SENTIMENT ANALYSIS IN TURKISH: RESOURCES AND TECHNIQUES

by

RAHIM DEHKHARGHANI

Submitted to
the Graduate School of Engineering and Natural Sciences
in partial requirements for the degree of
Philosophy of Doctorate

SABANCI UNIVERSITY

August 2015

APPROVED BY:

Assoc. Professor Yücel Saygın ....
(Thesis Supervisor)

Assoc. Professor Berrin Yanıkoğlu ....
(Thesis Co-Supervisor)

Professor Kemal Oflazer ....

Assoc. Professor Hakan Erdoğan ....

Asst. Professor Peter Schüller ....

DATE OF APPROVAL: ....

ABSTRACT

SENTIMENT ANALYSIS IN TURKISH: RESOURCES AND TECHNIQUES

RAHIM DEHKHARGHANI

CS, PhD Dissertation, August, 2015

Thesis Supervisor: Assoc. Prof. Dr. Yücel Saygın
Thesis Co-Supervisor: Assoc. Prof. Dr. Berrin Yanıkoğlu

Keywords: Sentiment Analysis, Polarity Extraction, Polarity Lexicon, Natural Language Processing, Machine Learning, Turkish

Due to the ever-increasing amount of online information, manual processing of data is impractical. Social media such as Twitter play an important role in storing such information and helping people share their ideas. Extracting the attitude and opinion of people from user entered data is worthwhile for companies. Sentiment analysis attempts to extract the embedded polarity from a segment of text (or other data types) with many commercial and con-commercial applications.

Companies are interested in opinions of their customers. On the other hand, customers are interested in opinions of other customers. Politicians and policy makers are also interested in public's feedback on political events. The above mentioned opinions can be (semi)automatically extracted from social media such as Twitter or Facebook by the help of sentiment analysis techniques.

Sentiment analysis is a language (e.g. English) dependent task that relies on natural language processing techniques. The richest language in terms of resources and research in sentiment analysis is English, while many other languages such as Turkish suffer from a lack of resources and techniques for sentiment analysis. In this thesis, we try to fill this gap by designing and implementing a framework for sentiment analysis in Turkish. This framework can also be adapted to other languages with some minor changes. In the scope of the framework, we have built a few Turkish polarity lexicons for the first time in the literature. We also comprehensively investigated the problem of sentiment analysis in Turkish and suggested some solutions. Experimental evaluation shows the effectiveness of the proposed resources and techniques for Turkish.

ÖZET

TÜRKÇEDE DUYGU ANALİZİ: KAYNAKLAR VE TEKNİKLER

RAHIM DEHKHARGHANI

Bilgisayar bilimleri ve Mhendislii, Doktora Tezi, Ağostos, 2015

Tez Danışmanı: Doç. Dr. Yücel Saygın
Tez Eş-danışmanı: Doç. Dr. Berrin Yanıkoğlu

Anahtar Kelimeler: Duygu Analizi, Duygu Sözlüğü, Doğal Diller şleme, Duygu Sınıflandırması, Türkçe

Günlük hayattaki verilerin artış hızından dolayı, bu verilerin üzerine manual olarak analiz yapmak yöntemleri kullanışsız olmaya başlıyorlar. Sosyal media (örneği Twitter) bu alanda bilgi depolamsı ve insanlara kendi fikirlerinin paylaşması konusunda önemli bir rol oynuyamaktadır. Insanların düşüncelerini sosyal mediadan çıkarmak, şirketler için önemli bir amac sayılır. Duygu analizi metinlerin (veya diğer veri tiplerin) olumlu veya olumsuz olduklarını çıkarmaya çalişıyor. Bu işlem, ticari ve gayri-ticari bir çok alanda kullanışlı olabilir.

Şirketler kendi ürünleri ve servisleri hakkında müşterilerin yorumlarını bilmek istiyorlar. Aynı zamanda müşterilerde diğer müşterilerin fikirlerini ürünlere göre öğrenmek isterler. Başka bir örnek verilecek olursa, siyasi partilerde insanların politik olaylara karşı fikir ve düşüncelerine önem göstermek zorundadırlar. Bunların otomatik veya yarı otomatik yöntemlerle yapılmaları gerekmektedir.

Duygu analiz teknikleri her dilde o dilin yapısına göre farklılık gösterir. Diğer dillere oranla daha fazla araştırma kaynağına ve sözlüklere sahip olduğundan dolayı, bu alanda en zengin dil İngilizce olarak gösterilebilir. Yapılan araştırmaların çoğu İngilizce üzerine olduğundan dolayı, diğer diller bu alandaki araştırma kaynaklarının eksikliğini hissediyorlar. Bu nedenden dolayı Türkçe duygu analizi alanında daha fazla kaynak sunabilmek için bu doktora tezi bu konuda yapmaya karar verdik. Bu çalışmamda Türkçe duygu anlizi yapabilmek için kapsamlı bir sistem tasarlıyıp ve geliştirdik. Bu sistemde bir kaç Türkçe sözlük üretip, bunları duygu analizi yapmak için kullandık. Bunun dışında, problemi kapsamlı bir şekilde araştırıp, onu daha küçük problemlere böldük. Üzerine küçük değişiklikler yapılırsa tasarladğımız sistem, diğer diller için de kullanılabilir. Tüm problemleri bu çalışmamızda çözememiş olsak bile, her problem için farklı bir çözüm yöntemi önerdik. Elde ettiğimiz sonuçlar, uyguladığımız yöntemlerin başarılı olduğunu kanıtlamaktadır.

Benim güzel anneme ve babamın güzel ruhuna. . .
To my lovely mother and the soul of my father . . .

Acknowledgements

I would like to thank

My advisers: **Dr. Yücel Saygın**, and **Dr. Berrin Yanıkoğlu** for their scientific support.

Professor Kemal Oflazer for sharing his helpful ideas and language resources with me.

My evaluating committee members: **Dr. Hakan Erdoğan** and **Dr. Peter Schüller** for their helpful comments on this thesis.

The former and current heads of Wring Center at Sabancı University: **Mrs Nancy Karabeyoğu** and **Mr Daniel Lee Calvey** for teaching me how to write an academic paper.

Pioneers in the field of sentiment analysis: **Dr Bing Liu** and **Dr Erik Cambria** for sharing their resources/works with me.

Sabanci University Academic Support Program for funding my graduate education for four years.

My colleagues and friends: **Saeid Agahian, Arsalan Javeed, Hanefi Mercan, Gizem Gezici, and my other friends** for their friendship, help and guidance during my research years at Sabanci university.

My mother for her unending support from the beginning of my life.

Rahim Dehkharghani

Contents

List of Figures

List of Tables

Abbreviations

SA	**S**entiment **A**nalysis
OM	**O**pinion **M**ining
WN	**W**ord**N**et
SWN	**S**enti**W**ord**N**et
SN	**S**entic**N**et
MPQA	**M**ulti **P**erspective **Q**uestion **A**nswering
STN	**S**enti**T**urk**N**et
PWS	**P**olar**W**ord **S**et
WSD	**W**ord **S**ense **D**issambigution

Chapter 1

INTRODUCTION

Sentiment analysis (SA), also known as opinion mining, sentiment extraction and polarity estimation, deals with extracting the sentiment (polarity) from given data which can be in different formats e.g. video, audio, image or text. This research area has been very popular since the year 2000. The terms "sentiment analysis" and "opinion mining" have been proposed also after the year 2000 [1]. We will use the term "sentiment analysis" throughout this dissertation as the general name of this research area. In this dissertation we only deal with textual data.

SA has been using in several areas such as management, politics, marketing and psychology. Because people are related to almost all issues in the real life, this area gets more popular everyday.

Most effort in SA has been dedicated to analyse natural language texts which implies that SA strongly depends on natural language processing (NLP) area. This makes sense because the sentiment is embedded in words in a segment of text and NLP techniques extract this sentiment by analysing the text; however, advancement in NLP does not necessarily imply advancement in SA because:

- A word may have different polarities in different domains or even in the same domain. For example the word "uzun" [long] is positive for *battery life* but negative for *zooming time* in the camera domain.
- There can be polar expressions/phrases which are composed of neutral (objective) terms. This is common in idioms. Normally the polarity of an idiom cannot be extracted by using the polarity of each term included in it. For

example the Turkish idiom "göz boyamak" [deceiving] is a negative idiom while its parts "göz" [eye] and "boyamak" [colouring] are neutral terms.

- Some neutral expressions/sentences may have polar terms. For example this sentence "güzel ve verimli bir araba almak istersen ilk önce interneti araştır." [if you want to buy a good and efficeint car, search in Internet first], has two positive terms: "güzel ve verimli" [good and efficient] but it is a neutral one.

Analysing the text to extract the sentiment in each natural language requires unique techniques. For example, in order to cover the negation in English, the word "not" (is not, does not, would not etc.) should be checked in the text but in Turkish, word suffixes such as "me" in "sevmedim" [I did not like] or "sız" in "kullanışsız"[useless] should be considered.

In spite of a great demand for efficient techniques in SA, the existent research is far from perfect even in English. Some branches in SA such as spam detection (detecting the fake reviews) suffer from this gap; while the situation is even worse for non-English languages.

Our motivation for choosing this research area as the topic of this PhD dissertation is to fill the above mentioned gap in Turkish. We built a few polarity lexicons and designed and implemented a sentiment analysis system for Turkish, which are explained in the following chapters. In this chapter, we discuss about the applications and sub-problems of sentiment analysis.

1.1 Sentiment Analysis Applications

The attitude of people towards different issues in the real life is worthwhile because everybody likes to know other's opinions whenever (s)he wants to make a decision. This aim could be achieved by questionnaires in the past but due to the ever-increasing amount of information it is impractical today . After emerging the world wide web, Internet became the main source of such information. Social media such as Twitter play an important role in sharing people's ideas. People discuss almost all topics in social media, which makes it a useful platform to mine public attitude towards an issue.

Marketing companies may be the main customers of SA systems. They are interested in customers' ideas about the products or services sold/proposed. If companies can collect ideas and attitude of the customers, the quality of products/services can be improved to satisfy customers.

Politicians and policy makers are also interested in public's feedback on political events. For example, in political objection of Turkish people to the government in the year 2013, *Twitter* played an important role in reflecting the public's opinions about the mentioned topic. Moreover, Political parties can understand the attitude of people towards their party and opponent parties from social media before a political election to estimate the results.

1.2 Research Areas

The broader problem of SA can be divided into simpler and more specific sub-problems. Below, some of these sub-problems are listed.

- *Resource Generation*: Polarity resources are essential for SA because many existing approaches depend on these resources. These resources also known as polarity lexicons are list of polar terms. There exist several polarity resources in English but the majority of other languages suffer from the lack of such lexicons. There exist three methods for generating lexicons [1]: Manual methods, dictionary-based methods, and corpus-based methods. Manual methods are not popular because they are very time-consuming; other two methods are discussed in Chapter 4.

- *Spam Detection*: The possibility of posting reviews by individuals to social media and online marketing systems such as *Amazon* gives opportunity to spammers post their fake reviews. *spam* is an unfair review towards an issue, e.g a product or service; it usually exaggerates in two ways: undermining a good product or service, or advertising a low quality service or product as a high quality one. The author of *spam* reviews is called *spammer*. *Spammer* can be a person who has been hired for this purpose or a computer program. Recognizing spam reviews or spammers is a new and challenging research area. Even human being cannot always recognize fake opinions from non fake ones.

- *Cross-domain SA*: Domain in SA, is an area/topic such as Hotel, Movie, or camera domain, on which SA is applied. An approach or resource designed specifically for a domain may not work also other domains. There exist different sentiment clues–domain dependent indicative keywords–in each domain such as "izleyin" (watch it) or "izlemeli" (watchable, should be watched) in movie domain; but they cannot be used in for example hotel domain. The same situation holds for the polarity lexicons: in hotel domain, the word "küçük" (small) is negative for "oda boyutu" (room size) but in camera domain, it is positive for "pil boyutu" (battery size).

- *Cross-lingual SA*: Natural languages are the basis of SA because they should be processed to extract the embedded sentiment from words, phrases, or sentences. Cross-lingual SA attempts to extract the polarity from a text by translating it to another language. This task is always erroneous because translation task itself is not perfect. Cross-lingual SA is useful only if one language has no resource or method in SA, then it has to get help from rich languages in SA such as English.

- *SA on Twitter*: Twitter may be the first choice for many people sharing their spontaneous thoughts and reactions with others. The brevity of tweets, informal language and easy accessibility make it a popular platform. Due to the rapid and brief nature of tweets, people often make spelling mistakes as well as use special characters to express meaning and use emoticons to express feelings. Tweets require preprocessing before getting analysed by SA methods. Preprocessing may include removal of URLs and hash-tags and replacing acronyms with their extended version.

