HELPING METONYMY RECOGNITION AND TREATMENT THROUGH NAMED ENTITY RECOGNITION

(ADLANDIRILMIŞ VARLIK ILE AD AKTARMASI ÇÖZÜMLEME)

by

HATİCE BURCU KÜPELİOĞLU, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

JULY 2016

Approval of the thesis:

HELPING METONYMY RECOGNITION AND TREATMENT THROUGH NAMED ENTITY RECOGNITION

submitted by **HATİCE BURCU KÜPELİOĞLU** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Galatasaray University** by,

Examining Committee Members:

Prof. Dr. TANKUT ACARMAN
Supervisor, **Computer Engineering Department, GSU**

Prof. Dr. BERNARD LEVRAT
Computer Engineering, University of Angers

Assist. Prof. Dr. R.BURAK ARSLAN
Computer Engineering Department, GSU

Date:

ACKNOWLEDGEMENTS

The goal of this work is to recognize and treat metonymies while considering named entity recognition.

The reason which drove me to effect this study is my choice of domain. I wish to proceed to further academic research on this domain which is natural language processing.

I thank Bernard LEVRAT to accept me as student and give effort on my education.

I thank Tankut ACARMAN to be my supervisor and help me through graduation process.

I thank my family and my friends for their support whatever the conditions.

July, 2016
H.Burcu KÜPELİOĞLU

TABLE OF CONTENTS

LIST OF SYMBOLS

BNC	British National Corpus
NEBMR	Named Entity Based Metonymy Resolution
NER	Named Entity Recognition
NLP	Natural Language Processing
PM	Possible Metonymy
SRV	Selectional Restriction Violation
WSD	Word Sense Disambiguation

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Internet and computers has become more and more important in human life. Especially with Web 2.0, the permission given to surfers of not only reading but also changing the content has encouraged humans to change their habits of explorations. Nowadays mankind use internet instead of encyclopaedias and books. This undeniable growth of digitalism put forth the importance of automatically processing of the data. Automatically processing the content means turning unstructured data into structured data. By unstructured data we connote human language – natural language data. This treatment is necessary for computers to understand the natural language content. Natural Language Processing (NLP) is the discipline behind this process. NLP is a subcategory of artificial intelligence and computer linguistics.

Named Entity Recognition (NER) and Word Sense Disambiguation (WSD) is two of the major tasks of NLP. NER is the classification and extraction process of word(s) considered significant in a text. This significant word(s) can differ according to field. For example, this entities may be percentages, dates as well as person names, location names and company names, etc. WSD is an open problem in NLP. It consists of identifying the sense of a word, when having multiple meaning, in a sentence. WSD tries to identify litteral expressions of a word, not figurative expressions. Figurative Language Processing is the study similar and all but subfield of WSD. Figurative Language Processing concentrates on determining figurative expressions except litteral expressions.

Our project is based on metonymy recognition and resolution through named entity recognition. Metonymy is a figure of speech which consists by using a concept b to refer to concept a, without intending analogy. The existing methods of metonymy resolution depends on supervised and unsupervised methods as well as statistical approaches. The commonly used approaches are catching the Selectional Restriction Violations (SRVs) and deviations from grammatical rules. We consider our project having three parts. First part is to pre-process the given text.

Pre-processing is necessary for further treatment. Pre-processing consists of lemmatization, part-of-speech tagging, NER tagging, dependency tagging and WSD treatment. The second part is metonymy recognition, in other words detections of possible metonymies. Metonymy recognition is realized via named entities' SRVs. It is done by a rule based algorithm. The last and the third part is metonymy resolution which consists of determining metonymic relation.

Keywords : METONYMY, NAMED ENTITY, STANFORD CORENLP, WORDNET

ÖZET

İnternet ve bilgisayar kullanımı insanların hayatında gittikçe daya büyük bir yer kaplamaya başlamıştır. Özellikle Web 2.0'la birlikte kullanıcılara içerik yaratma ve değiştirme imkanın da sağlanması, insanların araştırmalarını ansiklopedi ve kitaplar yerine internet üzerinden yapmaya teşvik etmektedir. Bu denli dijitalleşme bilgisayarlardaki verilerin boyutunun son yirmi yılda tahmin edilemeyecek kadar büyük boyutlara ulaşmasını sağlamıştır. Veri hacmindeki bu büyüme verilerin otomatik işlenmesinin ne kadar önemli ve elzem bir hale geldiğini kuşkusuz ortaya koymaktadır.

Bilgisayar ve internet üzerindeki veriler yapılandırılmamış verilerdir. Diğer bir deyişle doğal dil olarak adlandırılan, bireylerin kullandığı insan dilidir. Dolayısıyla bu veriler insanlar için anlam taşırken bilgisayarlar için anlamsızdır. Bilgisayarların bu verileri otomatik işleyebilmesi için anlaması şarttır. Bilgisayarların insan dilini anlaması ve çözümlemesi için yapılandırılmamış bu verilerin yapılandırılmış bir hale getirilmesi gerekmektedir. Yapılandırılmamış verilerin yapılandırılması ile ilgilenen bilim dalına Doğal Dil İşleme (Natural Language Processing - NLP) adı verilir. Doğal Dil İşleme yapay zeka ve dilbilim çalışma alanlarının alt kategorisidir.

Adlandırımış Varlık Tanıma (Named Entity Recognition - NER) ve Kelime Anlamda Anlam Ayrımı(Word Sense Disambiguation - WSD) birçok alt dalı bulunan Doğal Dil İşlemenin ana görevlerinden ikisidir. Adlandırılmış Varlık Tanıma, bir metinde geçen anlamlı kabul edilebilecek kelime ve kelime gruplarını tanıyarak sınıflandırma ve özütleme işlemidir. Sınıflandırılmak istenen sözcük öbekleri çalışılan alana göre değişiklikler göstermektedir; örneğin bu öbekleri yüzdeler, tarihler, mesafeler, kişi isimleri, yer isimleri, firma isimleri vs şeklinde çeşitlendirmek mümkündür. Kelime Anlamda Anlam Ayrımı ise Doğal Dil İşlemenin açık sorunlarından birisidir. Kelime Anlamda Anlam Ayrımı bir kelimenin, birden fazla anlamı olması halinde, bir cümle içerisinde hangi anlamında kullanıldığını tespit etmeye yarayan çalışma alanıdır. WSD kelimelerin sözlük anlamları, bir diğer deyişle gerçek anlamları arasında araştırma yapar. WSD'na benzer ve alt kategorisi olarak kabul edilebilecek bir diğer çalışma ise Figüratif Dil İşleme (Figurative Language Processing)'dir. Figüratif Dil İşlemenin amacı kelimelerin ve sözcüklerin sözlük anlamları dışında mecazi kullanımlarını tespit etmektir.Bizim üzerinde çalıştığımız proje Adlandırılmış Varlık Tanıma yardımıyla ad aktarması (mecaz-ı mürsel) tanıma ve çözümleme projesidir. Ad aktarması benzetme amacı güdülmeden, varolan bir ilişkiye referans vererek bir kavramın başka bir kavram yerine kullanılmasıyla oluşan mecazdır. Ad aktarması tanıma ve çözümleme çalışmaları denetimli, denetimsiz ve istatistiksel metotlar kullanılarak gerçekleştirilir. Ad aktarması çalışmaları genelilkle Seçmesel Kısılama İhlalleri (Selectional Restriction Violations – SRVs) ve dilgisi kurallarında aykırılıklar incelenerek yapılır. Projemiz üç bölüme ayrılmıştır. Bunlardan birincisi, mevcut etiketleme araçları kullanılarak ad aktarması çözümlemesi yapılacak metnin ön işlemesini yapmaktır. İkinci bölümde ön işlemesi yapılmış, etiketlenmiş metne seçmesel kısıtlama ihlalleri ve dilbilgisi kuralları gözetilerek oluşturduğumuz kural tabanlı algoritma uygulanır. Bu uygulama verilen bir adlandırılmış varlığın ad aktarması olup olmadığı tespit etmeye çalışır. Üçüncü ve son kısımda ise eğer sonuç ad aktarması ise bu ad aktarmasının türünü belirlemek, bir diğer deyişle gizli kavramı bulmaktır.

Anahtar Kelimeler : AD AKTARMASI, ADLANDIRILMIŞ VARLIK, STANFORD CORENLP, WORDNET

1 INTRODUCTION

The internet has become a large data collection through the past two decades. As a significant amount of this data is unstructured - human language text data, automatic processing of this information is of great importance.

