting(0), create(1)	Mastered(0), erecting(1), create(2)	Mastered(Y), erecting(Y), create(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	3	0	2	Mastered(1), erecting(0), create(1)	Mastered(3), erecting(2), create(3)	Mastered(Y), erecting(Y), create(Y)	NA
Illinois Curator	3	0	4	Mastered(1), erecting(0), create(0)	Mastered(2), erecting(2), create(3)	Mastered(Y), erecting(Y), create(Y)	NA
Boxer (DRT)	3	0	NA	Mastered(1), erecting(1), create(1)	NA	NA	NA
Stanford NLP	1	0	1	Mastered(1)	NA	NA	NA
Manual annotation	Mastered, erecting, create	Potemkin villages(location)	Answer, Congress, art(political), appearance(regulation), banks(large, global), institutions(financial, international)	Mastered(Congress, erecting), erecting(Potemkin villages, create), create(appearance)	Mastered(2), erecting(3), create(1)	NA	Mastered(5), erecting(3), create(6)

Table A.92: Experiment 1 - Sentence 92

- The Panama Papers reveal that the law firm wiped its Nevada computer files and moved papers to Panama to conceal some of its conduct from U.S. authorities trying to recover money hidden in shell companies.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	0	6	Reveal(0), moved(1), conceal(0), recover(0)	Reveal(0), moved(2), conceal(0),recover(1)	Reveal(Y),wiped(Y),moved(Y), conceal(Y), recover(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	5	1	3	Reveal(1), wiped(2), moved(2), conceal(0), recover(0)	Reveal(2), wiped(2), moved(4),conceal(3), recover(1)	Reveal(Y),wiped(Y),moved(Y), conceal(Y), recover(Y)	NA
Illinois Curator	5	0	5	Reveal(1), wiped(2),moved(0), conceal(0),recover(0)	Reveal(2), wiped(2), moved(1),conceal(2), recover(2)	Reveal(Y),wiped(Y),moved(Y), conceal(Y), recover(Y)	NA
Boxer (DRT)	5	3	NA	Reveal(0), wiped(2), moved(3), conceal(2), recover(2)	NA	NA	NA
Stanford NLP	2	3	2	moved(1), conceal(1)	NA	NA	NA
Manual annotation	Reveal, wiped, moved, conceal, recover	Panama Papers(product-other), Nevada(location), Panama(location), U.S.(org)	Firm(law), files(computer), papers, conduct, authorities, money(hidden), companies(shell)	Reveal(Panama Papers, firm, wiped), wiped(firm, files(Nevada)), moved(papers, Panama, conceal), conceal(conduct, authorities(U.S.), recover(money, companies)	Reveal(3), wiped(2), moved(3), conceal(2), recover(2)	NA	Reveal(3), wiped(1), moved(17), conceal(2), recover(6)

Table A.93: Experiment 1 - Sentence 93

- But Obama did win plaudits from the Russian strongman for having admitted that his biggest mistake in office was having failed to plan adequately for the aftermath of the toppling of the Moammar Gadhafi regime in Libya.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	4	0	3	Win(2), admitted(0), plan(0), toppling(0)	Win(3), admitted(1), plan(0), toppling(0)	Win(Y), admitted(Y), plan(N), toppling(Y)	NA
SHALMANESER	Unavailable						
LTH SRL	3	1	4	Win(2), admitted(0), plan(0)	Win(3), admitted(1), plan(2)	Win(Y), admitted(Y), plan(Y)	NA
Illinois Curator	3	0	6	Win(2), admitted(1), plan(0)	Win(3), admitted(2), plan(3)	Win(Y), admitted(Y), plan(Y)	NA
Boxer (DRT)	3	2	NA	Win(2), admitted(1), plan(2)	NA	NA	NA
Stanford NLP	1	3	1	plan(0)	NA	NA	NA
Manual annotation	Win, admitted, plan, toppling	Obama(person), Russian strongman(person), Moammar Gadhafi(person), Libya(location)	Plaudits, his, mistake, office, aftermath, regime	Win(Obama, plaudits, Russian strongman), admitted(Russian strongman, mistake, plan), plan(failed, aftermath(toppling)), toppling(regime(Moammar Gadhafi, Libya))	Win(4), admitted(3), plan(2), toppling(2)	NA	Win(7), admitted(8), plan(7), toppling(2)

Table A.94: Experiment 1 - Sentence 94

- Russian tennis champion Maria Sharapova was suspended from play by the International Tennis Federation after testing positive for the drug at the Australian Open in January.

	Event Identification	Named Entity Recognition	Object Identification	Predicate Relationship - Ident.	Predicate Relationship - Count	WSD - Ident.	WSD - Count
SEMAFOR	2	3	2	Suspended(1), testing(0)	Suspended(1), testing(1)	Suspended(Y), testing(Y)	NA
SHALMANESER	1	0	3	testing(2)	testing(5)	testing(Y)	NA
LTH SRL	2	0	3	Suspended(3), testing(1)	Suspended(4), testing(4)	Suspended(Y), testing(Y)	NA
Illinois Curator	2	0	2	Suspended(3), testing(0)	Suspended(4), testing(1)	Suspended(Y), testing(Y)	NA
Boxer (DRT)	2	3	NA	Suspended(3), testing(2)	NA	NA	NA
Stanford NLP	1	3	2	Suspended(2)	NA	NA	NA
Manual annotation	Suspended, testing	Russian(org-nationality), tennis(product-sport), Maria Sharapova(person), International Tennis Federation(org), Australian Open(event-game), January(timex-date)	Champion, play, drug(positive)	Suspended(Maria Sharapova(Russian, champion(tennis)), play, International Tennis Federation), testing(drug, Australian Open, January)	Suspended(4), testing(3)	NA	Suspended(7), testing(10)

Table A.95: Experiment 1 - Sentence 95