1.3 Outline of Thesis

We attempted to provide a comprehensive approach to expand the border of knowledge in SA for the Turkish language. The relation between natural language processing, sentiment analysis, Turkish and our contribution to sentiment analysis in Turkish is illustrated in Figure 1.1. Most of this dissertation is dedicated to the "Our contribution" part of this diagram. Each box of this part will be explained in each chapter with detail. In this dissertation, Chapter 2 formally defines the problem and discusses preliminaries for SA. The state of the art efforts in Turkish, English and other languages are provided in chapter 3. In chapter 4, the polarity

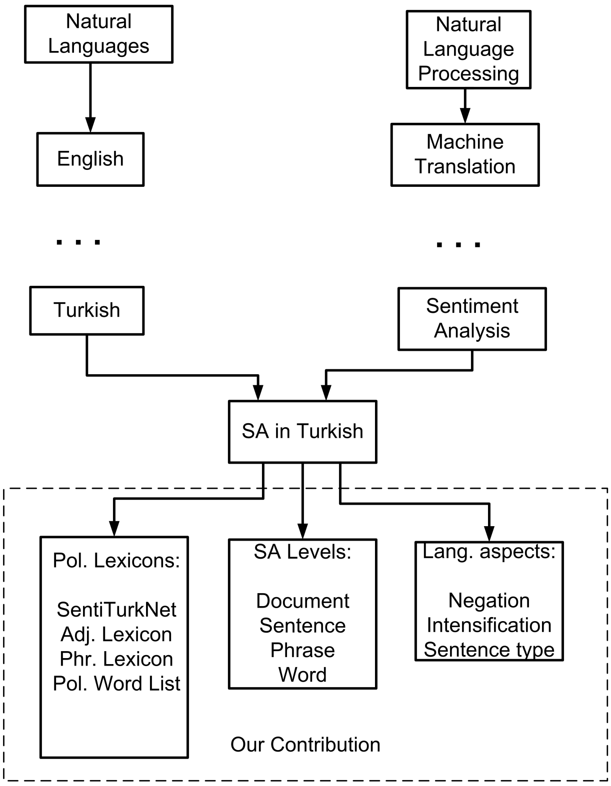

Figure 1.1: Research tree of this dissertation

lexicons and the methodologies for building them are presented. Chapter 5 discusses different levels and NLP issues in SA. Chapter 6 explains the framework that we have designed and implemented for SA in the Turkish language. Chapter 7 includes experimental evaluation and finally chapter 8 argues the conclusions and future work in Turkish SA.

Chapter 2

PROBLEM DEFINITION AND PRELIMINARIES

In this chapter, we provide a general structure for SA problem to give a big picture of what is going to be solved and which aspects of the mentioned problem are more important than the others. After formally defining the problem, active research areas in SA are explained and finally challenges of the Turkish language in SA are introduced. The problem is to extract the opinion towards a target from a segment of text, which is formally defined by Liu [2012] as:

An opinion is a quadruple S=(g, s, h, t), where s is the sentiment regarding the target g expressed by the opinion holder h at the time t.

Example: Extracting the polarity towards "oda kiralama fiyatı" (room renting price) from the sentence "Oda kiralama fiyatı otellerde daha ucuz olacak". (The renting price of rooms in hotels will be cheaper) is a simple SA problem. The target t is *Room renting price*, the sentiment is estimated based on the word *cheap*, the time and also the opinion holder are not specified in the sentence.

Having the above mentioned definition and example, we provide more explanation about some concepts and issues:

- Opinion target g is an entity such as *hotel* or an aspect of the entity such as *room renting price*. The entity can be a product, service, topic, person, event, organization etc.

- The above mentioned aspects can be explicit or implicit. Explicit aspects such as "oda kiralama fiyatı" in example above is clearly stated in the review but implicit aspects do not appear explicitly in the reviw; they rather can be extracted based on other words in the review. For example in the sentence "Bu kamera çok pahalı" (This camera is too expensive), the hidden aspect is the *price of camera* and it can be extracted based on the adjective *expensive*. We addressed only the explicit aspects in this work.

- Sentiment s can be a label such as positive, negative, or neutral, or a real number between 0 and 1 indicating the strength of positivity, negativity, or neutrality. Both cases are considered in this dissertation.

- The perspective of the reader is not included in the above mentioned definition. Perspective is the situation of the reader towards the target g. For example reducing the *room price* in hotels is positive news for travellers but probably negative for hotel owners. This issue has not been considered in this dissertation since we have only one perspective in experimented reviews: reading reviews as a company that reads its customers' ideas.

2.1 Terminology

Although we explain all new terms in their first appearance in text, here we provide a short overview on frequently used terms.

- *Opinion.* is the attitude of a person towards an issue, which has two types: regular and comparative. In regular opinion, the author states his/her opinion towards a target e.g. "Bu kamera çok iyidir" (this camera is very good) but in comparative opinion, two entities are compared e.g. "bu kamera diğer kameradan daha iyi" (this camera is better than the other one).

- *Polarity.* is a quantity indicating the positivity and negativity of a segment of text–word, phrase, sentence, or document. It can be binary or a continuous value between for 0 and 1.

- *Sentiment analysis or opinion mining.* refers to the process that extracts the polarity from data. This process is usually automatic or semi-automatic. Manual SA is possible but time consuming.

- *Objective vs subjective.* The term subjective means something that has taken place in one's mind but the term objective relates to an existing fact or reality [2]. In many papers, the term subjective and sentiment-bearing have been considered equivalent but they are actually different. A subjective sentence may not express any sentiment e.g. *I think you were in Turkey last year*; on the other hand, objective does not mean bearing no sentiment e.g. the sentence *My laptop stopped working two days after I bought it* is objective but it caries an implicit negative sentiment for the laptop.

2.2 Turkish and Its Challenges in Sentiment Analysis

Turkish is a member of the Turkic family of Altaic languages. Particular characteristics of Turkish make natural language processing (*NLP*) and SA tasks difficult for this language. Morphologically, Turkish is an agglutinative language with morphemes attaching to a root word as "beads-on-a-string". Words are formed by very productive affixations of multiple suffixes to root words, from a lexicon of about 30K root words (not counting proper names.) Nouns do not have any classes nor are there any markings of grammatical gender in morphology and syntax. When used in the context of a sentence, Turkish words can take many inflectional and derivational suffixes. It is quite common to construct words which correspond to almost a sentence in English: For example, the equivalent of the Turkish word: *"sağlamlaştırabileceksek"* in English can be expressed with the fragment *if we will be able to make [it] become strong (fortify it)* [3].

For Turkish, the morphological structure of a word is also necessary for SA in addition to the root word, as suffixes may change the polarity of a word. For instance, the word *iştahsız (having no appetite)*, is negative (due to suffix *-sız*), while its antonym, *iştahlı*, is positive (due to suffix *-lı*). Note that the root word itself, *iştah*, is also positive. This issue is handled in our system by using morphological analysis to extract and analyze suffixes of Turkish words.

Chapter 3

RELATED WORK

In this chapter we attempt to give a survey on sentiment analysis separately for English, Turkish, and other languages.

3.1 Related Work on English

There is a good deal of research on English SA because both English and non English researchers have worked on it. The most comprehensive survey in sentiment analysis are the books of Bing Liu [1] [2]. He discuss discusses almost all branches of SA problem and provides a complete survey on the topic. Below, we categorize the existent research in more popular branches and report a few work in each branch.

3.1.1 Polarity Lexicons

Polarity lexicons are language resources similar to dictionaries where instead of the sense or meaning, a polarity score or label has been assigned to each word or to a sense of word. Existing approaches to Sentiment analysis can be broadly divided into lexicon-based approaches and supervised (machine learning based) approaches. The first group of approaches benefit from sentiment lexicons. There exist a few sentiment lexicons for English which are reported below.

SentiWordNet [4] is based on Princeton WordNet [5] which assigns three polarity scores–positivity, negativity, and objectivity–to each synset (set of synonyms) in

WordNet such that their sum equals to 1. This resource has a high coverage in English because it is based on WordNet (a high coverage language resource in English) but it is somewhat noisy. The key point in building this resource was analysing the gloss (natural language explanations) of each synset. In this resource, each term has different senses and consequently different polarity scores. In order to distinguish the correct sense of a term in a context, word sense dissambiguation is required. For example the positivity, negativity, and objectivity scores of one of the adjective senses of *good* are *(P:0.75, N: 0, O: 0.25)* while these scores for one of its noun synsets are *(P:0.5, N: 0, O: 0.5)*.

SenticNet [6] assigns different numerical values to each term as its *pleasantness, attention, sensitivity, aptitude* and also the *overall polarity*. Each one of these aspects has a value between *-1* and *+1*. *-1* stands for the most negative and *+1* stands for the most positive polarities.

NRC-Emotion Lexicon [7] investigates words and expressions in terms of emotion. Not similar to above mentioned resources, this one assigns binary values to terms. It investigates each word according to the embedded emotions in it. Eight emotions are considered for each word: anger, fear, anticipation, disgust, joy, sadness, surprise, and trust. For example, the value *1* for the *joy* feature of the word *happy* means that it has the feeling of *pleasantness*.

Multi-perspective Question Answering (MPQA) [8] contains articles from a variety of news sources which have been manually annotated for opinions. This lexicon is created to support answering to opinion based questions. The method used for building MPQA is based on machine learning and rule-based subjectivity and opinion source filters. MPQA consists of three lexicons: the Subjectivity Lexicon, Subjectivity Sense Annotations, and Arguing Lexicon. These resources are available under the terms of GNU General Public License.

3.1.2 Sentiment Analysis on Twitter

Twitter is a popular microblogging and social networking website with a registered user base of around 650 millions as of 2013, which allows its users to send text messages of at most 140 characters (*tweets*). Twitter users tweet about everyday subject of life and especially in recent years, for launching political campaigns. Because of the importance of Twitter, we report some related work in this branch.

There are a few free tools on the Internet that do SA on Twitter such as [9]. sentiment140 [10]. The proposed approach in this tool uses tweets with emoticons for distant supervised learning. The authors obtained the advantage of machine learning classifiers such as Naive Bayes, Maximum Entropy, and Support Vector Machines. They also used unigrams and bigrams as features extracted from a tweet message. The authors used a method to build a data model by Twitter hash-tags. The features extracted from tweets in this work include n-grams, POS tag of words, and polar word frequency according to MPQA subjectivity lexicon. These researchers conclude that POS features are less useful than are other features such as presence of the intensifiers and the positive/negative/neutral emoticons and the abbreviations. Agarwal et al. [11] did sentiment analysis in Twitter with a different approach. The contributions of this work are introducing POS-specific prior polarity features, and also exploring the use of a tree kernel to obviate the need for tedious feature engineering. Dehkharghani and Yılmaz [12] studied the application of sentiment analysis on extracting the quality attributes of a software product based on the opinions of end-users that have been stated in microblogs such as twitter. They benefit from NLP techniques such as POS tag of words and also data mining techniques such as document frequency of words in a large number of labelled tweets.

3.1.3 Different Levels in Sentiment Analysis

The most common level in sentiment analysis is the document level. Many researchers have worked on this level to classify documents from different domains (e.g. hotel) as positive, negative, or neutral.

Pang et al. [13] investigated the document level by using machine learning approaches, Naive Bayes, maximum entropy classification, and support vector machines which were experimented on English movie reviews.

In sentence level, Meena and Prabhakar [14] investigated the sentences and their impact on document level. They also addressed the effect of conjunctions (e.g. "and" or "but"), and semantic relations between sentences in presence of such conjunctions. The highest obtained accuracy in binary classification of sentences in this work is 78%.

In aspect-level sentiment analysis, Ding et al. [15] estimated the polarity of aspects (e.g. room size in hotel domain) by analysing the polarity of neighbour words for each aspect in a window. The proposed method depends on the distance of polar words from the aspect and their sentiment strength.

There exist two well-known research in phrase level SA both by Wilson et al. [16] [17]. The authors propose an approach to phrase-level sentiment analysis that first classifies an expression as subjective or objective and then estimate its polarity in the case of subjectivity. The authors estimate the contextual polarity of an expression by using a large number of subjectivity clues and the prior polarity of appeared words in the expression. This work mostly relies on statistical methods.

Deng and Wiebe [2014] developed a graph-based model based on implicature rules to propagate sentiments among entities. The authors extract the implicitly stated sentiment by rule-based methods. For example "The bill would lower health care costs" has an implicit positive sentiment. They could increase the precision by 10 points with the help of this approach.

3.2 Related Work on Turkish

The Turkish language suffers from the lack of research and resources in SA. In terms of polarity lexicons, we (Sentiment analysis group of Sabancı university [1]) have produced four lexicons for Turkish which are explained in Chapter 4. To the best of our knowledge, no published work exists on sentiment analysis of Turkish tweets. We believe that the following papers are the only published research on Turkish sentiment analysis up to the year 2015.

Yıldırım et al. [19] accomplished a sentiment analysis task on Turkish tweets in the telecommunication domain. They applied a multi-class ternary (positive, negative, neutral) classification by support vector machines on tweets using features such as inverse document frequency, unigrams, and adjectives. They also benefit from NLP techniques such as Normalization, stemming and negation handling. The best accuracy in classifying tweets as three classes is reported as 79%. Vural et al [20] presented a framework for unsupervised sentiment analysis in Turkish text documents. They customized SentiStrength–a sentiment analysis framework on

[1] http://sentilab.sabanciuniv.edu/

English–for Turkish by translating its polarity lexicon to Turkish. SentiStrength [21] assigns a positive and a negative score to a segment of text in English. This work could achieve 76% accuracy in classifying Turkish movie reviews as positive and negative. Kaya et al. [22] investigated the Turkish political news in media. In this work, the unigrams and the bigrams together with polar Turkish terms are used as classification features, which in turn are used to train a classifier to classify unseen documents. The authors used four different classifiers: Naive Bayes, Maximum Entropy, SVM, and the character based n-gram language model, and compared their efficiency with each other. They conclude that Maximum Entropy and the n-gram language model are more efficient than SVM and Naive Bayes classifiers. The classification accuracy in different cases ranges from 65% to 77%. Aytekin [23] designed a model which assigns positive and negative polarities to text-based opinion data in Turkish blogs in order to present a general view on products and services. The model is a semi-supervised learning model based on Naive Bayes method. Training set comprises of English words stating sentiments. In order to calculate a word's probability to be in positive or negative sets, polarities are assigned to the words. Also color-word meaning correlation is provided for Turkish terms through a repetitive test-investigation process. Eroğul [24] also worked on Turkish sentiment analysis in his MSc thesis. He investigated language characteristics such as POS tag of words, bag-of-words, the unigrams, the bigrams, and negation . The structure and grammar of Turkish is also discussed in this work. Zemberek [25], as an NLP tool for Turkish, analyses the words in this work. Movie reviews are used as dataset in this thesis. The reported accuracy in classifing Turkish movie reviews as positive and negative is 85%. Boynukalın [26] worked on emotion analysis of Turkish texts by using machine learning methods. She investigated four types of emotions: joy, sadness, fear, and anger. Due to the lack of an appropriate Turkish dataset for this work, she built a new one for this purpose. The highest achieved accuracy in classifying documents into four emotions in this work is 78%.