Natural Language Processing (NLP) is a computer linguistics and artificial intelligence domain which tries to understand and work with human language. Researchers has been working on this discipline since the second half of the 20th century. This treatment consists of transforming unstructured data into structured data which consists of managing interactions between computers and natural languages. Named Entity Recognition (NER) and Figurative Language Processing are two of the major tasks of NLP. In the scope of this work, we are concentrated on resolving metonymies making use of NER tools.

Metonymy is a figure of speech in which a concept is replaced by another one connected logically (Wilks, 1978). Metonymy draws reference to an existing relation between two concepts such as an artist for his artwork or a producer for his products. Although metonymy can be came across often in everyday speeches but also in literature such as poetry.

Even by humans metonymy is often elusive. The task is even more challenging for computers. Previous work has been done by the years, with different approaches. There are supervised and unsupervised methods, knowledge-based and statistical approaches. The Selectional Restriction Violations (SRVs) and abuse of grammatical rules are commonly used clues for metonymy resolution task (Roberts and Harabagiu, 2011). These approaches are detailed in Section 2.1.

In our work, we will concentrate on metonymy resolution through named entities. Named entity recognition is realized by named entity recognition tools. We use SemEval 2007 Task 8 (Markert and Nissim, 2007) as our test and key data which is annotated for metonymy resolution. To process these data we use Stanford CoreNLP (Manning et al., 2014) toolkit. We use WordNet (Miller et al., 1990) as thesaurus for our semantic rules.

The basis of our metonymy resolution work depends on POS, NER and dependency tagging annotators which are used to process the sentences, then interpreting the dependencies of the named entity to be metonymic. We call it potential metonymy, PM. To interpret we have to identify the sense of PM's dependencies in the sentence. This is one of the principal tasks in our method. To realize this task we use WordNet's lexicographer files for nouns and verbs.

This document is organised as following : In the next section (section 2) we are presenting the metonymy resolution task, previous approaches and metonymy types we are interested in the scope of this work. In the section 3, our algorithm and our rules are detailed alongside the tools and thesaurus we use. Then (section 4) we describe the corpus we used SemEval 2007 Task 8. Section 6 contains our results and the analysis of the results when applying our algorithm to SemEval corpus. Finally, as conclusion (section 6) we present the balance sheet of our work, our contributions and limitations and we discuss possible improvements.

Note : all of the examples presented in this document are extracts from SemEval 2007 Task 8 except Figure 2.1.

2 METONYMY RECOGNITION AND RESOLUTION

Metonymy Recognition and Resolution is the main task of our project. In order to detect implied concept we must search for the metonymic word and then extract the metonymic relationship. These detections and classifications are majorly due to Natural Language Processing tools and Word Sense Disambiguation methods. In this section, we define the principle of metonymy recognition, and also define metonymy types of which we are interested as part of this project.

2.1 Definition

Metonymy Recognition is a subtask of Figurative Language Processing. Metonymy is a figure of speech in which the name of one thing is used for another with which it is logically associated (see Figure 2.1).

Metonymy Recognition presents two main difficulties. First, it is difficult to identify if a word is metonymic. Second, it is more difficult to identify the metonymic relationship. Metonymy Recognition task different approaches.

One significant step is taken by Markert and Nissim (2007). It consisted of metonymy resolution for countries and companies. A large corpora containing country and company names is annotated especially for metonymy resolution task. However, this corpora is annotated manually, which is considered to be very expensive by means of time because human annotation are laborious.

Lapata (2003) and Shutova (2009) proposed a probabilistic model to deal with metonymy. Lapata approaches the problem with an unsupervised method.

Markert and Nissim (2002) approach the problem as a classification task shaped by co-occurences of metonymic readings in training data to form a classification for location

`He read Shakespeare.`

Figure 2.1: *Shakespeare* metonymic for an artwork of Shakespeare.

– country names.

Amghar et al. (1995) make use of conceptual graphs with handling Selectional Restriction Violations (SRVs). SRV consist of a semantic restriction that enforce for a predicate to control the semantic categorization of its arguments.

Nissim and Markert (2005) proposes a supervised classification method for organization names which consists of training the algorithm by a set of key instances of different metonymic words of one semantic class in order to work with new test instances for metonymy of the same semantic class.

An analysis of metonymies in discourse is proposed by Markert and Hahn (2002). A learning algorithm is presented by Birke and Sarkar (2007). Bogdanova (2010) and Nastase et al. (2012) propose an approach by cluster based sense differentiation and usage of SRVs.

Another approach is the use of feature vectors. Each sample is expressed as a feature vector. It is possible to predict metonymy by using like Naïve Bayes machine learning classifiers (Russell and Norvig, 1995). These feature vectors are formed by human annotators by selecting informative and distinctive characteristics. It is a classified probabilistic model representing assigned features.

It is possible to apply WSD classification algorithms and features to metonymy resolution with semantic classifiers. The decision lists may be formed with likelihood ratios. When forming these lists, collocations, co-occurrences and grammatical roles are considered. Our rules can be considered as applying semantic classifiers to the sentences containing our possible metonymic named entity.

Our motivation to work with named entities and lexicographer files is to eliminate massive human work. With our method, we only need a corpus and a lexical database. Metonymies can be divided into two categories :

— Conventional Metonymy : We define Conventional Metonymy as a metonymy detectable by a large scale of humans. It is described by popular proper and common nouns, numbers, expressions, in another saying named entities. In our work, we are interested in metonymies with proper nouns such as country names, company names.

— Unconventional Metonymy : We define Unconventional Metonymy as a metonymy detectable by a group of person, a clique, or in a context-based situation. (In the context of our work we are not interested in extracting such metonymies because of the need of other major tasks in Natural Language Processing.

2.2 Metonymic relation types

Metonymy is based on a logic association in other words drawing reference between the predicate and the argument. This logic association can be shown in form of different relations. For the final reason of our work, which is detection of metonymic word and its relationship, we concentrate only on 7 of the 9 relations metonymic relations considered in SemEval 2007 Task 8 :

- Place-for-people : location names which are associated to people or members
- Place-for-event : location names which are associated to event
- Place-for-product : location names instead of products produced at that place
- Organisation-for-members : organization names which are associated to people or members
- Organisation-for-event : organization names which are associated to events
- Organisation-for-product : organization names instead of products produced at that place
- Organisation-for-facility : organization names which are associated to facility
- Organisation-for-index : organization names which are associated to index
- Other met : metonymies which don't fit any of the relations cited above

It is possible to define many other metonymic relations like artist for artwork, etc. but our data set is limited to LOCATION and ORGANIZATION typed named entities and these metonymic relations. In the following subsections, each metonymic relation is described in detail with an example extract of the SemEval 2007 Task 8 data set.

2.2.1 Place-for-people

Place-for-people metonymy relation is encountered by LOCATION typed named entities. In these relations the hidden idea is one or a group of person associated with location name (see Figure 2.2).

Hundreds of women and children were left behind. The racket was exposed at a refugee centre in Croatia. Britain and America are close to agreeing the best way of protecting relief supplies to Bosnia, Foreign Secretary Douglas Hurd said yesterday.

Figure 2.2: *Britain* for British Government, place-for-people metonymy.

And a warning. Northern Ireland's experience in Spain and Mexico in the last two World Cups taught them that heat can pose more problems than the opposition for British-based players. Bingham said: The Republic have the potential to do well in Italy. Just how well will probably depend on the temperature.

Figure 2.3: *Italy* for World Cup to take place in Italy, place-for-event metonymy.

2.2.2 Place-for-event

Place-for-event metonymy relation is encountered when an event is expressed with its location name (see Figure 2.3). This relation is encountered with LOCATION typed named entities.

2.2.3 Place-for-product

Place-for-product metonymy relation is present when a location name is used instead of a product produced there (see Figure 2.4).

2.2.4 Organization-for-members

Like place-for-people relations organisation-for-members metonymy relation is encountered when member(s) of this organisation is the implied concept with ORGANIZATION typed named entities (see Figure 2.5).

2.2.5 Organization-for-event

Organisation-for-event happens when an event is implied by an organisation to which it belongs (see Figure 2.6). This type of metonymy is eligible for ORGANIZATION

In this form the Art Fair reached a more numerous and a far wider public; and it is hoped that the Tabernacle will provide a permanent home for this venture, which does so much to help young artists of promise in the difficult years after leaving Art College until they make their name. On Monday 18th June sixty people met in Painters' Hall in the City to take part in a "tutored wine tasting" of sparkling wines, under the most knowledgeable guidance of Pamela Vandyke Price, author and wine journalist of fabulous expertise, whose distinctions include having been the first woman admitted on equal terms into the male-dominated world of wine. Pamela had chosen six reasonably priced sparkling wines; a cosmopolitan collection, ranging from France via Portugal to Australia. While we slipped and spat with deep concentration, Pamela gave a running commentary on the wine in question, combining information about its history and manufacture with many practical and amusing tips about wine management.

Figure 2.4: *Australia* for sparkling wines, place-for-product metonymy.

SunSoft announced a VAR program designed to educate resellers on selling 32-bit computing. It includes training and education, sales and lead generation; marketing and merchandising support. SunSoft Inc. claims Solaris-on-Intel is being evaluated by Amoco (10,000 units), AT&T Universal Card, DuPont Pixel Inc., Foxboro, Philip Morris, Superior National Insurance and 3M.

Figure 2.5: *DuPont Pixel Inc.* for its administration, organisation-for-members metonymy.

One reason for this reluctance to take action against the process of monopolization is the difficulty of distinguishing acceptable and unacceptable behaviour. As we saw in the previous section, there is an understandable reluctance to move against firms that have competed successfully and won market share. The case history of US antitrust Alcoa, United Shoe, AT& T, and IBM) in the post-war period is a witness to the difficulties that competition authorities face in this area. In both Alcoa and United Shoe the courts acknowledged that the defendants had built their market shares by legal means, but none the less found them guilty because their respective dominant positions were due to conscious choice.

Figure 2.6: *Alcoa* for its antitrust from US, organisation-for-event metonymy.

Ian Miller, 34, was swamped by the problems of small businesses, an inquest heard. The strain finally became too much for the father-of-three who worked in John Major's Huntingdon constituency in Cambridgeshire. He drove his Volvo to a beauty spot last month, drank most of a bottle of whisky and gassed himself with fumes from the car's exhaust. "There is bit of an anti-bank campaign going on and it worried Ian," his manager Jean Temple told the inquest.

Figure 2.7: *Volvo* for a car of the brand, organisation-for-product metonymy.

typed named entities. This metonymic relation is not considered in the scope of this work due to its rare occurrence.

2.2.6 Organization-for-product

This type of metonymic relation exists if a product of an organisation is represented by this organisation name (see Figure 2.7). Also eligible for ORGANIZATION typed named entities.

2.2.7 Organization-for-facility

When the name of an organisation is used instead of its facility it is an organisation-for-facility metonymic relation (see Figure 2.8). It is suitable for ORGANIZATION typed

The driver saw two men attacking a middle-aged woman outside a bank cash machine, but instead of stopping to help her, carried on driving, detectives believe. And yesterday they issued a direct appeal to the person to come forward with information so the thugs can be caught before they strike again. The victim, aged 44, is still recovering from her ordeal which happened outside Barclays Bank in Kingsway, Dovercourt, on February 22. She was grabbed and thrown against a wall by two teenagers who snatched the "substantial" amount of cash she had just withdrawn.

Figure 2.8: *Barclays Bank* for its premises, organisation-for-facility metonymy.

Page 30 NY cellular telephones: McCaw Cellular has bought the half stake owned in the New York cellular telephone franchise by Metromedia. Page 31 Accounting mergers: Deloitte Haskins and Sells has pulled out of its merger with Touche Ross and is expected to join up with Coopers & Lybrand. This page and View from City Road, page 31 Mitsubishi listing: Mitsubishi has became the first Japanese general trading company to be listed on the London Stock Exchange. Dealings in its shares begin today.

Figure 2.9: *Mitsubishi* for page 31, organisation-for-index metonymy.

named entities as well.

2.2.8 Organization-for-index

This type of metonymic relation exists if an index of an organisation is represented by this organisation name (see Figure 2.9). Also eligible for ORGANIZATION typed named entities. We didn't focused on this metonymic relation for this work.

2.2.9 Othermet

If we don't encounter any of the previously described relations we consider the relation to be other met, this is all to say we are unable to describe the relation (see Figure 2.10).

In Cato "1976" 1 WLR 110 gross negligence and recklessness were used synonymously. Even after Seymour gross negligence and recklessness are used interchangeably. In Sargent "1990" The Guardian, 3 July, Boreham J at Leeds Crown Court is reported as saying: "You were so negligent as to be reckless as to this woman's welfare" by pumping so much oxygen into her during an operation that she swelled up like a Michelin man. Note the reference to "welfare", a low test which is below that required for reckless manslaughter.

Figure 2.10: *Michelin* for its advertising mascot, other met metonymy.

3 METHODOLOGY

In this section, we describe the tools and libraries we use in order to achieve our main goal and also our method. Generally we use 1 toolkit, 1 thesaurus and 1 WSD algorithm. Our algorithm is detailed at section 3.3. The named entities are commonly used metonymically. 50% of the named entities is observed being used metonymically by a small scale experiment (Markert and Hahn, 2002) conducted on BNC (Leech, 1992). Especially in everyday speech and journal headlines and articles it is possible to find named entities used metonymically. The frequency of named entities being metonymic shaped our motivation to work with NER for metonymy resolution task.

3.1 General Structure

In this study, our main goal is to define and evaluate an algorithm which finds out if a given word in a text is metonymic, and if so outputs its associated metonymic relationship. In order to facilitate this process, we divide the algorithm into smaller subtasks. First step is to pre-process the given text. Then, the processed text is evaluated using multiple rule functions. Finally, use a decision function to aggregate rule function results (see Figure 3.1). Pre-processing step consists of splitting the text into sentences and then into tokens using a tokenizer. Then each token is part-of-speech (POS) tagged, NER tagged and dependency tagged. Then this processed input is sent to multiple rule functions. Each rule function uses a different algorithm on the given text to determine metonymic relations (if one exists) for the given word. Rule functions have access to WordNet database. Results of all rule functions are summarized by an aggregator algorithm.

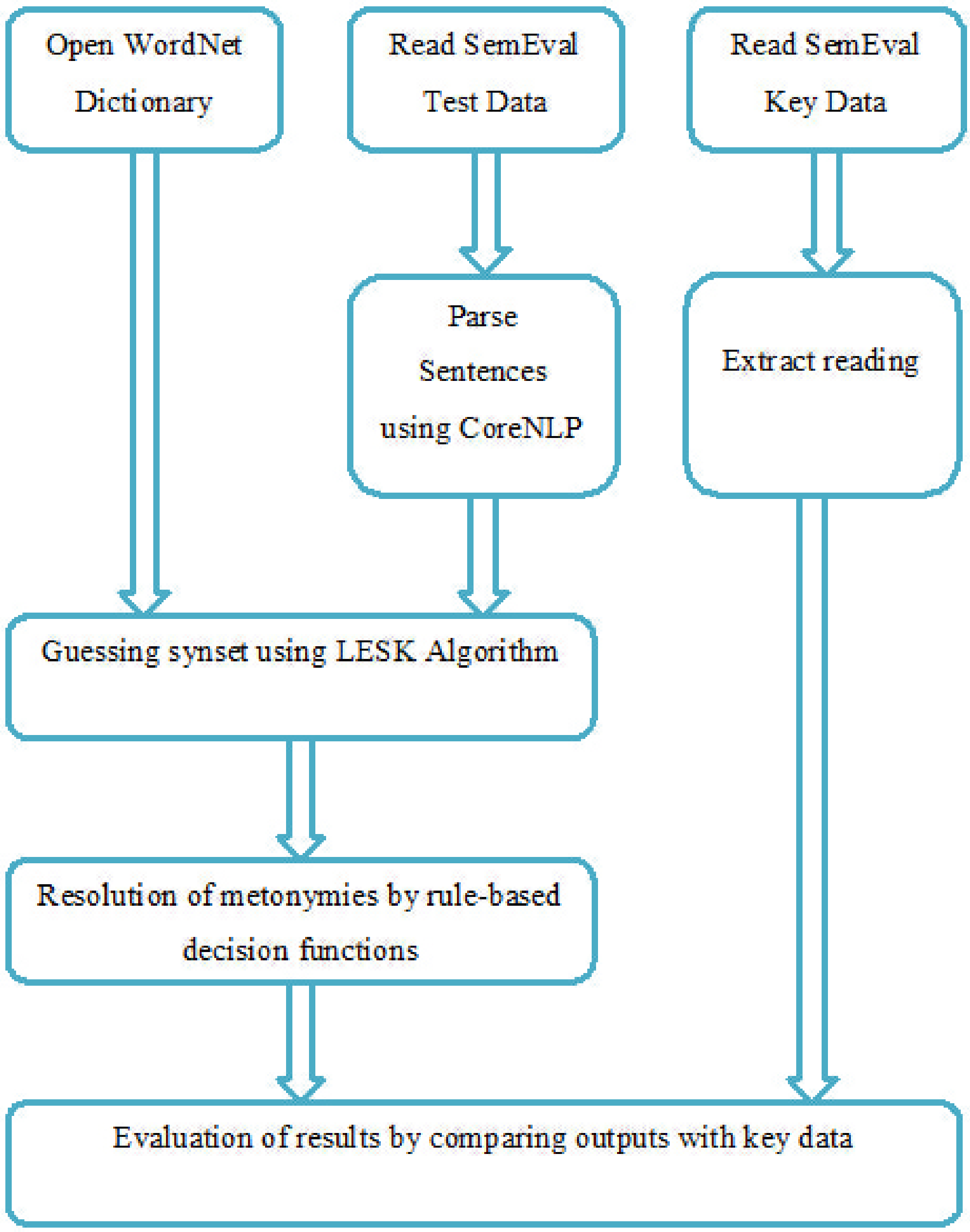

Figure 3.1: General flow diagram of NEBMR

3.2 Dependencies

Our general structure depends on a few major third party libraries and existing tools. In the following subsections we describe these dependencies and their use cases in order of importance.