3.3 Related Work on Other Languages

Because reporting the related work from all languages is impractical, in this section we report only one work from these languages: Chinese, Indian, German, and Spanish as four active languages in SA area.

Lin Pan [27] worked on Chinese reviews using two sets of positive and negative terms, each of which includes more than 4000 words. This work use predefined templates in sentences. It is applied on different review categories such as hotel reviews and was able to achieve accuracies higher than 85% in classifying reviews as positive and negative.

Das and Bandyopadhyay [28] propose a method for building SentiWordNet(s) for three Indian languages: Hindi, Bengali, and Telugu. The key focus in this work is translating English SentiWordNet and the Subjectivity Word List (list of polar English terms) [16] to a target language so as to build a polarity resource. They also provide a game which lets a player assign polarity values to each term.

Brooke et al. [29] investigate the problem of adapting English polarity resources to Spanish. They adapt an English semantic orientation system to Spanish and also compare it to existing approaches based on translation or machine learning methods, and show the effectiveness of proposed approach over the existent ones. For this purpose, they benefit from language aspects such as negation, intensification, and irrealis expressions.

For the German language, Remus et al. [30] built a German sentiment resource named SentimentWortschatz. It assigns positive and negative values in interval of [-1, 1] and also part of speech tags to each word, which result in over 3500 polar German words.

Chapter 4

POLARITY LEXICONS

Polarity lexicons are commonly used in estimating the sentiment polarity of a review based on the polarity of its constituent words obtained from the lexicon. There exists a good deal of work on polarity lexicon generation which is grouped by Liu [2012] into two categories: lexicon-based methods and Corpus-Based methods. Lexicon-Based methods start with a small seed word list and expand it upon synonymy and antonymy relations by using dictionaries such as WordNet [5]. In Corpus-Based methods, semantic relations between terms in a corpus are employed to generate polar terms. These relations include pointwise mutual information [31] considering the co-occurrence of words in a window (e.g. a sentence), conjoined adjectives (by “and”, “but”) [32], and delta *tf-idf* [33]. All three polarity resources that we have built and explained in this chapter, benefit from a hybrid methodology that consists of both lexicon-based and corpus-base methods.

In lexicon-based approaches, dictionaries such as WordNet play the main role. These methods start with a small seed set (e.g. 20 terms) and expand the list by using existing relations–such as synonymy and antonymy–among terms in dictionaries. Hu and Liu [2004] used this method to generate a list of polar English terms and then manually cleaned up the generated list to remove errors. The same approach was used by Dehkharghani et al. [2015] to build a polarity lexicon for Turkish (Section 4.2.1). A similar approach was proposed by Kim and Hovy [36] which assigns also a sentiment score to each word by using a probabilistic method.

In corpus-based approaches, having a seed list of words with known polarity and a linguistic corpus, new polar words are extracted based on the existing semantic relations in the corpus. One of the early ideas was proposed by Hatzivassiloglou and McKeown [1997]. The authors used conjunctions in a corpus to find new polar adjectives. They showed that conjoined adjectives by "and" usually have the same polarity while they will have the opposite polarity when conjoined by "but". Some extra relations such as "Either-or" and "Neither-nor" were also used for this purpose. This assumption holds also for Turkish as experimented in the current dissertation. Kanayama and Nasukawa [2006] followed this approach and improved it by adding the idea of consecutive sentences usually have the same polarity.

Another popular method was proposed by Turney [2002] by introducing the Pointwise Mutual Information (*PMI*) concept. He computed the *PMI* score of adjectives with "excellent" as a pure positive and with "poor" as a pure negative word co-occurred in a sequence of words as a window. Wu and Wen [38] dealt with the problem of comparative sentences in Chinese by relying on the proposed method by Turney and also Web search hit counts.

Apart from the above mentioned categorization, polarity lexicons can be divided into domain-independent (general-purpose) and domain-specific. General-purpose polarity lexicons such as SentiWordNet [39] are domain-independent and have the shortcomings that they do not capture sentiment variations across different domains or cultures, nor can they handle the changing aspects of the language; however, these lexicons do provide a fast and scalable approach to sentiment analysis.

A typical example for the shortcomings of domain-independent polarity lexicons is the term "big" that is positive for *room size* in the *hotel* domain but negative when referring to the *battery size* in the *camera* domain. As for cultural–dependence, one can give the example of the noun "Atatürk" (a former Turkish leader) which is mostly positive in Turkish culture, while it may be neutral in others. In order to solve these issues, domain-dependent and language-dependent (or culturally-dependent) lexicons are required. Another issue is that while languages are changing, polarity resources also need to be updated to reflect the changes. However doing so manually is time consuming, costly and open for bias. Finally, the polarity of an idiomatic phrase may differ from the polarity of its parts. For example, "costing an arm and a leg" has a negative sentiment while no single word has negative polarity in the phrase. Hence, a polarity lexicon should handle idioms separately.

he domain dependence problem is addressed by some researchers as an adaptation problem where a general purpose polarity lexicon is adapted to a specific domain using some domain-specific data [40]. Others have worked on constructing a lexicon in a given domain starting from a seed word set [41].

Numerous polarity resources already exist for English, e.g., SentiWordNet *(SWN)* [42], SenticNet *(SN)* [6], and *NRC* Emotion Lexicon [7]. On the other hand, the absence of polarity resources in many other languages such as Turkish, hampers the development of sentiment analysis tools and applications in these languages. In order to close this gap in Turkish, we have undertaken the development of some polarity resources for Turkish.

A simple approach for building polarity resources for non-English languages has been to translate available polarity resources from English. The reason why we did not take the same approach and translate English lexicons such as SentiWordNet to Turkish is two-fold:

- Meaning between languages is often lost in translation. Translating a Turkish word into an English word only implies that this English word is the closest term in English for the given Turkish word, rather than their meaning being equivalent. Indeed, the meaning of many words only exist within a native context: The Turkish word *"gönül"* which is translated to English as "heart/soul/feelings" lacks a single equivalent term in English.
- Translation of meaning does not necessarily correspond to translation of the polarity strength in language dependent terms. For example, "Tanrı" [*God*] is a positive term in Turkish although the term may be objective in another language. Indeed, polarity scores given in SentiWordNet for the synset "supreme-being, God" are *(pos, neg, obj)=(0, 0, 1)*, supporting this observation.

In this chapter, we propose three semi-automatic methods for building polarity lexicons and specialize them for the Turkish language. Although we applied the proposed methodologies on Turkish, our methods are language independent and can be applied on other languages.

In the next section, we propose the first methodology for building the first polarity resource for Turkish named SntiTurkNet which is based on WoedNet.

4.1 SentiTurkNet

SentiTurkNet [35] is the largest and first polarity lexicon for Turkish that we have built. A few polarity resources have been used in building SentiTurkNet which are listed below.

4.1.1 English Resources

We have used the following three English resources during the construction of SentiTurkNet.

- *English WordNet* [5]: This lexical resource groups synonym terms in a set called *synset* that includes a *gloss* (natural language explanation) for each synset. There are about 117,000 synsets in English WordNet.

- *SentiWordNet* [39] : This resource is built with the purpose of supporting sentiment analysis tasks in English. Three polarity scores summing to one are assigned, indicating the positivity, negativity, and objectivity of each English Wordnet synset.

- *SenticNet* [6]: This resource assigns numerical values to each term according to its *pleasantness, attention, sensitivity, aptitude* and also the *overall polarity strength*. We have translated this resource to Turkish by a bilingual dictionary [1] and used the *overall polarity strength* as features in our algorithm.

4.1.2 Turkish Resources

We have used only one Turkish resource in this work: Turkish WordNet [43]. This resource consists of about 15,000 synsets along with the gloss, equivalent English synset, POS tag and so on [43]. Each synset includes these fields:

- *Synonyms* are the synonym terms in a synset.

[1] http://www.seslisozluk.net

- *Gloss* is the Turkish gloss for the synonym list. *Gloss* is not available for all synsets; therefore we added them some explanations from the TDK (Turkish Language Organization) monolingual dictionary [2].
- *Synset ID* is a unique identifier for each synset.
- *ILI ID* is the Interlingual Index used for mapping the Turkish synset to its equivalent English synset in English WordNet.
- *POS tag* is the part of speech tag of the terms in the synset –noun, verb, adverb, or adjective.
- *Hypernym synset ID* is the synset *ID* of the hypernym synset (denoting a more general concept). This ID is not available for all synsets; therefore we used only those available.
- *Near-antonym synset ID* is the synset *ID* of the near-antonym synset. This ID is not available for all synsets; therefore we used only those available.

A sample entry from Turkish WordNet is provided in the top part of Table 4.1. The bottom part shows information derived from the manual labelling (Section 4.2.2) and WordNet mapping (Section 4.1.3).

TABLE 4.1: **A synset from the Turkish Wordnet extended with sentiment polarity and English correspondent information (below the line)**

field	**value**
Synonyms	*güzelleştirmek, süslemek*
Gloss	*daha güzel hale getirmek*
POS tag	*Verb*
Synset label	*Pos*
Hypernym synset label	*Pos*
Near-antonym synset label	*Neg*
Equivalent English synset	*ameliorate, improve, better, amend...*

In the original version of Turkish WordNet, some of the synsets do not have Turkish gloss. As our approach requires this gloss, we extracted Turkish explanations for synsets from a Turkish dictionary (TDK). This mono-lingual dictionary consists of over 80,000 entries.

[2] http://www.tdk.gov.tr

4.1.3 WordNet Mapping

Turkish Wordnet has been already mapped (one to one) to English WordNet by using the *ILI*s (Inter-Lingual Identifiers). In this mapping, some Turkish synsets have a mapping to English WordNet v2.0 and some others to WordNet v2.1. Since all synsets among different versions of English WordNet have been mapped to each other, we used the existing mappings between Turkish to English synsets, to map the Turkish WordNet to English WordNet 3.0.

As SentiWordNet 3.0 is based on WordNet 3.0, we could extract the polarity scores of the equivalent English synset of each Turkish synset from SentiWordNet. These polarity scores are used as two features in Section 6.1.4.

4.2 Building SentiTurkNet

The problem addressed in this work is to build a polarity lexicon for Turkish, indicating the polarity scores for all (14,795) the synsets in the Turkish WordNet. The assigned polarity scores are triplets indicating the positivity, negativity, and objectivity strength of each synset, summing to 1 as in SentiWordNet.

The proposed methodology starts manually assigning one of the three polarity classes (positive, objective/neutral, or negative) to each one of the synsets. Note that this is a relatively easy step compared to the ultimate goal of assigning sentiment polarities to each synset, not just class labels.

After the manual labelling, we extract various features about the synsets from the resources indicated in Sections 4.1.1 and 4.1.2. The extracted features include some characteristics of the synonyms and gloss of the synset, as indicated by different resources. We then build a classifier to learn this classification given the features extracted from the synsets. In other words, the classifier learns the mapping from extracted features to polarity classes and once it is trained, the confidence scores returned by the classifier for a given synset s_i are used as the polarity strength values $pos(s_i), obj(s_i), neg(s_i)$.

The process is illustrated in Figure 4.2 and can be summarized in four steps that are explained in the following subsections:

- Step 1: Manually labelling all synsets in Turkish WordNet as positive, negative, or objective (Section 4.2.2).
- Step 2: Extracting features related to each synset (Section 6.1.4).
- Step 3: Learning the mapping between synsets described by the extracted features and the three class labels (positive, negative, objective/neutral) through machine learning techniques (Section 4.2.4).
- Step 4: Combining output of the classifiers to obtain more accurate results. (Section 4.2.5)

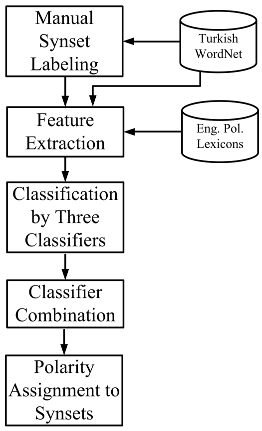

FIGURE 4.1: **Flow diagram of the proposed methodology for building SentiTurkNet**

4.2.1 Resource Generation

In addition to the resources mentioned in Section 4.1.2, we developed and used two small polarity lexicons in extracting features for the classification.

Polar Word Set (PWS): We have semi-automatically generated a list of polar Turkish terms including 1000 positive and 1000 negative terms using the method proposed by Hu and Liu [2004]. This method uses the synonymy and antonymy relations between terms to generate a large polar word set starting from a small seed set.

Polar words with PMI scores: We have assigned polarity scores to each word in PWS using Pairwise Mutual Information (PMI) score between that word and pure positive or negative Turkish words listed in Table 4.2.

Table 4.2: **Pure positive and pure negative Turkish words used in the *PMI* formula**

Pos.	*harika (excellent), güzel (beautiful/fine), mükemmel (perfect), sevgi (love), inanılmaz (unbelievable), muhteşem (gorgeous), iyi (good), şahane (fantastic), hayırlı(good), olumlu(positive)*
Neg.	*berbat (terrible), korkunç (terrible), iğrenc (disgusting), rezil (abject), felaket (disaster), kötü (bad), yetersiz (inadequate), üzgün (sad), fena (bad), olumsuz (negative)*

The *PMI* concept was first introduced by Turney [2002]. Our *PMI* scores are calculated according to co-occurrence of two terms in a database of 10,000 Turkish sentences that have been manually labelled as positive, negative, or objective (neutral). The *PMI* score of two terms t_1 and t_2 is given in Equation 4.1.

$$PMI(w_i, w_j) = \frac{P(w_i, w_j)}{P(w_i) * P(w_j)} \tag{4.1}$$

where $P(w_i)$ is the probability of seeing w_i in the above mentioned 10,000 labelled Turkish sentences. Similarly $P(w_i, w_j)$ is the probability of seeing w_i and w_j in a sentence (as a window) in the same database.