3.2.1 WordNet

WordNet is a lexical database created by George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. The most distinctive feature of WordNet is the existence of synsets. The synsets are groups of synonyms which serve to characterize a concept in WordNet. The words are lexically classed in WordNet such as verbs, nouns, adjectives and adverbs. Each lexical category is grouped with semantic relations called as lexical semantics to form synsets (example, noun.artifact, verb.possession, etc.). Other semantic and ontological relations exist between words such as hypernymy-homonymy, merynymy, troponymy, etc. In the concept of our work we use WordNet for its syntactic categories and its lexical semantics. Also the implementation of the Lesk Algorithm uses WordNet too.

3.2.2 Word Sense Disambiguation and Lesk Algorithm

Word Sense Disambiguation (WSD) is one of the major tasks in Natural Language Processing. A word is polysemous when it has multiple meanings. Word Sense Disambiguation consists to identify which sense of this polysemous word is used in a given sentence. For humans it is often easy to figure out the sense of a given word. But in some cases two different humans can select two different meaning of a given word in a same sentence. Word Sense Disambiguation has two main approaches, deep approaches and shallow approaches. Deep approaches assume full understanding of a context whereas shallow approaches do not try understand the context. The conventional approaches are listed below :

— Dictionary based methods : usage of dictionaries, lexical databases ; Lesk Algorithm

— Supervised methods : usage of annotated training corpus

- Semi-supervised methods : primarily usage of a small annotated corpus alongside of lexical knowledge
- Unsupervised methods : usage of raw corpus (unannotated)
- Other approaches : various approaches exist like domain-driven disambiguation, WSD using cross-Lingual evidence, etc.

Lesk Algorithm is a dictionary-based algorithm which identifies the meaning of a polysemous word in a given sentence (Lesk, 1986). Basically, the polysemous word's and its neighbours' senses are looked upon a glossary and then they are compared. The closest meaning is selected for the polysemous word. In our work, we are concentrated in verbs, nouns and adjectives to recognize a metonymy (see Section 3.3). If these words have more than one meaning in a given sentence, it is essential to find the sense currently used to maintain our goal. For that purpose we use an adoption of Lesk Algorithm for WordNet (Banerjee and Pedersen, 2002), (Ekedahl and Golub, 2004) after we find the dependencies.

3.2.3 Stanford CoreNLP

Stanford CoreNLP is a JVM-based natural language toolkit. It consists of multiple annotators for several languages. In the scope of this work we integrated tokenization, sentence splitting, lemmatization, parts of speech, named entity recognition and dependency parsing annotators. The dependency parser analyses the grammatical structure of a given sentence. It indicates the grammatical relations between words. The dependency parser is one of the most crucial annotator for our rule algorithms. The second important annotator is named entity recognition annotator. Stanford CoreNLP named entity recognition parser detects named, numerical and temporal entities. These entities are PERSON, LOCATION, ORGANIZATION and MISC for named entities and MONEY, NUMBER, ORDINAL and PERCENT for numerical and finally DATE, TIME, DURATION and SET for temporal entities.

3.3 Algorithm

The algorithm containing our rule functions is implemented in Java, same as Stanford CoreNLP. It is called NEBMR after Named Entity Based Metonymy Recognition. SemEval (2007, task 8) is the corpora that is used in the algorithm which is formatted in

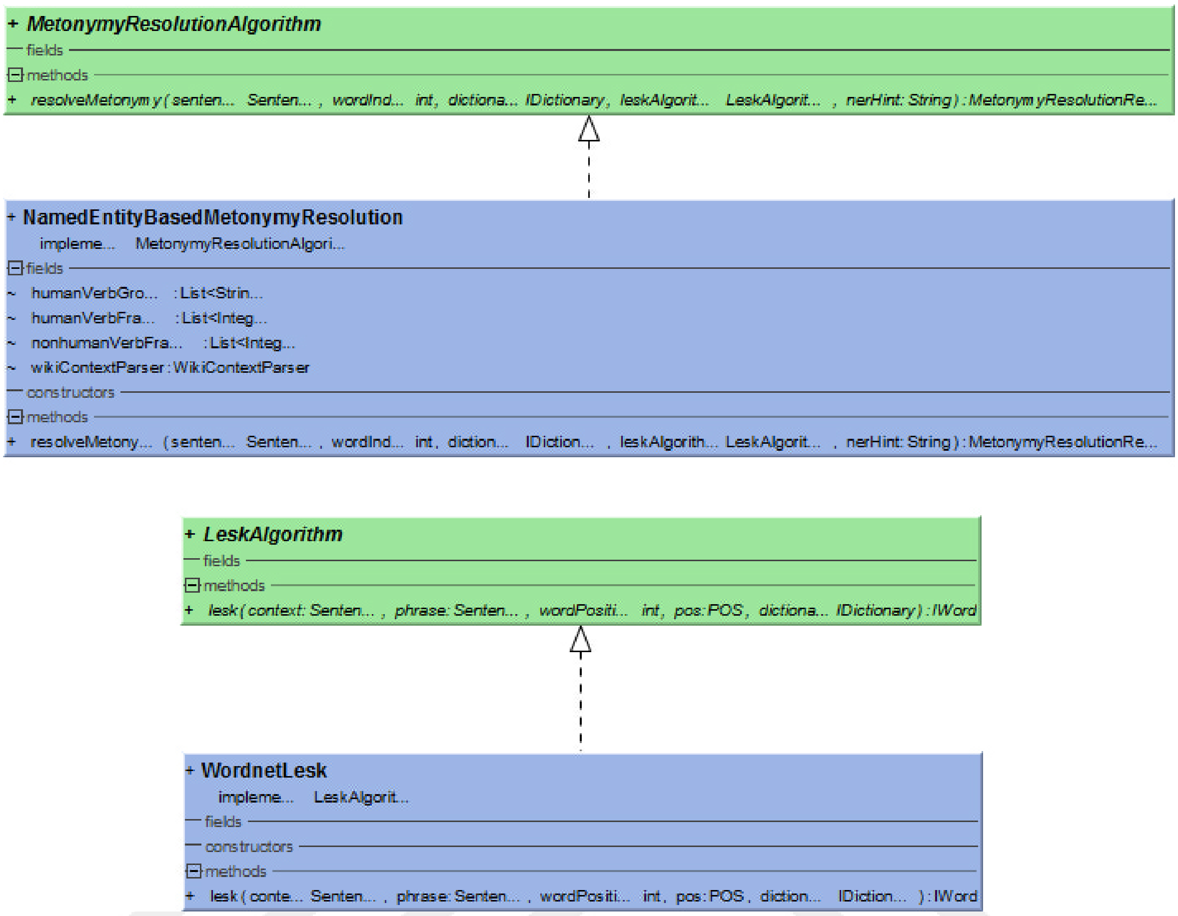

Figure 3.2: Class Diagram of NEBMR

XML. We use WordNet as thesaurus and Lesk Algorithm for WSD. We have four major class, `MetonymyResolutionAlgorithm`, `NamedBasedMetonymyResolution`, `LeskAlgorithm` and `WordnetLesk` (see Figure 3.2).

3.3.1 Pre-processing

The Sentence class of Stanford CoreNLP tokenizes automatically the input string. Then for each token postag, nertag, dependency tag and dependent token information are given.

3.3.2 Rule functions

All rule functions take an ordered list of tokenized and POS tagged, nertagged and dependency tagged sentences alongside an index for the word in question. The result may appear as either *litteral*, *metonymic* or *mixed* or not applicable. If the outcome is

Figure 3.3: Processed text with outcome

not applicable the next rule is being processed (see Figure 3.3, Figure 3.4 and Figure 3.5). Our rules functions rely on named entity –PM and the lexicographer files of its dependent verbs or nouns. These informations are provided by WordNet library. But, for possible polysemous verb or noun we first have to identify the synset with Lesk Algorithm before applying the rules (Section 3.2.2). Mainly we use verb or noun groups for metonymy detection. In sections 3.3.1.1 and 3.3.2.2 we describe verb related detections and in sections 3.3.2.3 and 3.3.2.4 we present noun related detections.

3.3.2.1 Named Entities as Agents

In these cases the most significant distinction is due to verbs. If named entities are agent in a sentence, we check the verb group of the verb which the entity is the subject. Like said earlier, having multiple sense for a word is very possible. The same applies to verbs. To detect the verb group we have to identify the meaning of this verb in the sentence. Some acts can only be considerable for humans, animals or objects but other may be suitable for locations or organisations. The meaning in the given context of the verb whose subject is the named entity belongs to cognitive or communicative verbs it

Figure 3.4: *Fujitsu*, PM, is subject to verb *talking*, which is a verb of communication. For that reason NEBMR decides that *Fujitsu* is used metonymically in this sentence.

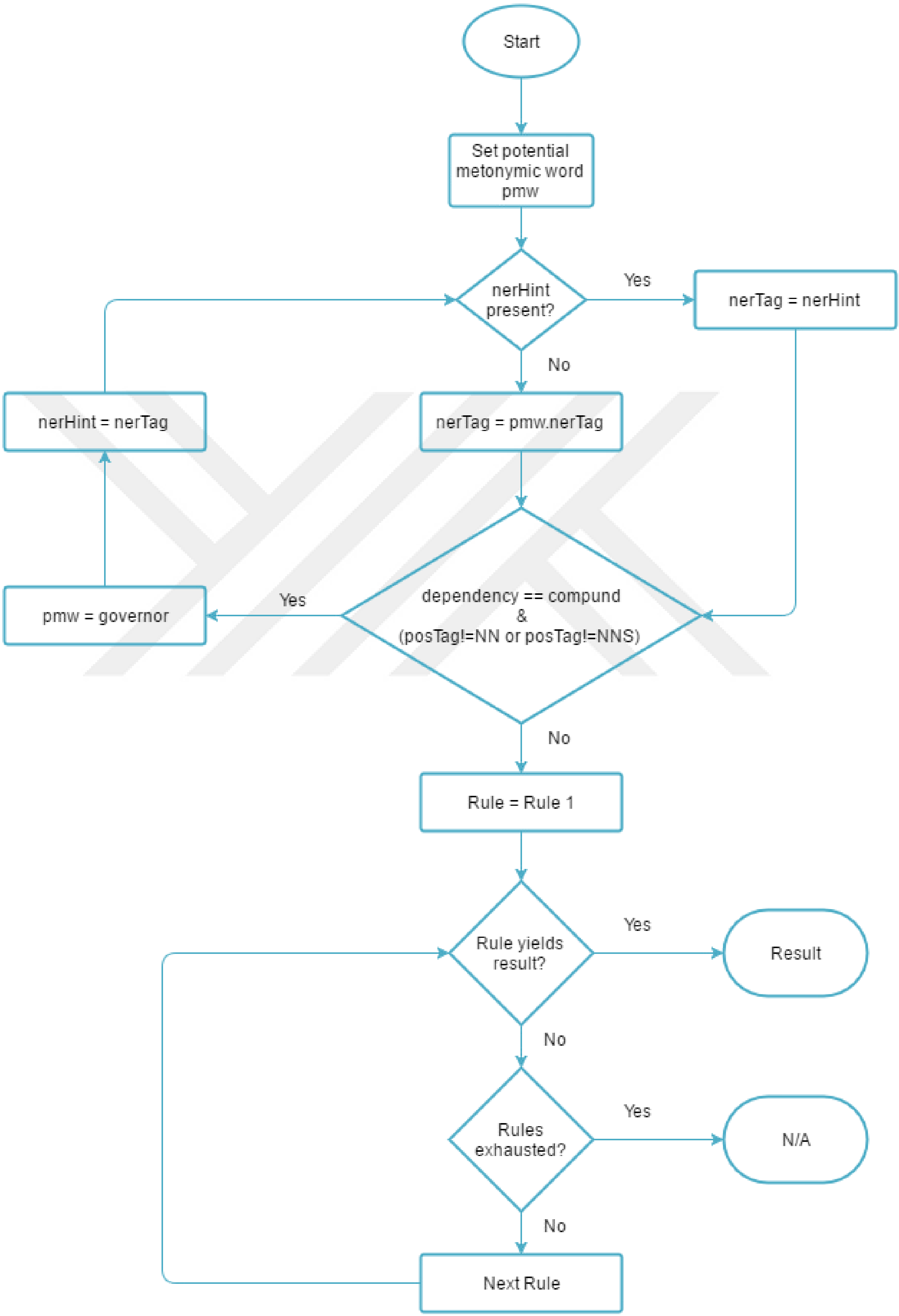

Figure 3.5: Rule-based decision function of NEBMR

is possible to say that PM is metonymic.

3.3.2.2 Named Entities as Predicates or Passive Agents

In these cases also verbs are our biggest clue. When some actions are possible between humans, some can take objects like predicates. Usually very few verb groups are suitable for locations and organisations. We make our decisions entirely based on verb groups as same in the previous section. Again, we must identify the synset of the verb in order to know the corresponding verb group.

3.3.2.3 Named Entities Having Compound Dependencies

In some cases named entities are neither agents nor predicates like when they are compounds. If so, we track compound dependency until we find a common noun. An organisation can have a member or a worker like a location can have an address. For that to detect if PM is either metonymic or not we have to analyse noun groups of this noun. Again, first we determine the synset of the noun and then decide. Like in verbs cases, some nouns are unique to humans or animals like arm, leg or feelings. If so we decide PM to be metonymic.

3.3.3 Implementation

We implemented our NEBMR algorithm alongside existent algorithms, tools and thesaurus. Stanford CoreNLP is the toolkit we use for sentence splitting, NER, tokenization, lemmatization parsing. Lesk Algorithm and WordNet are combined together to define PM's dependency word's semantic context. All the information provided by the existent base are aggregated in NEBMR to recognize and resolve metonymies (see Figure 3.3.3 1).

3.3.3.1 Dependency tags

As we use Stanford CoreNLP POS Tagger our dependencies are in standard of Stanford CoreNLP standard (De Marneffe and Manning, 2008) (see Table 3.1). This standard is also known as Universal Dependencies[1]. The motivation of universal dependency creation is to help researchers study multilingual and cross-lingual easier.

Table 3.1: Some universal dependencies used in NEBMR

Dependency	Definition
nsubj	Nominal subject
nsubjpass	Passive nominal subject
dobj	Direct object
iobj	Indirect object
amod	Adjectival modifier
nmod	Nominal modifier
compound	Compound
conj	Conjunct

3.3.3.2 Verb groups

WordNet's lexicographer files are classified by synset meanings especially for verbs and nouns. One verb synset can only have one verb group corresponding. We have chosen some of the verb groups for the reasons already explained in the previous subsection (see Table 3.2).

Table 3.2: WordNet verb groups considered in NEBMR

Human Agent Verb Groups	Human Predicate Verb Groups	Copular Agent Verb Groups
verb.communication	verb.body	verb.stative (be, become, get, remain, seem, etc.)
verb.cognition	verb.communication	
verb.emotion	verb.possession	
verb.social	verb.social	
verb.possession	verb.weather	
verb.consumption	verb.competition	
verb.competition	verb.consumption	
verb.creation	verb.contact	
verb.body	verb.creation	
verb.perception	verb.motion	
verb.motion		

3.3.3.3 Noun groups

Like verbs, nouns also have groups that they belong according to their synsets. For our work we have chosen some of these noun groups (see Table 3.3).

1. http ://universaldependencies.org/language-en

Table 3.3: WordNet noun groups considered in NEBMR

Human Related Noun Groups	Mixed Noun Groups
noun.act	noun.Tops
noun.body	noun.artifact
noun.cognition	noun.attribute
noun.communication	noun.event
noun.feeling	noun.group
noun.motive	noun.process
noun.object	noun.phenomenon
noun.possession	

3.3.4 Evaluation

Our evaluation depends on two tasks, delimitation and type which correspond respectively to metonymy recognition and metonymy resolution.

We define 4 predicted conditions for metonymy recognition : true positive, false positive, true negative and false negative (see Table 3.4, Table 3.5). We divided the evaluation in two parts. In one evaluation, we consider mixed readings also. True positives are the cases when our result is either metonymic or mixed and the reading is either metonymic or mixed. It is a true negative if the outcome is literal or mixed and the reading is literal or mixed. The false positives exist when the result is metonymic but the reading is literal. And finally when the outcome is literal but the reading is metonymic it is a false negative.