In our case, w_i is each one of the polar words in PWS and w_j is a pure positive or negative word in Table 4.2. Note that a higher PMI score between the term w_i and positive (or negative) terms indicates a higher positive (or negative) polarity for w_i.

We calculate the PMI score of each word, w_i, in PWS with ten pure positive words and assign the average of these scores to w_i as its positivity score (Equation 4.2). The negativity score (*NegPMI*) is computed in similar way by using the ten pure negative word list.

$$PosPMI(w_i) = \frac{\sum_{w_j \in PurePos} PMI(w_i, w_j)}{10} \tag{4.2}$$

where *PurePos* is the above mentioned ten pure positive word list in Table 4.2.

The word w_i is then assumed to be positive according to the PMI scores, if $PosPMI(w_i)$ is greater than its $NegPMI(w_i)$.

4.2.2 Manual Labelling of the Polarity Lexicon

As the first step, all 14,795 synsets in the Turkish WordNet are manually labelled (each synset by one person) to indicate only their polarity *class* as positive, negative, or objective. The manual labelling is done by native Turkish speakers. Labelling the synsets in this simple manner, *without* assigning polarity strengths, is needed to train the classifier, whose output scores are then used as polarity values.

In order to evaluate the labelling task, we randomly chose 10% of synsets and asked two more native speakers to label them; then we compared three labels assigned to each synset. As a result, labels in 87% of synsets were agreed by three labellers; and in 13% of labels, only two persons agreed on the assigned label.

Table 4.3: **Features are extracted for each synset using SenticNet (SN), PolarWordSet (PWS) and SentiWordNet (SWN).**

Feature name
f_1: *Avg. polarity of pos. synonyms based on PMI*
f_2: *Avg. polarity of neg. synonyms based on PMI*
f_3: *Avg. polarity of pos. synonyms based on SN*
f_4: *Avg. polarity of neg. synonyms based on SN*
f_5: *Number of pos synonyms based on PWS*
f_6: *Number of neg. synonyms based on PWS*
f_7: *Number of synonyms that are adjectives*
f_8: *POS tag of the synset*
f_9: *Number of capitalized synonyms*
f_{10}: *Number of pos. synonyms in gloss according to PWS*
f_{11}: *Number of neg. synonyms in gloss according to PWS*
f_{12}: *Avg. polarity of pos. terms in gloss based on PMI*
f_{13}: *Avg. polarity of neg. terms in gloss based on PMI*
f_{14}: *Avg. polarity of pos. terms in gloss based on SN*
f_{15}: *Avg. polarity of neg. terms in gloss based on SN*
f_{16}: *Number of pos. terms in gloss based on PWS*
f_{17}: *Number of neg. terms in gloss based on PWS*
f_{18}: *Number of adjectives in gloss*
f_{19}: *Number of capitalized terms in gloss*
f_{20}: *Pos. score of equivalent synset in SWN*
f_{21}: *Neg. score of equivalent synset in SWN*
f_{22}: *Label of hypernym synset*
f_{23}: *Label of near-antonym synset*

4.2.3 Feature Extraction

We extract 23 features shown in Table 6.5 for each synset. The extracted features include some characteristics (e.g. average polarity) of the synonyms and gloss of the synset, as indicated by different resources.

Before feature extraction, the gloss of each synsets are tokenized, then each token is stemmed to extract its root word and suffixes.

- $f_1 - f_4$: The first four features compute the average polarity scores of synonyms in a synset using different resources. The first two features are the average PMI score of positive and negative terms, as classified according to their PosPMI and NegPMI scores. The next pair of features uses the polarity scores of SenticNet. In SenticNet, we assume a term (or phrase) is positive if its polarity score is greater than or equal to zero or as negative otherwise. Note that simply using the average polarity of all synonyms would require also using the purity measure. We take a different and more symmetric approach and use the average polarity of positive and negative synonyms separately.
- $f_5 - f_6$: These features capture the frequency of positive and negative polar terms in each synset according to PWS.
- $f_7 - f_9$: These features cover certain characteristics of synonyms. f_7 captures the number of synonyms in a synset that are adjective. Generally, those synsets with higher number of adjectives are more subjective. Adverbs are not considered in f_7 because less than 1% of the synsets are tagged as adverbs. f_8 captures the part of speech tag of the synset. The rationale behind f_8 is that adjective and adverb synsets have a tendency to be more subjective than do noun or verb synsets. f_8 is different from f_7 in that, some synsets tagged as adjective have non-adjective synonyms. f_9 is the number of synonyms that start with a capital letter. These synonyms (generally proper nouns) are most probably objective e.g. "Milli Gvenlik Kurulu" (National Security Corporation).
- $f_{10} - f_{11}$: Similar to $f_5 - f_6$, this pair represents the frequency of positive and negative polar terms in a gloss.

- $f_{12} - f_{15}$: Similar to $f_1 - f_2$, this set computes the average polarity scores of the terms (unigrams and bigrams) in a gloss.
- $f_{16} - f_{17}$: Similar to $f_5 - f_6$, this pair represents the frequency of polar terms in a gloss.
- $f_{18} - f_{19}$: Similar to f_7 and f_9, these features represent the number of adjectives and (first letter) capitalized terms in gloss.
- $f_{20} - f_{21}$: This pair indicates the positivity and negativity scores of equivalent English synset (in SentiWordNet). The result of WordNet mapping between English and Turkish is utilized in this set.
- $f_{22} - f_{23}$: The polarity (label) of hypernym and near-antonym synsets of a given synset is indicated by these features. Most of the synsets in Turkish WordNet have hypernymy and near-antonymy relations with other synsets which can be used to estimate the polarity of the given synset. Some synsets in Turkish WordNet lack the hypernymy or near-antonymy relations; if these relations are not available, a default value (e.g. -1) is assigned to f_{22} and f_{23}.

4.2.4 Synset Classification

We trained three different classifiers to learn the mapping between features and polarity classes: Logistic Regression (*LR*) [44], Feed-forward Neural Networks (*NN*) [45], and Support Vector Machine with sequential minimal optimization algorithm (*SMO*) [46]. These three classifiers are some of the most commonly used classifiers for various reasons, such as good generalization accuracy (SVM, NN) and simplicity and computing posterior probabilities (LR). We used *Weka 3.6* [47] for implementing these classifiers.

4.2.5 Classifier Combination

After training the base classifiers, we used a classifier combination method called *stacking*, to learn how to combine the individual classifier results. Classifier combination is a commonly used technique for improving generalization accuracy [48].

In this approach, the output of these three base classifiers are given as input to a final classifier which learns to map them to the desired polarity classes.

In our case, the training set of the new classifier receives input samples that consist of confidence scores obtained from three base classifier as features ($3 \times 3 = 9$ features), along with the label (the known polarity class of the corresponding synset). During testing, given a synset, the classifier assigns different confidence values to each of the three classes; we then interpret the output o_i as the polarity strength of the synset for the corresponding class i (positive, negative, and objective). Classifier combination brought an increase of 8% percentage points in classification accuracy, over the base classifiers.

4.2.6 Example

In Table 4.4, we provide a real example for the proposed methodology. The top part of the table shows the information obtained from the extended Turkish WordNet, while the bottom part shows the scores assigned by mapping from SentiWordNet and the proposed method. For the latter, we give the results of the three base classifiers and the combination (indicated as SentiTurkNet score). As can be seen with this language/cultural dependent synset, the result of the proposed method is in accordance with the term that is accepted as mostly positive in Turkish. On the other hand, polarities obtained from translations from SentiWordNet indicate it as objective (neutral).

4.2.7 Summary and Contributions

The two contributions of this work are building the first comprehensive polarity lexicon for Turkish (SentiTurkNet) and proposing a semi-automatic approach to do this for other languages as well. The developed lexicon contains polarity score triplets for all synsets in the Turkish WordNet, containing almost *15,000* synsets. SentiTurkNet is thus based on Turkish WordNet and is mapped (one to one) to English WordNet and consequently to SentiWordNet.

The quality of the lexicon is established using different approaches, including low mean absolute error between the estimated and the manually assigned polarities for a small portion of the lexicon for which a groundtruth exists. Furthermore,

Table 4.4: **An entry from SentiTurkNet, together with assigned polarities.**

field	value
Synonyms	*Cuma namazi* [Friday Prayers]
Gloss	*Müslümanların Cuma günleri yaptığı ibadet* [Worship muslims perform on Friday]
POS tag	*Noun*
Synset label	*Pos*
Hypernym synset label	*Pos*
Near-antonym synset label	*Not specified*
Equivalent English synset	*salat, salah, salaat...*
SentiWordNet scores	***(P, O, N)=(0,1,0)***
score by NN	*(P, O, N)=(0.52,0.45,0.02)*
score by LR	*(P, O, N)=(0.54,0.45,0.01)*
score by SMO	*(P, O, N)=(0.33,0.66,0.01)*
SentiTurkNet scores	***(P, O, N)=(0.49,0.44,0.06)***
SentiTurkNet label	***Pos***

we showed that the use of the generated lexicon results in higher classification accuracy in sentiment classification, compared to using translated resources.

The shortcoming of the developed lexicon is its relatively small coverage size. As for the proposed methodology, it is applicable to any language for which a WordNet exists, but it is time consuming to manually label the polarity classes of the synsets.

Here we compare SentiTurkNet with SentiWordNet because it is the most similar resource to SentiTurkNet and the main idea for building SentiTurkNet has been derived from SentiWordNet. The similarities and differences are as follows:

- Both resources benefit from the polarity of the gloss of a synset as a feature to estimate the polarity scores for the synset.
- Both resources assign polarity scores to each synset in WordNets of different languages such that the sum of these scores equals to one.
- English WordNet (and consequently SentiWordNet) has around 117,000 synsets while Turkish WordNet (and SentiTurkNet) has 15,000 synsets.
- In SentiWordNet, the polarity level of a synset is estimated as one of eight categories; hence, polarity scores in SentiWordNet are multiples of 0.125, while the polarity scores in SentiTurkNet are continuous values in [0, 1].

4.3 Adjective Polarity Lexicon Generation

In this section, another polarity lexicon and the methodology used for building this resource is explained. As mentioned earlier, proposed methods for polarity lexicon generation are grouped by Liu [2012] into two categories: Lexicon-Based methods and Corpus-Based methods.

The above mentioned methods have been separately used in the literature; however, they could be combined to design a more effective approach which has been accomplished in this work. Each method contributes to our hybrid method as a classification feature in classifying adjectives as positive, negative, or neutral. Experimental evaluation approves the effectiveness of the hybrid approach when compared to each method in isolation.

In spite of the existing work, the current work differs from them in its hybrid approach, input and output. Moreover, despite the good deal of work in polarity lexicon generation for English, there are only two previous attempt for Turkish [35] [23]. We expanded our previous work by the current one which results in first adjective polarity lexicons for Turkish.

In order to generate an adjective polarity lexicon, we downloaded a list of 11,000 Turkish adjectives from an online Turkish lexicon [3]. Note that we covered unigrams and bigrams (adjective phrases) which are very scarce compared to unigrams. A bigram adjective (adjective phrase) is composed of two words appearing together as an adjective e.g. "akla yatkın" (advisable). Our methodology differs from the existing research in that it receives a list of raw adjectives as input and classifies them as three classes (positive, negative, and neutral) while the existing approaches extract these adjectives from linguistic corpora or lexicons. Different methods have been used in adjective classification, each of which contributes to the classification tasks as a feature.

4.3.1 Classification Features for Adjectives

In this section, we introduce a few polarity estimator methods, which are used as features in classifying adjectives into polarity classes.

[3] *http://tr.wiktionary.org*

- Pointwise Mutual Information (PMI): This method captures the co-occurrence of two terms in a corpus. The main idea is that positive terms generally co-occur with positive adjectives and negative ones co-occur with negative adjectives. This concept was first proposed by Turney [2002] to extract the co-occurrence of terms with two positive and negative words: *excellent* and *poor*. He proposed an equation (4.3) for computing the PMI score of two terms.

$$PMI(w_1, w_2) = log_2\left(\frac{P(w_1, w_2)}{P(w_1) \times P(w_2)}\right) \tag{4.3}$$

  $P(w_1)$ is the probability of seeing w_1 and $P(w_1, w_2)$ is the probability of seeing both w_1 and w_1 in a specified window. We computed the average PMI value of each adjective with 1,000 positive and 1,000 negative words that we had already generated for Turkish [35]. This co-occurrence is searched among 270,000 Turkish sentences in Turkish movie reviews [4] as the corpus.

- Delta *tf-idf*: In this technique, the *tf-idf* (Term Frequency-Inverse Document Frequency) score of an adjective in positive sentences is subtracted from its *tf-idf* score in negative sentences. Equations 4.4 and 4.5 are used for computing the *tf-idf* score of an adjective in a set of documents.

$$tfidf(adj, s, S) = tf(adj, s) \times idf(adj, S) \tag{4.4}$$

$$idf(adj, S) = -\log(\frac{N}{\{|s \in S, adj \in s\}|}) \tag{4.5}$$

  adj stands for a given adjective, s for sentence and S for a dataset of sentences. We assumed that $tf(adj, s)$ has a binary value. If an adjective appears several times in a sentence (unlikely), still we suppose $tf(adj, s)$ as 1. This feature has been experimented on about 6000 manually labelled sentence extracted from Turkish Movie Reviews and also Twitter.

- Translating to English: In this feature, we translated all adjectives to English by a bilingual dictionary [49] and extracted first three English translations of each Turkish adjective. Then we searched these English words in three English polarity lexicons: Polar word set generated by Hu and Liu [2004], SentiWordNet [42], and SenticNet [6], and checked their polarity label/score in these lexicons. Polar word set has already separated positive list from the negative one. In SentiWordNet, a word is assumed as positive if the

[4] This dataset is collected from www.beyazperde.com

average positive polarity of all synsets of the word disambiguated by parts of speech tags is higher than its negative score. We did not go more deeply into Word Sense Disambiguation (WSD) problem. In SenticNet, if the overal polarity score of a word is positive (or negative), we assumed it as a positive (or negative). Note that the weight of the i_{th} translation is higher than the $i+1_{th}$ translation. Finally a Turkish word is labelled as positive (or negative), if English polarity lexicons label it as positive (or negative) by using the majority voting method. This feature has been used as the baseline for adjective polarity lexicon generation.

- Hit number in Google: In this feature, the expressions "*adj* ve güzel" [*adj* and good/beautiful], and "*adj* ve kötü" [*adj* and bad] are searched in Google search engine, where *adj* is an adjective in the adjective list. As conjoined adjectives by "ve" [and] generally have same polarity, an adjective is expected to be positive (or negative), if its hit number in Google for the clause "adj ve güzel" is greater than that of the clause "adj ve kötü". Equation 4.6 is used for this purpose. *hit(clause)* gives the number of hits in Google returned for the searched *clause*.

$$DeltaHit(adj) = \log(hit(adj\ ve\ güzel) - hit(adj\ ve\ kötü)) \tag{4.6}$$

Table 4.5 lists the classification features explained above, plus linguistic techniques (conjunctions and suffixes) for classifying the adjectives.

TABLE 4.5: **Classification features and linguistic techniques for classifying adjectives.**

Classification Features
Delta *tf-idf*
Hit number in Google
Translating to English
Pointwise mutual information
Linguistic Techniques
Conjunctions
Suffixes

4.3.2 Classification of Adjectives

In this phase, suggested features in Section 4.3.1 are combined to train a classifier. For this purpose, we manually labelled 1100 (10% of all data) adjectives as positive, negative, or neutral and fed their feature vectors as well as their polarity label to the classifier. Then the polarity of each adjective in the test set (about 1,0000 terms) is estimated by classifying it as one of the above mentioned three classes. At the end of this phase, about 1500 positive, 1200 negative, and 7300 neutral adjectives are obtained. The classifier used in this step is logistic regression [44]; evaluation method is 5-fold cross validation on training data; and the classification tool is WEKA [47]. A correctly classified positive adjective is "zevkli" [pleasant] and a positive adjective which is incorrectly classified as negative is "fantastik" [fantastic] . Afterwards, we expanded the obtained polarity lexicon by two linguistic techniques explained in the following subsection.

4.3.3 Improvement Phase on Classification

This phase consists of two tasks: (1) adding new polar adjectives by using conjunctions, and (2) adding new polar adjectives by adding/removing suffixes to/from already generated adjectives.

4.3.3.1 Conjunctions for adjective extraction

As mentioned earlier, conjoined adjectives by "ve" [and] are expected to have same polarity; however, they will most probably have opposite polarity when conjuncted by "ama" [but]. Using this method, we extracted all conjunctions from 270,000 Turkish sentences in Turkish movie reviews. We extracted new polar adjectives based on patterns listed in Table A.5. In this Table, if the polarity of an adjective

TABLE 4.6: **Patterns used for extracting new polar adjectives.**

Negative Adjective extraction	Positive adjective extraction
adj and *NegAdj*	*adj* and *PosAdj*
adj ama *PosAdj*	*adj* ama *NegAdj*

(*adj*) in one side of the conjunction is unknown to our system, and the other side (*NegAdj* or *PosAdj*) is known, a polarity tag is assigned to *adj* based on the above mentioned patterns. In other words, the unknown adjectives in the right hand side of this table are supposed to be positive while the extracted adjectives by the left hand side patterns are expected to be negative. This technique could generate only about 100 new positive and 100 new negative adjectives that did not already exist in the polarity lexicon.

4.3.3.2 Using suffixes for adjective extraction

In Turkish, as an agglutinative language, suffixes can be added to the root word to build new words, such as adding suffixes to nouns to generate adjectives. A suffix can also change the polarity of word; the Turkish noun "kullanış" [usage], for example, is neutral and its polarity changes to negative by adding the suffix "sız" [-less]: "kullanışsız" [useless], while due to the suffix "lı" [-ful], its antonym, "kullanışlı" [useful], has a positive polarity.

One method for assigning a polarity tag to a Turkish word is to decompose it into the root word and suffixes. Then the root word is searched in polarity lexicons and the word polarity is changed if suffixes shift the polarity of the root word. For example, "sevgisiz" [loveless] has the root "sev" [love (infinitive verb form)], and the suffix "gi" transforms the verb to noun with the same meaning and polarity but the suffix "siz" [without] transforms the noun to adjective and flips the polarity. Another approach to extract the polarity of this kind of words is to add the whole word (e.g. sevgisiz) to polarity lexicons. We followed the second method by generating polar words and providing them for Turkish sentiment analysis systems.

we use only two sets of suffixes–siz, sız, suz, süz [without], and li, lı, lu, lü [with]– because they generate new polar words with a negligible error rate. For example, if an adjective ended by [li, lı, lu, lü] is positive, replacing the suffix with [siz, sız, suz, süz] will generate a negative adjective. Note that there are many other suffixes for transforming nouns and adjectives to each other, but almost all of them generate erroneous (or irrelevant) words when adding/removing them to/from words. As

Table 4.7: **Patterns used for extracting new polar adjectives by changing their suffixes.**

Adj+(li, lı, lu, lü)=> Adj+(siz, sız, suz, süz)	insaf**lı** => insaf**sız** (pos=>neg)
Adj+(li, lı, lu, lü) => Adj+(siz, sız, suz, süz)	korku**lu** => korku**suz** (neg=>pos)
Adj+(siz, sız, suz, süz) => Adj+(li, lı, lu, lü)	keder**siz** => keder**li** (neg=>pos)
Adj+(siz, sız, suz, süz) => Adj+(li, lı, lu, lü)	vicdan**sız** => vicdan**lı** (pos=>neg)

an erroneous example, the affix "sel" [related to] can transform a noun to an adjective, but in this work, if it was used to obtain new adjectives from nouns, the newly generated words should be manually checked for their validity in Turkish. Therefore, we used only the above mentioned suffixes which could achieve the highest accuracy in automatic generation of new polar adjectives by replacing a suffix with another.

By employing this method, we generated about 250 positive and 150 negative adjectives which were new to already generated polarity lexicon.

Furthermore, a small set of polar nouns is also generated by removing suffixes from adjectives. For example by removing the suffix "sız", the negative adjective "heyecansız" [without excitement] changes to positive noun "heyecan" [excitement]. A small polar noun set generated by this technique can be also useful for sentiment analysis systems.

4.4 Phrase Polarity Lexicon Generation

There is not enough attempt in generating phrase polarity lexicons; two work have been accomplished by Wilson et al. [2005] and [2009]. In 2005, the authors propose an approach to phrase-level sentiment analysis that first classifies an expression as subjective or objective and then estimate its polarity in the case of subjectivity. The authors estimate the contextual polarity of an expression by using a large number of subjectivity clues and the prior polarity of appeared words in the expression. This work mostly relies on statistical methods. The obtained accuracies in classifying expressions as objective/subjective and also positive/negative range from 61% to 75%.

The authors expanded their work in 2009. The focus of this work is understanding which features are more important in automatically distinguishing between prior and contextual polarity. Multi-perspective Question Answering (MPQA) is used as the opinion lexicon in this work.

For phrase lexicon generation, we modified the above explained methodology for adjective lexicon generation. The main difference is adding a pre-processing step for extracting phrases from Turkish sentences. This pre-processing step as well as the whole approach are explained in the following subsections.

4.4.1 Phrase Extraction

A phrase is defined as "a small group of words standing together as a conceptual unit, typically forming a component of a clause" in Oxford dictionary[5]. As another

[5] http://www.oxforddictionaries.com/

definition [6], a phrase is a sequence of two or more words arranged in a grammatical construction and acting as a unit in a sentence. Phrases can be divided into two main categories: noun and verb phrases. We did not cover Adjective and adverb phrases because they are not prevalent. According to Oxford dictionary a noun phrase is a word or group of words containing a noun and functioning in a sentence as subject, object, or prepositional object such as "'inanılmaz bir performans" (an unbelievable performance), while a verb phrase is a verb with another word or words indicating tense, mood, or person such as "'gözlerimizi boyadılar" (they deceived us). Both phrase types are addressed in this work. Unlike the adjective list which was directly downloaded from an online Turkish dictionary, we could not find an existing list of Turkish phrases; therefore we generated such a list by extracting collocations–trigrams and quadrigrams–using patterns in Table 4.8. The employed patterns provide a large and generally meaningful list of phrases. We did not include separated expressions because using them in sentiment analysis tasks as a polarity lexicon is much more difficult than the collocated ones. Note that the collocated expressions are not necessarily compositional. As defined by Manning and Schütze [50], an expression is compositional if its overall meaning can be estimated based on the meaning of its parts. As not all extracted phrases

Table 4.8: **Patterns used for extracting phrases from sentences.**

triples	quadruples
adv adj verb	adv adj adj noun
adv adj noun	adv adj noun noun
adv adv verb	adv adj noun verb
adj noun verb	adj adj noun verb
	adv adv adj verb

are correctly formed phrases, we trained a classifier to classify unseen phrases as correct and incorrect. For this purpose, we relied on three features listed below.

- N-gram language model: This method computes the co-occurrence probability of terms (words) with each other in a phrase. The goal is to distinguish correctly formed phrases from incorrectly formed ones. If the co-occurrence probability of included terms in a phrase is high, most probably they constitute a commonly used phrase. As mentioned in [50], N-gram language model can be computed by probabilities given in equation 4.7.

$$\log(P(t_i, t_j, t_k)) = \log(P(t_i)) + \log(P(t_j|t_i)) + \log(P(t_k|t_i t_j)) \tag{4.7}$$

[6]dictionary.reference.com

$P(t_i)$ is the probability of seeing the term t_i in a phrase, $P(t_j|t_i)$ and $P(t_k|t_it_j)$ are conditional probabilities of seeing t_j and t_k after seeing the given terms t_i and t_it_j in a phrase, and $P(t_i, t_j, t_k)$ is the probability of having correctly formed phrase with three terms: t_i, t_j, and t_k. A similar equation could be written for quadruples. For example, in the phrase "daha fazla olmalıydı" [it should be more (better) than this], extracted by the pattern [Adv Adv Verb], $\log(P(daha))$, $\log(P(daha|fazla))$, and $\log(P(olmalıydı|dahafazla))$ are computed.

- Hit number in a search engine: In this feature, each phrase is searched in Google search engine to capture its hit number. The higher the number of hits for a phrase, the higher the probability of correct formation.
- Document frequency: This feature simply counts the appearance of each phrase among 11,000 Turkish sentences (unlabelled) extracted from Turkish movie reviews.

After training the classifier by using the above mentioned features, we classified all phrases as correctly formed and incorrectly formed. By the help of this classification, incorrect phrases are removed from the list. This classification have been trained by 1000 phrases manually labelled as correct and incorrect which was kept separated from the test set. The input of this classification task (test set) is a set of 5213 phrases and the output is a set of 4950 common phrases. A correctly classified sample is "üstüne yok doğrusu" [Actually there is no higher level upon it] and an incorrectly formed phrase which was misclassified as correctly formed is "bir film günün en ..." [the most ... of a movie day]. Note that an incorrectly classified sample does not make sense in Turkish or very unlikely appears in a Turkish sentence.

4.4.2 Polarity Classification Features for Phrases

The list of features for phrase extraction and polarity classification is provided in Table 4.9. First set of features have been used for phrase extraction (explained in Section 4.4.1) and the rest of features have been used for polarity classification of phrases which are explained below.

- Appearing in Positive/Negative sentences: We counted the appearance of each phrase in 6,000 positive and negative sentences which has been extracted from Turkish movie reviews and Twitter and have been manually labelled as positive, negative, or neutral.

- Positive/negative word count: This feature captures the number of positive and negative terms appeared in a phrase. We used two Turkish polarity lexicons for this purpose: Polar Word list and SentiTurkNet [35]. In polar word list, words are already separated as postive and negative; In SentiTurkNet, similar to SentiWordNet, three polarity scores are assigned to each Turkish synset. We assumed a Turkish word as positive if the positivity score of its synset is higher than the negativity score. Similar to the WSD method for SentiWordNet, part of speech tags are used for WSD of Turkish terms in SentiTurkNet. This feature is assumed as baselinefor phrase lexicon generation as it simply counts the number of positive and negative terms in a phrase.

Table 4.9: **Features extracted for classifying phrases as positive, negative, or neutral.**

Phrase Extraction	N-gram language model Hit number in Google Document frequency
Polarity classification	Appearing in Pos/Neg sentences Pos/neg word count

4.4.3 Polarity Classification for Phrases

After each phrase is classified as "correct" or "incorrect", we attempted to classify the correct phrases as positive, negative, or neutral. For this purpose, two classifications (listed below) are carried out by using features listed in Table 4.9. The classifier, evaluation method, and classification tool are the same as those in adjective classification task: logistic regression, 5-fold cross validation, and WEKA. The proposed methodology for building phrase polarity lexicon is illustrated as a flowchart in Figure 4.2.

- *Classifying phrases as subjective and objective (neutral)*: In this classification, the output of previous step (phrase extraction) are classified as objective and subjective; in other words, objective phrases are removed from

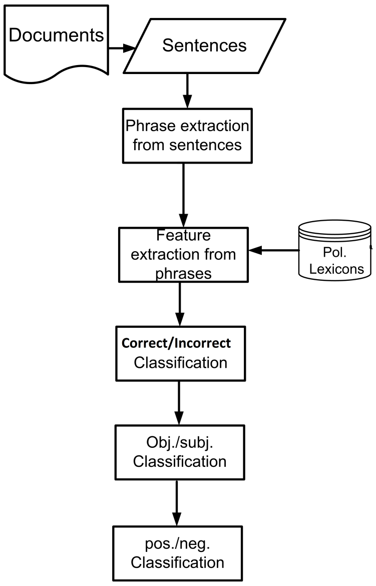

FIGURE 4.2: The proposed methodology for phrase lexicon generation as a flowchart

the list. The input of this classification is a set of 4950 phrases and the output is a set of 2092 subjective phrases. A correctly classified sample is "nasıl böyle saçma" [how silly like this] and an objective phrase which is incorrectly classified as subjective is "tabii romantik komedi" [Naturally a romantic comedy]. The training set for this classification is a set of 800 correctly formed phrases which have been manually labelled as subjective and objctive, and the test set is 4950 phrases obtained from phrase extraction phase.