Table 3.4: Predicted conditions including mixed readings and outcomes

Predicted Condition	Annotation	Result
True Positive	Metonymic, Mixed	Metonymic, Mixed
True Negative	Literal, Mixed	Literal, Mixed
False Positive	Literal	Metonymic
False Negative	Metonymic	Literal

Table 3.5: Predicted conditions excluding mixed readings

Predicted Condition	Annotation	Result
True Positive	Metonymic	Metonymic
True Negative	Literal	Literal
False Positive	Literal, Mixed	Metonymic
False Negative	Metonymic, Mixed	Literal

We choose to include mixed outcomes and readings as positive cases because even

A typical CSC workstation consists of a computer with a microphone and video camera. This year, BT and IBM plan to market a # 3,000 video phone card which fits into a PC. Olivetti also plans to launch its Personal Communications Computer. This will enable users to conduct live video conversations, send faxes and use a multimedia electronic mail service -LRB- with sound, pictures and text -RRB- , and write messages or draw diagrams on an electronic white board.

Figure 3.6: *Olivetti* reading mixed in key data, but NEBMR outcome is metonymic because of verb.communication *plan*

for humans the mixed cases are not always clear to determine. Namely some mixed cases can be considered as metonymies (see Figure 3.6) and some literal readings by other human annotators. The readings of the key data correspond to human annotators which can be considered as subjective (in some cases) for that reason we felt it could be interesting to examine the algorithm on these readings also.

In another evaluation, we choose to exclude mixed readings because of the incertitude of our key data. If the reading for metonymy recognition is uncertain, for another and more accurate evaluation, we thought it might be more reasonable to exclude these ambiguous readings. The mixed outcomes can be either false positive or false negative regarding the key data reading to be respectively metonymic or literal.

We defined two predicted conditions for metonymy resolution ; TPC for true positive correct and TPI for true positive incorrect. For metonymy resolution, we only consider the true positive cases of metonymy recognition. TPC means that we correctly resolve the metonymic relation as indicated in our key data and TPI means that the metonymy resolution is incorrect.

A complete metonymy resolution consists of identifying the metonymic relation with hidden/implied concept but we content ourselves to identifying only the metonymic relation because our key data does not include implied concepts for metonymy resolution.

4 CORPUS

One of the challenging aspects of NLP is the need of human annotated corpus. Manual annotation of unstructured data is expensive in the terms of time. Besides linguistics must work alongside computer scientists for these annotations.

SemEval (Semantic Evaluation) is a continuing series to make automated semantic analysis. SemEval is derived from Senseval (Edmonds, 2002). Senseval is a corpus created for WSD.

SemEval has semantic evaluation tasks. In our work we use Task 8 of SemEval 2007 which is a task annotated for metonymy resolution. This task is an organized lexical sample for English and has 2 particular semantic class ; countries (see Table 4.1) and companies (see Table 4.2). There are 3000 country names and 1000 company names in the existing dataset. Overall 4000 sentences has been annotated in XML format. The content is provided through British National Corpus Version 1.0 (BNC). For each potential metonymy 4 sentences is framed (2 sentences before and 1 sentence after the sentence containing the PM).

Table 4.1: Test and key data annotation numbers by readings for countries, LOCATION type named entitie

Reading	Test Data
Literal	721
Mixed	20
Othermet	11
Obj-for-name	4
Obj-for-representation	0
Place-for-people	141
Place-for-event	10
Place-for-product	1
Total	908

Table 4.2: Test and key data annotation numbers by readings for companies, ORGANIZATION type named entities

Reading	Test Data
Literal	520
Mixed	60
Othermet	8
Obj-for-name	6
Obj-for-representation	0
Org-for-members	161
Org-for-event	1
Org-for-product	67
Org-for-facility	16
Org-for-index	3
Total	842

4.1 Key Data

Key data is divided into 2 groups ; key data for countries and key data for companies. There are three possible outcomes for aforementioned annotated sentences ; metonymic (see Figure 4.1), literal (see Figure 4.2) and mixed (see Figure 4.3). If the outcome is metonymic the metonymic relations are also included in annotations.

```
<sample id="samp3826">
<bnc:title> Unigram x </bnc:title>
<par>
There's a theory being nurtured in certain quarters that
Microsoft Corp Windows NT is less a strategic product than
it is a dike against Unix and that <annot><org
reading="metonymic" metotype="organisation-for-members">
Microsoft </org></annot> is trying to freeze the
marketplace long enough to bring on Cairo, the Taligent
Inc/Sun Microsystems Inc Project Distributed Objects
Everywhere-like object-oriented environment it's working
on.
Others say no, that's impossible.
</par>
</sample>
```

Figure 4.1: Metonymic reading with metonymic relation from key data for companies

```
<sample id="samp1002">
<bnc:title> Contemporary Britain: A Geographical
Perspective </bnc:title>
<par>
Kondratieff presented few data for growth from 1789 to
1814 from countries other than Britain because they had
little growth to analyse.
Parts of Belgium, Germany and France were economically
active, but real growth came with the mid-century period
of railway building.
&bquo; It seems possible to place France in the company
of those countries &mdash; <annot><location
reading="literal"> Germany </location></annot> and the US
included &mdash; which at mid century experienced a &bquo;
railway Kondratieff &equo; with emphatic growth-industry
capabilities &equo; (Trebilcock, 1981).
Having achieved modern technological growth in the period
1849 to 1873, several of these countries then consolidated
their competitive position during the &bquo; Victorian
depression &equo; of the 1880s.
</par>
</sample>
```

Figure 4.2: Literal reading from key data for countries

```
<sample id="samp1055">
<bnc:title> Keesings Contemporary Archives. April 1991
</bnc:title>
<par>
No later than six months after signature of the Treaty a
new Union constitution would be promulgated, followed by
fresh elections to the Congress of People's Deputies.
The declaration made clear that the Union Treaty need only
be signed by the nine republics which were party to the
Novoye Ogarevo meeting.
The only specified penalty for the six republics which had
not participated and were not expected to sign the Treaty
(Armenia, Estonia, Georgia, Latvia, <annot><location
reading="mixed" notes="litpeople"> Lithuania
</location></annot> and Moldavia), was that they would
thereby be excluded from a new &quot; common economic
space &quot;.
Government anti-crisis programme
</par>
</sample>
```

Figure 4.3: Mixed reading from key data for countries

4.2 Test Data

Just as key data, test data are also divided into 2 groups ; countries and companies. The difference between test and key data is test data's readings are unknown (see Figure 4.4, Figure 4.5)

```
<sample id="samp3111">
<bnc:title> Unigram x </bnc:title>
<par>
AIX Expo and the Power Solutions Conference, the
exclusively IBM shindig, has scheduled for October
20&ndash;23 at the San Jose Convention Centre in
California: they expect to change the world.
According to the latest tittle-tattle, Sparc licensees
<annot><org reading="unknown"> Fujitsu </org></annot> and
LSI Logic (who is into MIPS as well) have been talking to
DEC about Alpha.
</par>
</sample>
```

Figure 4.4: Literal reading from key data for countries

```
<sample id="samp1005">
<bnc:title> Keesings Contemporary Archives. Feburary 1990
</bnc:title>
<par>
Inflation rate 12% (1988)
Unemployment 12% (1987 est.)
Principal trading partners (1988) Exports: USA (41%), El
Salvador, West Germany; Imports: USA (39%), Mexico,
<annot><location reading="unknown"> Japan
</location></annot>
New economic programme
</par>
</sample>
```

Figure 4.5: Mixed reading from key data for countries

5 ANALYSE

5.1 Results

Generally we obtained promising results. Metonymy Recognition and Resolution results vary depending on the mixed readings and outcomes including or excluding. But again there is limitations and other cases to consider.

As said in the section about evaluation, we evaluated the algorithm in two different ways which consist of including and excluding the mixed results.

We noticed that for delimitation our false negative number is high. We think the reason is the restrictiveness and insufficiency of our rules.

When it comes to the type evaluation, we observe that only rule functions are not efficient enough to resolve metonymic relations but also we have to consider the context which means to analyse the sentences covering the sentence where the PM is present.

Table 5.1, Table 5.2, Table 5.3, Table 5.4, Figure 5.1 and Figure 5.2 correspond to an evaluation including the mixed readings.

Table 5.1: Test results for countries and companies including mixed readings

Predicted Condition	Countries	Companies	Total
True Positive	95	168	263
True Negative	614	395	1009
False Positive	78	68	146
False Negative	102	138	240

Table 5.2: Precision, Recall and Accuracy for countries and companies

	Countries	Companies	Total
Precision	0.549	0.711	0.643
Recall	0.482	0.549	0.522
Accuracy	0.797	0.732	0.767

The highness of true negative cases' (literal readings) numbers is a disadvantage for our evaluation, especially for recall formulation. Table 5.5, Table 5.6, Table 5.7, Table 5.8,

Figure 5.1: Results for countries and companies including mixed readings

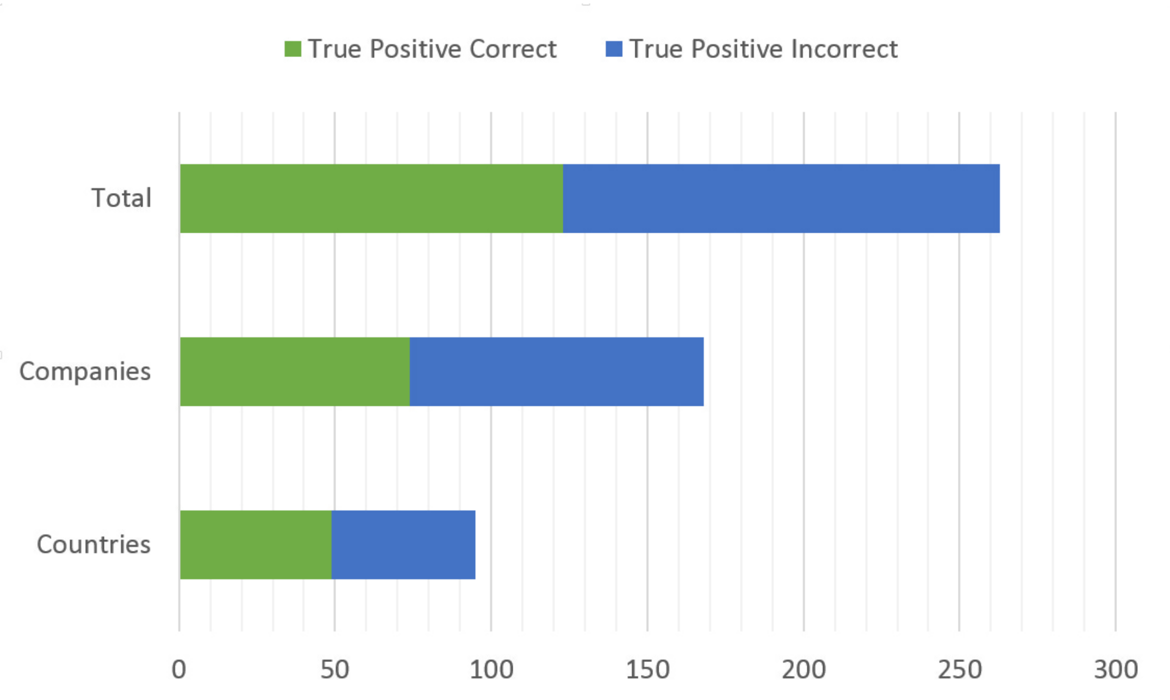

Figure 5.2: Results for metonymy resolution including mixed cases

Table 5.3: Metonymy resolution results for countries and companies including mixed readings

Predicted Condition	Countries	Companies	Total
True Positive Correct	49	74	123
True Positive Incorrect	46	94	140

Table 5.4: Accuracy for countries and companies including mixed readings

	Countries	Companies	Total
Accuracy	0.515	0.440	0.467

Figure 5.3 and Figure 5.4 correspond to an evaluation excluding the mixed readings.

Table 5.5: Test results for countries and companies excluding mixed readings

Predicted Condition	Countries	Companies	Total
True Positive	55	94	149
True Negative	596	360	956
False Positive	111	113	224
False Negative	107	145	252

Excluding the mixed readings caused regression for our precision, recall and accuracy numbers in metonymy recognition but our accuracy for metonymy resolution has increased. It is due to eliminating mixed outcomes and/or readings as true positive ; if the mixed reading is considered as true positive it is impossible to compare metonymic relation because there is none in key data. Vice versa, counting mixed outcomes as true positive or negative was adding success to our precision, recall and accuracy numbers for metonymy recognition, but counting mixed outcomes as errors, mixed outcomes came as disadvantages.

We compared our results with another algorithm using the same corpus as ourselves (see Table 5.9). While comparing results we noticed that with counting the mixed readings our metonymy recognition is better than excluding the mixed readings but metonymy resolution is better than REFERENCE. As said earlier, when the mixed readings are included, metonymy resolution failure is caused by not finding the metonymic relation for mixed readings in key data. Again, we find that some of the mixed readings were metonymic because human annotation is subjective.