- *Classifying subjective phrases as positive and negative*: In this classification, the output of previous step (subjective phrases) are classified as positive and negative. The input of this classification is a set of 2092 phrases and the output is a set of 1591 positive and 501 negative phrases. The lower number of negative phrases may be caused by the lower number of negative reviews and sentences in movie reviews. The training set for this classification is a set of 500 correctly formed phrases which have been manually labelled as positive and negative, and the test set is 2092 phrase obtained from the previous (subjective/objective) classification task.

  A correctly classified positive phrase is "tek işe yarar ..." [the only useful ...]; a correctly classified negative phrase is "kesinlikle çok gereksiz bir..." [Absolutely a very unnecessary ...]. Finally a positive phrase that has been misclassified as negative is “izlediğim en iyi gerilim” [The best intensity movie that I have ever watched].

Chapter 5

GRANULARITY LEVELS AND NLP ISSUES IN SENTIMENT ANALYSIS

5.1 Granularity Levels in Sentiment Analysis

The most common level in sentiment analysis is the document level, in which we predict a polarity label (positive, negative, or neutral) for the whole document. This approach may lead to some information loss in documents that have mixed sentiment; for example in movie reviews, if an aspect (e.g., action) is positive but another aspect (e.g., director) is negative, the sentiment analyser may classify this document as neutral while in fact it has mixed sentiment. Finer grain analysis is required to address this issue:

- *Word level:* Assigning a sentiment polarity to a word is not very easy, as a word may have different polarities in different domains or even in the same domain. For example the word "long" has a positive polarity for the aspect *battery life* but a negative polarity for the aspect *zooming time* in the camera domain.
- *Phrase level:* A phrase is an ordered (not necessarily consecutive) list of n terms within a sentence and a sentence is composed of one or more phrases, possibly with different sentiments. For example the sentence below has two verb phrases with two different sentiments (one shown in italic):

Ben beğendim, *ama herkes beğenmedi.*

(I liked it, *but not everyone did.*)

- *Aspect level:* Aspects are different perspectives relating to the review item, e.g., "room" in hotel reviews or "plot" in movie reviews. Each sentence in a domain may include several aspects and the polarity of each aspect may be different from the overall polarity of the sentence. For example the sentence below has two phrases about two separate aspects, one with positive and the other with negative sentiment (one shown in italic):

  oyunculuk iyi, *ama efektleri sevmedim*

  (the acting is good, *but I did not like the special effects*)

- *Sentence level:* Assigning an overall sentiment to a sentence is sometimes required. If the sentence has a mixed polarity (both positive and negative due to multiple aspects or phrases), one can assign an overall polarity based on relative sentiment strengths of the components.

- *Document level:* This is the coarsest level and attempts to estimate the overall polarity of a document. Often document polarity is aggregated from the estimated polarity of the constituent words or sentences. Previous work [14] has shown that initial and last sentences may have higher influence on document polarity, compared to sentences in the middle.

5.2 Natural Language Processing Issues in Sentiment Analysis

An effective sentiment analysis system must handle various linguistic markers such as negations, intensifications, and conditional constructions, in order to make more precise sentiment classifications. Most of these marker are language-specific and may need to be extracted using various language-specific tools (e.g., morphological analysers and parsers), while some others such as emoticons could be considered language-independent.

Below we group the issues that we rely on for Turkish sentiment analysis, into two subsets: language-specific and language-independent issues. Here, we present only the challenges, while proposed solutions are presented in Section 6.1.3 in Chapter 6.

5.2.1 Linguistics Issues

- *Negation:* Negation markers can switch the polarity of a predication or main verb in their scope. The following sentence is a simple negation form by using the predication negation marker "değil" (is/am/are not):

  ...20 defa izlemişimdir, *p*işman değilim.
  (...probably watched it 20 times, *I* am not regretful.)

  where we have a negative to positive change in the sentiment as "pişman" (*regretful*) is negated by "değilim" (*I am not*).

  The second example provides a more complicated negation form by two negated verbs where the underlined morphemes in words mark negation:

  *s*ev<u>me</u>dim diyen ¸cık<u>ma</u>dı
  *(*no one came out saying they *d*id not like it)

  where polarity first switches to negative with " sev<u>me</u>dim" (*I did not like*) and back to positive within the scope of "çık<u>ma</u>dı" (*no one came out*).

- *Intensification:* Intensifiers modulate the polarity of a term stronger or weaker. For example, the adjective "iyi" (good) is strengthened in "*çok* iyi" (*very good*) or weakened in "*biraz* iyi" (*so so good*).

- *Conditional sentences:* These sentences may change the *apparent* polarity of a sentence. For example the sentence below indicates a less positive sentiment than what is indicated by the existence of a score of 10.

  ¸Cok uzun olmasaydı, 10 verirdim.
  (If it was not too long, I would have given it a 10.)

- *Rhetorical questions:* The polarity of these sentences usually differ from what appears on the surface–that is the expression is formally a question sentence but is not used to elicit an answer; it rather is used to convey a variety of sentiments. For example, in Table 5.1, the overall sentiment is made positive with the addition of the question suffix (**mi**), while "sevmeyebilir" (cannot like?) has negative polarity.

Table 5.1: An example rhetorical question

İnsan	bu	filmi	sevmeyebilir	mi?
anyone	this	movie	cannot like	?
Can	anyone	not like	this movie?	

- *Idiomatic uses:* An idiom is a combination of words whose meaning is a compositional combination of the meanings of its constituent words. The challenging issue in idioms is that the polarity of an idiom cannot always be extracted automatically by using the polarity of terms included within the idiom. For example a commonly used idiomatic compound verb in Turkish is "göz boyamak" (*to deceive* – literally to paint the eyes) which has a negative sentiment while its constitutents "göz" (*eye*) and "boyamak" (*to paint*) are neutral terms when considered separately.

5.2.2 Other Issues

- *Emoticons:* Emoticons can help estimate the polarity of a sentence. Normally positive emoticons appear in positive sentences and negative ones in negative sentences. For example the *happy emoticon* ":)" may appear at the end of a positive sentence and the *sad emoticon* ":(" usually appears at the end of a negative sentence.

- *Sarcastic phrases:* Sarcasm detection may be the most challenging issue in language processing tasks. This task has obtained very low accuracy even in English (57%) [2]. A sarcastic statement such as "harika bir film olmuş!" (it was a great movie!) can only be detected by the disagreement with it and the whole of the (negative) review and slightly hinted by the exclamation mark.

- *Domain-specific indicative keywords*: The polarity of sentiment keywords can change across domains. Furthermore, each domain has some keywords that are good clues for estimating the polarity of a sentence/review that includes those keywords. For example the phrase "kaçırmayın" (do not miss it) at the end of a movie review is a commonly used positive phrase in the movie domain.

- *Conjunctions:* Conjunctions can help estimate the polarity of the two terms around the conjunct, with the help of the other. For example two adjectives conjoined by “ama” (but) are supposed to have opposite polarities, while they often have the same polarity when they are conjoined by “ve” (and). This observation was made and used to estimate word-level polarities in previous work [32].

- *Background knowledge:* Sentiment analysis systems require background knowledge for classifying special kinds of sentences such as: *“of those rare films that makes me feel that I am present in the film”.* In this sentence, the key issue is that the feeling of being present in the film is a positive emotion, which is the background knowledge necessary to understand the sentiment. It is however extremely hard with the current state of the art in natural language processing to extract such information.

Chapter 6

TECHNIQUES FOR SENTIMENT ANALYSIS IN TURKISH

In this chapter, we explain the proposed framework for sentiment analysis which has been adapted for Turkish. As already mentioned in previous chapters, English has the richest set of sentiment analysis resources such as the SentiWordNet [4], and SenticNet [6]; However, social media is proliferating in many other places where many other languages are used and sentiment analysis for data in those languages has developed significant demand. We focused on Turkish owing to significant penetration of traditional social media as a percentage of the population and proliferation of homegrown social media of local interest. The few earlier work on Turkish sentiment analysis however, have mostly focused on a binary (positive and negative) classification at the document level. The sole focus on binary classification does not appear sufficient as documents/sentences can be neutral and ignoring this class leaves out a large portion of reviews.

In earlier work, we have built polarity resources for Turkish such as SentiTurkNet [35] (Please refer to Chapter 4), and polar word list [35].

In this work, we propose a system for sentiment analysis in Turkish and apply it Turkish movie reviews.[1] Our method works at aspect, sentence, and document levels. Our contributions can be summarized as follows.

[1]These reviews are collected from `www.beyazperde.com`.

- We provide a comprehensive overview of issues for sentiment analysis in Turkish,
- We propose and evaluate a sentiment analysis system for Turkish covering linguistic issues and different levels of analysis granularity. This system exploits polarity lexicons such as SentiTurkNet and additional natural language processing techniques such as dependency parsing that have not yet been employed for Turkish sentiment analysis in the literature.

6.1 Proposed Methodology for Sentiment Analysis in Turkish

In this section, we first present an overview of our system and then elaborate on each component of the system, explained in the following subsections. The system consists of several components as illustrated in Figure 4.2. The input is a document (a movie review) which is segmented into sentences and then each sentence is fed to a parser [51] that provides the dependency tree structure of the sentence. This structure is used in aspect-level polarity classification (see Section 6.1.2.3).

We assign polarity scores to word unigrams and bigrams by using the polarity lexicons: SentiTurkNet, polar word list, adjective polarity lexicon (all explained in Chapter 4) and translation of the SenticNet (see Section 6.1.1).

After assigning polarity values to terms in a sentence, and covering linguistic and other related issues, we do a sentence level polarity classification (see Sections 6.1.5) by using 16 features listed in Table 6.5.

A document level sentiment classification (see Sections 6.1.6) is then accomplished by using features listed in Table 6.6 with four additional features (compared to Table 6.5) indicating the estimated polarities of the first and last sentences in the document.

6.1.1 NLP Tools and Polarity Resources

We rely on a parser and three polarity lexicons in this work.

Table 6.1: **Parse tree generated by using the ITU parser for the sentence "Bence hoş vakit geçirmek için seyredilebilir" (it can be viewed for an enjoyable time).**

Pos.	Word	Root	Major POS	Minor POS	Morph. Features	Dep. Head
1	Bence	ben	pron	pers	A1sg.pnon.equ	0
2	hoş	hoş	adj	adj	–	
3	vakit	vakit	noun	noun	A3sg.pnon.nom	4
4	–	geçir	verb	verb	Pos	5
5	geçirmek	–	noun	Inf1	A3sg.pnon.nom	6
6	için	için	postp	pcnom	–	0
7	seyredilebilir	seyredil	verb	Able	Pos.aor.a3sg	0

- *ITU Turkish Parser [51]:* This parser receives a Turkish sentence as input, produces a dependency tree with morphological analyses for every token in the sentence. The output of this parser for the sentence "bence hoş vakit geçirmek için seyredilebilir." (It can be viewed for an enjoyable time) is illustrated in Table 1. This parser is not perfect and may parse some sentences with errors. These potential errors will affect our methodology but because only a few features such as conditionality, and interrogativity of sentences are based on this parser, the erroneous parsed sentences will slightly affect our methodology. Another alternative could be using a morphology analyser such as Zemberek [25], but those kind of analysers do a word-level (rather than sentence-level) analysis.

- *Polar word list:* We have semi-automatically generated a list of polar Turkish terms including 1000 positive and 1000 negative terms using the method proposed by Hu and Liu [2004]. This method benefits from synonymy and antonymy relations between terms to generate a large polar word set starting from a small seed set. We also added the adjective polarity lexicon (explained in Chapter 4) to this polarity resource. The generated set is named *Polar Word Set* (PWS).

- *SentiTurkNet (STN):* We have developed the first Turkish polarity resource, STN, based on the Turkish WordNet [43], where three polarity scores are assigned to each Turkish synset (set of synonyms) indicating its positivity, negativity, and neutrality levels. This resource consists of 15,000 synsets and 1.47 terms per synset in average. For more information, please refer to Chapter 4.

- *SenticNet (SN):* This resource assigns different numerical values to each term as its *pleasantness, attention, sensitivity, aptitude,* and also the *overall-polarity*. Each of these features has a value between -1 and $+1$ as the most negative and the most positive polarities respectively. We translated this resource to Turkish by a bilingual dictionary named *seslisozluk* [49] and used only the *overall polarity* of each term (or phrase) as its sentiment strength. This lexicon contains about 14,000 entries (words and phrases).

6.1.2 Sentiment Analysis Levels in Turkish

Our system is designed to address different levels of sentiment analysis: words, phrases, aspects, sentences and overall document, as explained below.

6.1.2.1 Word level

We extract the polarity of a given word using the polarity lexicons described in Section 6.1.1 (PWS, SN and STN), to be used as features in sentence and document-level analysis. In PWS, we have a label (positive or negative); in SenticNet we have a polarity value ranging from -1 to 1; and in SentiTurkNet we have three polarity scores for each word. In the last case, we extract the positive and negative polarities and we use those separately in subsequent steps, rather than deciding the dominant polarity.

As each term may have different connotations in STN and only its contextual meaning is desired, word-sense disambiguation (*WSD*) is required. However there are no WSD systems for Turkish so we narrow senses by relying on the morphological features–mostly the part of speech (POS). WSD is an ongoing problem in Turkish and English, and exploting the part-of-speech tags for this purpose improves the efficiency of polarity extraction, when compared to randomly choosing the word-sense in a context. In SN, we use only the *overall-polarity* score of each word or phrase (sequence of words). In PWS, only the polarity label (positive or negative) is available, which indicates the overall polarity of words.