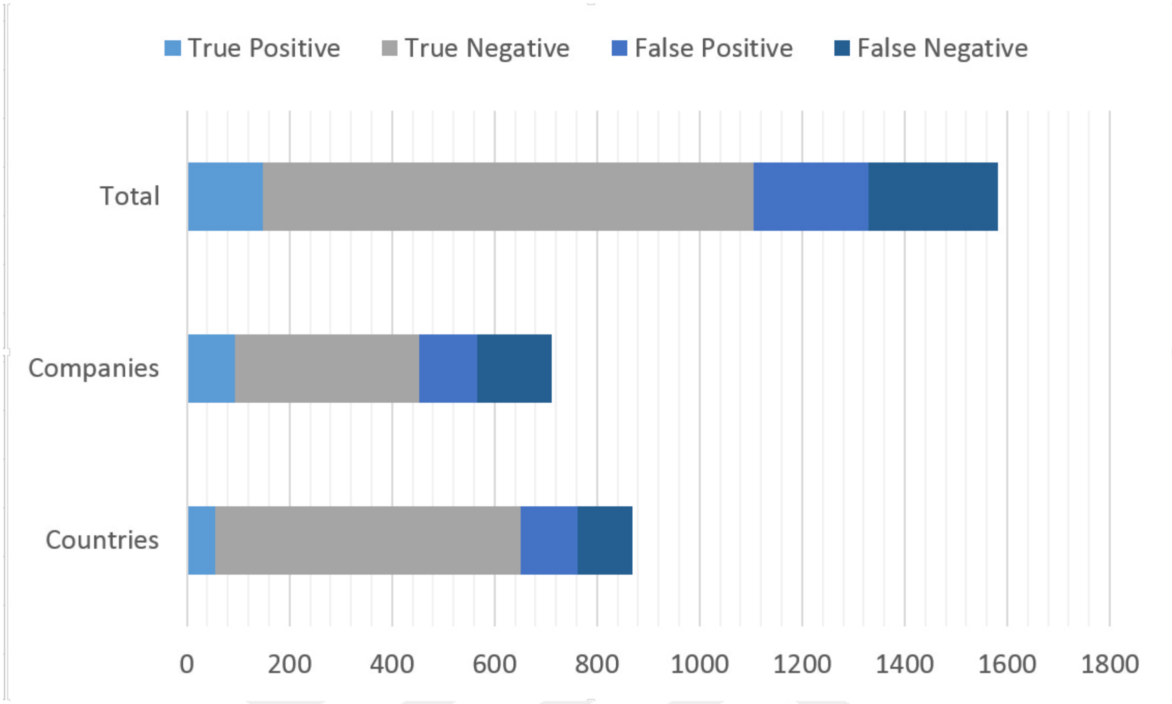

Figure 5.3: Results for countries and companies including mixed readings

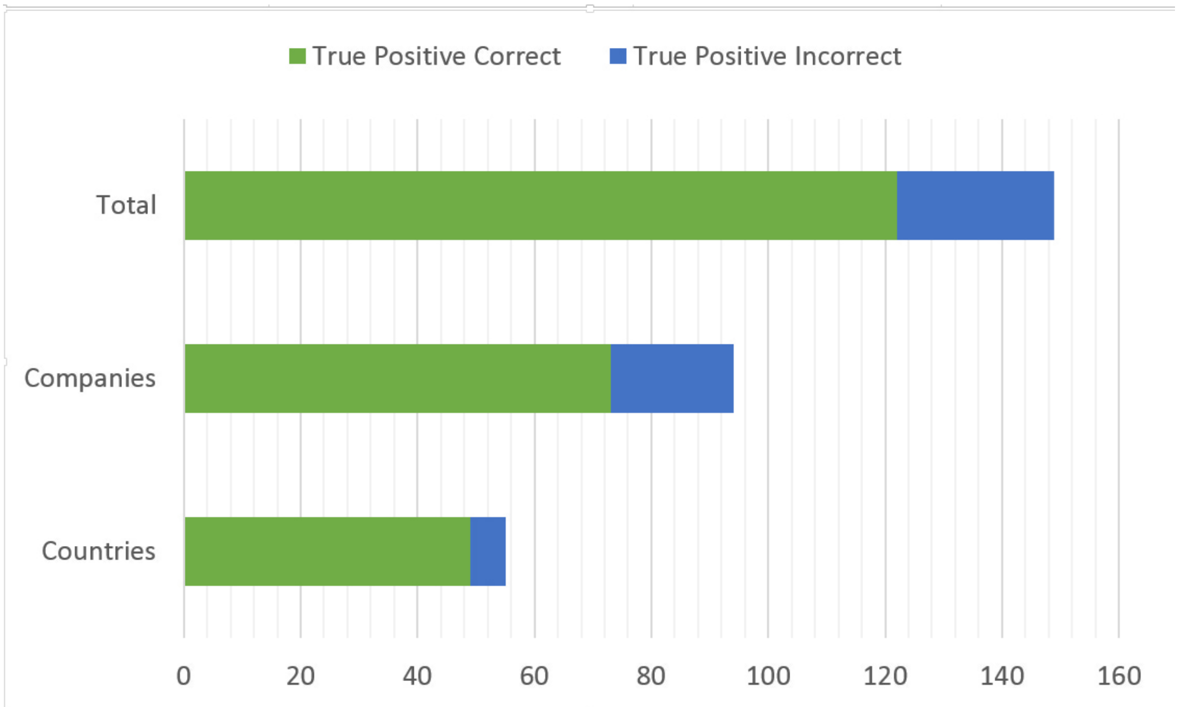

Figure 5.4: Results for metonymy resolution including mixed cases

Table 5.6: Precision, Recall and Accuracy for countries and companies

	Countries	Companies	Total
Precision	0.331	0.454	0.399
Recall	0.379	0.393	0.371
Accuracy	0.749	0.637	0.698

Table 5.7: Metonymy resolution results for countries and companies excluding mixed readings

Predicted Condition	Countries	Companies	Total
True Positive Correct	49	73	122
True Positive Incorrect	6	21	27

Table 5.8: Accuracy for countries and companies excluding mixed readings

	Countries	Companies	Total
Accuracy	0.890	0.776	0.818

Table 5.9: Accuracy scores compared with REFERENCE

Algorithm	Countries	Companies
NEBMR with mixed	0.515	0.440
NEBMR without mixed	0.890	0.776
Metonymy Resolution SemEval 2007	0.794	0.918

5.2 Limitations and Improvement Points

We have remarked that some sentences are not decently tagged. This is the case for headlines, broken or incomplete sentences. It will be wise to show special interest for this cases. And also, some named entities are not recognized by Stanford CoreNLP, it is possible to make special treatment for not recognized entities therefor it is possible to say we are limited to tools' success.

We are dependent also to corpus and thesaurus. Corpus are hard to create, they need generally a human annotator which makes the procedure long and expensive. The thesaurus we used does not include noun-adjective relations, adjective and adverb semantic categorization which is crucial for our SRVs based rules.

Again, dependency relations with prepositions, adjectives and adverbs are not considered. We propose to make a list for prepositions for further treatment and for the adverbs and adjectives the limitations depend on thesaurus.

One other improvement point is to reannotate the key data especially for mixed readings, metonymic relations and adding implied concept data for further metonymy resolution.

6 CONCLUSION

The main goal of this project is to detect metonymies and identify metonymic relations. We intended to reduce massive human work for feature vector labelling and inconsistence of statistical methods by our dependency rule-based algorithm which we apply directly to named entities. In order to achieve our main goal, we adapted the Lesk Algorithm WordNet adaption in java, we have implemented our rule functions for metonymy recognition and resolution and then we evaluated our proposed method. Existing NER tools have their advantages and disadvantages. For some named entities they can remain insufficient. Likewise, the existing base we used have some disadvantages like the success rate of the algorithms or insufficient semantic data. Obtained results are promising and they point there is still a lot of work with named entities. We had the opportunity to test our algorithm on two types because of test and key data limitations but also we propose the same method and approach for PERSON named entities. For further studies, it will be wise to considerate other types of dependencies and annotate data containing PERSON typed named entities and test our algorithm on this new data. Also, we restricted ourselves to detection of metonymic relations defined in our key data set for metonymy resolution. In order to make a complete metonymy resolution, the identification of the hidden concept must be made alongside the detection of metonymic relations. Besides our rule functions and patterns must be applied to the sentences surrounding the sentence containing the PM. This can also be considered as resolving the context or discourse.

REFERENCES

Amghar, T., Gayral, F. and Levrat, B. (1995). *Table 10 left without paying the bill! A good reason to treat metonymy with conceptual graphs*, Springer.

Banerjee, S. and Pedersen, T. (2002). An adapted lesk algorithm for word sense disambiguation using wordnet, *Computational linguistics and intelligent text processing*, Springer, pp. 136–145.

Birke, J. and Sarkar, A. (2007). Active learning for the identification of nonliteral language, *Proceedings of the Workshop on Computational Approaches to Figurative Language*, Association for Computational Linguistics, pp. 21–28.

Bogdanova, D. (2010). A framework for figurative language detection based on sense differentiation, *Proceedings of the ACL 2010 Student Research Workshop*, Association for Computational Linguistics, pp. 67–72.

De Marneffe, M.-C. and Manning, C. D. (2008). Stanford typed dependencies manual, *Technical report*, Technical report, Stanford University.

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

Ekedahl, J. and Golub, K. (2004). Word sense disambiguation using wordnet and the lesk algorithm, *Projektarbeten 2004* p. 17.

Lapata, M. (2003). Probabilistic text structuring : Experiments with sentence ordering, *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, Association for Computational Linguistics, pp. 545–552.

Leech, G. (1992). 100 million words of english : the british national corpus (bnc), *Language Research* **28**(1) : 1–13.

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries : how to tell a pine cone from an ice cream cone, *Proceedings of the 5th annual international conference on Systems documentation*, ACM, pp. 24–26.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S. and McClosky, D. (2014). The stanford corenlp natural language processing toolkit., *ACL (System Demonstrations)*, pp. 55–60.

Markert, K. and Hahn, U. (2002). Understanding metonymies in discourse, *Artificial Intelligence* **135**(1) : 145–198.

Markert, K. and Nissim, M. (2002). Metonymy resolution as a classification task, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, pp. 204–213.

Markert, K. and Nissim, M. (2007). Semeval-2007 task 08 : Metonymy resolution at semeval-2007, *Proceedings of the 4th International Workshop on Semantic Evaluations*, Association for Computational Linguistics, pp. 36–41.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. J. (1990). Introduction to wordnet : An on-line lexical database*, *International journal of lexicography* **3**(4) : 235–244.

Nastase, V., Judea, A., Markert, K. and Strube, M. (2012). Local and global context for supervised and unsupervised metonymy resolution, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, pp. 183–193.

Nissim, M. and Markert, K. (2005). Learning to buy a renault and talk to bmw : A supervised approach to conventional metonymy, *Proceedings of the 6th International Workshop on Computational Semantics, Tilburg.*

Roberts, K. and Harabagiu, S. M. (2011). Unsupervised learning of selectional restrictions and detection of argument coercions, *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 980–990.

Russell, S. and Norvig, P. (1995). Intelligent agents, *Artificial intelligence : A modern approach* pp. 31–52.

Shutova, E. (2009). Sense-based interpretation of logical metonymy using a statistical method, *Proceedings of the ACL-IJCNLP 2009 Student Research Workshop*, Association for Computational Linguistics, pp. 1–9.

Wilks, Y. (1978). Making preferences more active, *Artificial Intelligence* **11**(3) : 197–223.

BIOGRAPHICAL SKETCH

Education

Master of Science, Galatasaray University, Istanbul, Turkey, Computer Engineering, 2016
Bachelor of Science, University, Istanbul, Turkey, Computer Engineering, 2012
French Baccalauréat, Notre Dame de Sion, Istanbul, Turkey, 2004

Master Thesis

Title : Helping Metonymy Recognition and Treatment through Named Entity Recognition
Supervisors : Prof.Dr.Tankut ACARMAN, Prof.Dr.Bernard LEVRAT

Dissertation

Title : Exploitation De La Syntaxe Html Pour La Reconnaissance D'Entités Nommées
Supervisor : Asst. Prof.Dr. Vincent LABATUT

Experience

2012 - current : Managing Director, Monospect Yazılım Bilişim Hiz.Tic.Ltd., Istanbul, Turkey
2009 : Intern, Corvus İletişim Güvenlik Teknolojileri Elektronik San.Tic.Ltd., Istanbul, Turkey
2008 : Intern, Arıkanlı Holding, Istanbul, Turkey

Languages

French : Near-native
English : Advanced
Spanish : Intermediate

PUBLICATIONS

— 16th International Conference on APPLIED COMPUTER SCIENCE (ACS '16), Istanbul, Turkey