Word-level polarities found here are then combined taking into additional through linguistic markers and modified by the methods proposed in Section 6.1.3. The modified polarity scores/labels are then used in aspect, sentence, and document level classifications.

6.1.2.2 Phrase level

We use the dependency parse structures produced by the parser described above to identify disambiguated morphological analyses for all the words in the sentence and any relational structures (e.g., *subject-verb-object*) encoded by a dependency parse. Note that we generate structures with any number of terms, for example if term t_i is related to (dependent on) term t_j, t_j is related to t_k, and t_k is related to t_l, the phrase "$t_i t_j t_k t_l$" is extracted from the sentence. The relations let us focus on the main predications or relevant modifications or conjunctions in the sentence, ignoring words that may not be relevant for sentiment analysis. Looking up the words in the resource we have built, provide initial estimates of sentiment.

In this step, we do not explicitly do sentiment analysis in the phrase level; instead, we use the output–extracted phrases by dependency parse tree–in the aspect level sentiment analysis. For more detailed phrase-level sentiment analysis please refer to Chapter 4 (phrase polarity lexicon generation).

6.1.2.3 Aspect level

We compiled a list of aspects (A) in movie domain and proposed a novel method for estimating the polarity of each aspect. After identifying an aspect a_j in a sentence S, we identify those relations to encode basic predications. An example sentence is given below.

> Oyunculuk iyi, *ama efektleri pek sevmedim.*
> (The acting is good, *but I did not like the effects that much.*)

In this sentence, the two phrases "oyunculuk iyi" (*the acting is good*) and "efektleri sevmedim" (*I did not like the special effects*) are extracted from the dependency tree ignoring other words that do not necessarily have much effect on the sentiment. We then compute the average polarity (positivity and negativity) of all such relations involving the aspect a_j in sentence S by means of two terms $P(a_j)$ and $N(a_j)$ that indicate the average positivity and negativity scores of aspect a_j, using Equations (6.1) and (6.2).

$$P(a_j) = \sum_{\forall n_k \in NG, s.t. \ a_j \in n_k} \frac{\sum_{t_i \in n_k} pos(t_i)}{|n_k|} \tag{6.1}$$

$$N(a_j) = \sum_{\forall n_k \in NG, s.t.\ a_j \in n_k} \frac{\sum_{t_i \in n_k} neg(t_i)}{|n_k|} \tag{6.2}$$

where NG is the set of all relational structures generated by the dependency parse tree; and n_k is a relational structure in the sentence; $|n_k|$ is the number of tokens in n_k; and $pos(t_i)$ and $neg(t_i)$ are positivity and negativity scores of term t_i, as extracted from SentiTurkNet.

These relational structures consist of two, three, or more words that are structurally related together in the dependency parse tree. In these equations, if $P(a_j) > N(a_j)$, a_j is classified as positive, if $P(a_j) < N(a_j)$, a_j is classified as negative, or neutral otherwise. The list of aspects is provided in Table 6.2.

TABLE 6.2: **The list of chosen aspects from Movie domain for our system.**

aksiyon (action), oyuncu/aktor (actor), müzik (music), sahne (scene), efekt (effect), senaryo (scenario), oskar (oscar), yönetmen (director), animasyon (animation)

6.1.2.4 Sentence level

We start sentence level sentiment analysis by automatically segmenting each document to its sentences by using punctuation, capitalization, and emoticons. Then, we extract 16 features given in Table 6.5 from each sentence to be used in classification task. The classifier is trained with 2,700 labelled (as pos, neg, or obj) sentences in the Turkish movie reviews and evaluated using 5-fold cross validation. Note that in order to simplify the sentiment analysis task, as done in the literature, we also assumed that each sentence has a single sentiment towards a target. This assumption is not real and must be ignored in phrase or aspect level sentiment analysis, as we did.

6.1.2.5 Document level

We address the document level sentiment analysis similar to the sentence level analysis, using 20 features given in Table 6.6. The classifier is trained by 1000 feature vectors which have been extracted from 1000 labelled documents (as pos, neg, or obj) in the Turkish movie reviews. We also benefit from additional four features for this level to highlight the effect of the first and last sentences in the document. The evaluation method for this classifier is again 5-fold cross validation.

6.1.3 Issues in Turkish Sentiment Analysis

In this section, we propose our solutions for most of the linguistic issues discussed in Chapter 5 and leave some of them as future work. We also address some additional relevant issues.

The proposed methods in this section are applied on words and sentences to change their polarity if applicable. As mentioned earlier, the initial polarity scores and labels for words have been obtained from three polarity lexicons explained in Section 6.1.1, which can be changed through the studied (linguistic and other) issues.

6.1.3.1 Linguistic issues

- *Negation:* We covered different kinds of negation in Turkish and were able to increase the classification accuracy by about two percentage points.
  - The predication negation marker *"değil"* (is/am/are not) switches the sentiment of polar words in the sentence, preceding the verb "değil". For example in the sentence *"ama kötü bir film de değil"* [but it is not a bad movie at all) the marker *"değil"* switches the negative polarity of "kötü" (bad) to positive (not negative).
  - Morphemes "ma" and "me" in verbs negate the polarity of a verb. For example "sevdim" (I liked) has positive sentiment but sentiment changes to negative when the morphological negation is introduced with the morpheme "me" in "sevmedim" (I did not like). For this we rely on the disambiguated morphological representation of the verbs provided by the dependency parser.
  - Morphemes "lu" and "suz" derive adjectives from noun with the semantics of with or without respectively. For example the noun "kusur" (fault) is a negative term and morphemes "lu" and "suz" generate adjectives "kusurlu" (faulty) and "kusursuz" (flawless) which have negative and positive sentiments respectively.

- *Intensification:* We compiled a set of intensifiers in Turkish listed in Table 6.3. For strengthening intensifiers we double the sentiment value and for weakening intensifiers we halve it. This has contributed about a percentage points to our classification accuracy.

- *Conditional sentences:* We cover this only by adding a boolean feature to the classification features (Tables 6.5 and 6.6) indicating the conditionality of a sentence. This issue needs further investigation that we have left for future work.

- *Rhetorical questions:* We cover this by adding a boolean feature to the classification task, which indicates if a sentence is interrogative. Capturing only the rhetorical questions (not all interrogative sentences) needs further investigation that we have left as future work.

6.1.3.2 Other issues

In this work, we cover only three issues:

- *Emoticons:* We compiled a list of 50 positive and 50 negative emoticons and marked their presence with appropriate features.

- *Domain-specific indicative keywords:* We gathered a list of 20 keywords and key phrases that indicate positive sentiment in Turkish movie reviews. A subset of these keywords and keyphrases is listed in Table 6.4. Again we mark their presence with appropriate features.

- *Conjunctions:* we apply the proposed idea by Hatzivassiloglou and McKeown [1997] to Turkish, by using the conjunctions "ama/fakat" (but) and "ve" (and). Two adjectives conjoined by "and" are supposed to have same polarity while they will most probably have the opposite polarity when conjoined by "but". Two examples from Turkish movie reviews are given below:

  Film güzel ama çok uzun.
  (the film is good *but* too long.)

  Film güzel ve heyecanlı.
  (the film is good *and* exciting.)

In the former example, our approach estimates the polarity of "çok uzun" (very long) as negative because it already knows that "güzel" (beautiful/-good) is positive.

Conjoined adjectives (although rare) help to increase the classification accuracy only about 0.5 percentage points.

Table 6.3: **A subset of strengthening and weakening intensifiers.**

Strengthening (very/really):	*baya(ğı)*, *gayet*, *çok gerçekten*, *iyice*, *cidden*
Weakening (a little/almost):	*biraz*, *azcık*, *yaklaşık*

Table 6.4: **A subset of domain-specific indicative terms/phrases in Turkish movie reviews.**

izleyin (watch it), iyi seyirler (happy viewing),*izlemeli*, *izlemek gerek* (should be watched), *kaçırmayın* (do not miss it), *izlenebilir* (could be watched)

6.1.4 Features for Sentence and Document Classification

The 16 and 20 features used in sentiment classification of sentences and documents are listed in Tables 6.5 and 6.6. Features $f_1 - f_{16}$ in two tables are similar but $f_{17} - f_{20}$ are additional features used only in the document level. The term "review" used for feature explanations in the following paragraphs refers to a sentence in case of sentence level sentiment analysis, and refers to a document in case of document level sentiment analysis.

- $f_1 - f_4$: The first four features capture the average polarity of terms in a review, computed using two separate resources that assign numerical polarity scores to each term. In SenticNet, we label a term as positive if its polarity score is non-negative, otherwise it is negative. In SentiTurkNet, three polarity scores are assigned to each Turkish synset but we use only positivity and negativity values (neutrality score depends on these two scores).

- $f_5 - f_6$: These features count the number of positive and negative polar terms in each review, based on the PolarWordSet.

- $f_7 - f_8$: These features capture the appearance of positive and negative emoticons in the review.

TABLE 6.5: **Features used in sentiment analysis of a sentence, S. SN, PWS, and STN respectively stand for SenticNet, PolarWordSet, and SentiTurkNet.**

f_1: *average positive score of words in S using STN*
f_2: *average negative score of words in S using STN*
f_3: *average score of positive words in S using SN*
f_4: *average score of negative words in S using SN*
f_5: *number of positive words in S using PWS*
f_6: *number of negative words in S using PWS*
f_7: *occurrence of positive emoticons in S*
f_8: *occurrence of negative emoticons in S*
f_9: *number of adjectives and adverbs in S*
f_{10}: *number of (first letter) capitalized words in S*
f_{11}: *number of domain-specific indicative words in S*
f_{12}: *length of sentence (number of tokens in S)*
f_{13}: *is S a conditional sentence?*
f_{14}: *is S an interrogative sentence?*
f_{15}: *is S a negated sentence?*
f_{16}: *is S an exclamative sentence?*

- $f_9 - f_{12}$: These features model three assumptions: (1) the higher the number of adjectives and adverbs in a review, the higher the chances of its subjectivity; (2) the higher the number of initial capital words in a review, the greater the chances of neutrality for the review (capitalized terms are proper nouns which are generally neutral); and (3) the higher the number of domain-specific indicative terms in a review, the greater the chances of positivity for the review. *length of review* simply counts the number of tokens in the review

- $f_{13} - f_{16}$: These features capture the interrogative, conditional, negated, or exclamative form of a sentence. These features can be extracted from the output of the parser.

- $f_{17} - f_{20}$: These polarities of the first and last sentences in the document are used as features for document level sentiment analysis, following the sentence level analysis. Generally the first and last sentences are more subjective than the middle sentences because many people write their ideas more clearly in the first and last sentences.

  We analysed the relationship between the document polarity and the polarity of its first and last sentences. Table 6.7 shows the conditional probabilities

TABLE 6.6: **Features used in sentiment analysis of a document, D. SN, PWS, and STN respectively stand for SenticNet, PolarWordSet, and SentiTurkNet.**

f_1: *average positive score of words in D using STN*
f_2: *average negative score of words in D using STN*
f_3: *average score of positive words in D using SN*
f_4: *average score of negative words in D using SN*
f_5: *number of positive words in D using PWS*
f_6: *number of negative words in D using PWS*
f_7: *occurrence of pos. emoticons in D*
f_8: *occurrence of neg. emoticons in D*
f_9: *number of adjectives and adverbs in D*
f_{10}: *number of (first letter) capitalized words in D*
f_{11}: *number of domain-specific indicative words in D*
f_{12}: *length of document (number of tokens in D)*
f_{13}: *Does D contain a conditional sentence?*
f_{14}: *Does D contain an interrogative sentence?*
f_{15}: *Does D contain a negated sentence?*
f_{16}: *Does D contain an exclamative sentence?*
f_{17}: *avg. positive score of words in first sentence of D*
f_{18}: *avg. negative score of words in first sentence of D*
f_{19}: *avg. positive score of words in last sentence of D*
f_{20}: *avg. negative score of words in last sentence of D*

of the document polarity given the sentence polarity. For instance 76% of documents with positive sentiment have a positive first sentence.

6.1.5 Sentence Level Classification

For sentence classification, 16 features in Table 6.5 are used with a Logistic Regression (LR) classifier [44]. The evaluation is done using 5-fold cross validation over training data of 2700 sentences. Both binary and ternary classifiers are trained separately at this level.

6.1.6 Document Level Classification

Document sentiment analysis follows sentence level analysis and uses 20 features in Table 6.6. In order to show the importance of first and last sentences in estimating the polarity of the whole review, we computed two types of conditional probabilities: (1) Conditional probability of the document given the actual polarity of the first sentence, and (2) Conditional probability of the document given the actual polarity of the last sentence. These probability values are given in Table 6.7. The

Table 6.7: **Conditional probability of the document polarity given the polarity of the first or last sentence.**

Document	first sentence		
	positive	**negative**	**neutral**
positive	0.76	0.01	0.23
negative	0.01	0.79	0.20
neutral	0.13	0.04	0.83
Document	last sentence		
	positive	**negative**	**neutral**
positive	0.76	0.05	0.19
negative	0.03	0.56	0.41
neutral	0.13	0.10	0.77

classifier and evaluation methods are the same as in sentence level analysis, using logistic regression classifiers and 5-fold cross-validation for evaluation.

Chapter 7

EXPERIMENTAL EVALUATION

In this chapter we evaluate the quality of three polarity lexicons: SentiTurkNet, adjective polarity lexicons, phrase polarity lexicon and also the proposed sentiment analysis system.

7.1 Evaluation of SentiTurkNet

In this section, we explain the evaluated dataset and methodology used for building SentiTurkNet.

7.1.1 Dataset

In the evaluations, we either used a small test set, sequestered for this purpose, or all of the data (all 14,795 synsets) using cross-validation.

The test set is a small subset of the synsets (3%) that has been kept sequestered for testing purposes. For this subset, called the gold standard set, we manually assigned a quantized polarity strength value to each synset (in one of eight intervals between 0 and 1). The reason for using this categorization was so that we could compare our resource with SentiWordNet and because assigning a value in a finer resolution would have been difficult.

7.1.2 Methodology

We evaluated the proposed approach by:

- *Test 1:* Mean absolute error between manually assigned ground-truth polarities on a small test set and the polarities estimated by the proposed method;
- *Test 2:* Misclassification error of the proposed method as compared with class labels assigned by manual labelling, using five-fold cross-validation on all data;
- *Test 3:* Misclassification error of the mapping from SentiWordNet with class labels assigned by manual labelling on all data;
- *Test 4:* Sentiment analysis improvements when using *SentiTurkNet* instead of the mapped *SentiWordNet* to Turkish for classifying Turkish movie reviews.

The mapping from SentiWordNet is used as baseline and is done in this way: As Turkish WordNet has been mapped (one to one) to English WordNet, the polarity scores of an English synset are used as polarity scores of its equivalent synset in Turkish WordNet.

7.1.3 Results

The above mentioned four tests are explained below.

7.1.3.1 Test 1

In the first evaluation, we used the test set and compared the Mean Absolute Error *(MAE)* between the manually assigned ground-truth polarities on these synsets and the ones obtained with proposed methodology. The MAE values presented in Table 7.1 are computed using Equation 7.1.

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n}|f_i - y_i| \tag{7.1}$$

where f_i and y_i are estimated and ground-truth scores of the ith synset, computed separately for positive and negative cases and n is the number of evaluated synsets. Note that the ground-truth has polarity levels 0,1,...,7 while SentiTurkNet has continuous polarity values between 0 and 1, and polarity scores in SentiWordNet are multiples of 0.125 between 0 and 1. In order to be able to compare SentiTurkNet with SentiWordNet we discretized the SentiTurkNet continuous polarity values into eight equal ranges (with labels 0,1,...,7). We also mapped the SentiWordNet values which are multiples of 0.125 to eight levels where SentiWordNet score of 0 corresponds to 0, 0.125 corresponds to 1, and 0.250 corresponds to 2 etc.

TABLE 7.1: **Mean Absolute Error on Test Data**

Classifier	pos	neg
SentiWordNet mapped to Turkish	3.73	3.01
SentiTurkNet with SMO	2.95	2.21
SentiTurkNet with LR	2.81	2.25
SentiTurkNet with NN	2.99	2.14
SentiTurkNet *with classifier combination*	**2.45**	**1.95**

As can be seen in this table, the mean absolute error computed over all the synsets by the final system are 2.45 and 1.95 separately for positive and negative synsets. The error rate with the classifier combination is 1 point or more lower than the baseline (mapping *SentiWordNet* to Turkish). These results support the assumption that translating (mapping) *SentiWordNet* to another language is not very accurate.

7.1.3.2 Test 2

In the second test, we evaluated the classification of synsets into three polarity classes, in the final polarity lexicon. Note that even though these labels were manually assigned initially, here we are testing the outcome of the classifier. If the manually assigned label differs from the label of maximum polarity score out of three scores *(pos, obj, neg)* in *SWN*, this was counted as an error. We used 5-fold cross-validation where the mapping between features and three polarity classes is learned using 80% of the data (training set) and the system is tested with the remaining 20% of the data, for an unbiased testing. This process is repeated five times with different 80-20% splits of the data and the results are averaged, as displayed in Table 7.2.

As seen in this table, the best accuracy (91.11%) is achieved using all features and classifier combination of three classifiers. A discussion on this table is provided in Section 7.5.4.

TABLE 7.2: **Classification Accuracy by the Individual Classifiers using 5-fold Cross Validation on All Data(%)**

Feature Subset	**Accuracy (%)**			
	(SMO)	**(NN)**	**(Logistic)**	**(Classifier Combination)**
f_1-f_9	79.03	79.71	79.42	86.72
f_{10}-f_{19}	79.02	78.74	78.97	85.26
f_{20}-f_{21}	79.03	79.16	79.22	86.11
f_{22}-f_{23}	81.63	81.99	81.93	87.32
f_1-f_{19}:	79.05	79.79	79.56	85.07
f_1-f_{21}:	79.05	79.85	80.14	87.99
f_1-f_{23} :	81.90	82.44	82.01	88.82
All features:	**82.89**	**83.32**	**83.13**	**91.11**

7.1.3.3 Test 3

As we have done in Test 2, this time we evaluated the polarity class assignments obtained from the mapped *SWN* (to Turkish WordNet), as a baseline comparison. In this case, we obtained the error rates of (32%, 10%, 22.5%) for positive, objective, and negative synsets. Error rate of objective synsets is low because most of synsets are objective and SentiWordNet assigns an objective label to them.

7.1.3.4 Test 4

The last evaluation studies sentiment analysis improvements when using *STN* instead of the mapped *SWN* for classifying Turkish movie reviews. More specifically, we use polarity scores obtained from *STN* or from the mapped *SWN*, to classify 300 reviews from Turkish movie dataset [1].

The method simply tokenizes the reviews and extracts the average polarity of terms in each review, to feed to a simple sentiment analysis classifier (by Logistic Regression [44]) we had developed previously. The accuracy of ternary classification (positive, negative, objective) by Logistic Regression and 5-fold cross-validation

[1]This dataset is collected from `http://www.beyazperde.com`

method using *STN* is 66.7% while it is 61.3% by using the mapped *SWN* to Turkish.

The low accuracy may be caused by the lack of language features such a negations, conjunctions, and intensifiers; our goal was to show the difference between polarity scores in two polarity resources by using them in a sentiment classification task. An example review that was correctly classified as positive using *STN* but incorrectly classified as negative using *SWN* is *"Sadece müziği için bile izlenir"* [It can be watched just for of its soundtrack].

We did not do Word-Sense Disambiguation (WSD) within the sentiment analysis system, as WSD is an ongoing problem in Turkish and is out of the scope of this work. Instead, for a given term with a given POS tag, we simply used the average polarity of all of its synsets with a matching POS tag. No *NLP* technique except extracting the root of words and their POS tag is used for this purpose.

The misclassified reviews by our system generally are those that include words which are absent in *STN* or those that are subjective but need background knowledge to distinguish this subjectivity such as *"izlerken bana ordaymış hissi veren nadir filmlerden".* ["of those rare movies that gives the feeling of being there (in movie) while watching"].

7.1.4 Discussion on SentiTurkNet

Results presented in Section 7.1.3 indicate that the proposed methodology is reasonably successful in predicting the label of synsets. Table 7.2 lets us draw the following conclusions:

- The most efficient feature group in isolation is $f_{22} - f_{23}$. The labels of hypernym and near-antonym synsets are good indicators for the polarity of a synset. Also this feature set is the most efficient adjustment to other feature groups.
- Classification accuracies of different classifiers are quite close and the highest accuracy is obtained by classifier combination. The higher accuracy of combined classifier indicates that polarity scores (confidence values) achieved by classifier combination are better than those found by the base classifiers.

- The errors are mostly caused by features related to glosses. It is common for a positive (or negative) synset to be explained by a non-positive (or non-negative) sentence. In most of the synsets, this deficiency is compensated by other features. An example for this statement is given in Table 7.3.

TABLE 7.3: **A negative synset misclassifed as neutral (objective).**

fields	content
synonym	iştahsız
gloss	Yemek yeme isteği olmayan, boğazsız [no desire to eat]
actual label	Neg.
estimated label	Obj.

The distribution (in percent) of positive, objective, and negative synsets in each part of speech is illustrated in Figure 7.1. As can be seen, the majority of synsets are objective in all parts of speech.Also among four parts of speech, nouns constitute the majority. Note that because of the low percentage of adverbs (less than 1%), they do not appear in this chart.

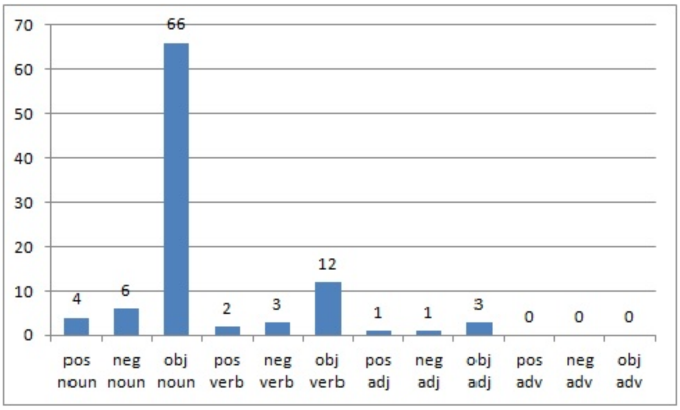

FIGURE 7.1: Distribution (%) of pos/neg/obj parts of speech in SentiTurkNet.

Similar to the above distribution, the overwhelming majority of all words in SentiWordNet are marked as objective.

7.2 Evaluation of Adjective Polarity Lexicon

We evaluated the proposed methodology by classification accuracy, and confusion matrix for classifications. The effect of each feature in each classification is shown in Table 7.4. Note that below the line is the accuracy of linguistic techniques which have been accomplished after doing the classification by using the features listed above the line. Adjective classification could be accomplished by two binary classification (objective-subjective and then positive-negative) or one ternary classification (positive-negative-objective). We experimented both cases and showed their effectiveness. Confusion matrices for both binary and ternary classification are provided in Tables 7.5 and 7.6. "Direct translation to English" (translation feature) is supposed as a baseline and our proposed methodology outperforms the baseline by about four percentage points in classification accuracy.

TABLE 7.4: **Binary and ternary classification accuracy for adjectives by Logistic Classifier using 5-fold Cross Validation on all Data(%)**

	Feature name	**ternary**	**com/uncom**	**subj/obj**	**pos/neg**
Adjectives	Hit number	52.45	-	53.01	52.03
	Delta *tf-idf*	54.24	-	57.23	59.17
	Translations	68.12	-	67.13	72.01
	PMI	53.20	-	56.50	56.17
	All features	**71.77**	-	**71.61**	**73.35**
	Conjunctions	**91.70**	-	-	-
	Suffixes	**73.50**	-	-	-

TABLE 7.5: **Confusion matrix for binary classification of adjectives with all features.**

True	Estimated	
	positive	**negative**
positive	0.78	0.22
negative	0.30	0.70

TABLE 7.6: **Confusion matrix for ternary classification of adjectives with all features.**

True		Estimated	
	positive	**negative**	**neutral**
positive	0.75	0.10	0.15
negative	0.12	0.66	0.22
neutral	0.10	0.18	0.72

7.3 Evaluation of Phrase Polarity Lexicon

For phrases, one ternary and three binary classifications have been accomplished. The intuition behind this is that incorrectly formed phrases must be excluded from the extracted list, then the remaining list should be classified as positive, negative, or neutral. The classification accuracies for binary classification of phrases as correct and incorrect are listed in Table 7.7, and similar classification accuracies for binary and ternary classification of correct phrases are listed in Table 7.8. Moreover, confusion matrices for both binary (pos/neg) and ternary classification of correct phrases are provided in Tables 7.9 and 7.10.

TABLE 7.7: **Binary classification of phrases as correctly formed and incorrectly formed by Logistic Classifier using 5-fold Cross Validation on training data(%)**

Feature name	**correct/incorrect**
N-grams	76.4
Hit number	72.20
Doc. freq.	70.45
All features	**79.40**

TABLE 7.8: **The accuracy of Binary and ternary classification of phrases by Logistic Classifier using 5-fold Cross Validation on training data(%)**

Feature name	**ternary**	**subj/obj**	**pos/neg**
pos/neg sentences	73.42	70.01	88.04
pos/neg words	71.02	68.22	85.16
Both features	**74.43**	**72.90**	**91.31**

TABLE 7.9: **Confusion matrix for binary (pos/neg) classification of phrases with all features.**

True	Estimated	
	positive	**negative**
positive	0.93	0.07
negative	0.18	0.82

Table 7.10: **Confusion matrix for ternary (pos/neg/neut) classification of phrases with all features.**

True		Estimated	
	positive	**negative**	**neutral**
positive	0.79	0.05	0.16
negative	0.11	0.68	0.21
neutral	0.17	0.15	0.68

7.4 Discussion on Adjective and Phrase Lexicons

To the best of our knowledge, the current work on building adjective and phrase lexicons is unique due to its hybrid approach. Previous work have benefited from some methods which are used as classification features in this work. Although previous work have used different datasets (making the comparison difficult), We report their performance.

Turney [2002] who proposed the *PMI* concept, achieved accuracies ranging from 66% in Movie reviews [13] to 84% in automobile reviews in classifying the reviews as positive and negative. Hatzivassiloglou and McKeown [1997] proposed the idea of conjoined adjectives by "and" and "but" which resulted in accuracies ranging from 78% to 82% in classifying adjectives as positive and negative extracted from 21 million word 1987 Wall Street Journal corpus[52].

As mentioned earlier, We evaluated the proposed approach by classification accuracy, and confusion matrix. The following conclusions can be extracted according to the obtained results presented in Tables 7.2, 7.13, and 7.14 .

- The proposed approach for adjectives, outperforms the baseline approach– direct translation to English– by about four percentage points. This issue approves the idea of building polarity lexicons specifically for a non-English language is more efficient that translating English polarity lexicons to non-English languages. This assumption was also approved in previous work [35].
- The proposed approach for phrases, outperforms the baseline approach– counting the number of positive and negative terms in phrase– by 1 to 3 percentage points. This issue emphasises the effect of non-compositional

phrases in sentiment analysis, in which the polarity of whole phrase cannot be estimated based on the polarity of its parts.

- In most cases, classifying the adjectives and phrases as positive and negative has the highest accuracy, while the ternary classification, as expected, has the lowest accuracy.

- In correct/incorrect classification of phrases, the N-gram feature obtained the highest accuracy. This finding approves the assumption that the higher the probability of co-occurrence of a word-pair, the higher the probability of a common phrase formation by this pair.

- In both binary and ternary classifications, phrase classification accuracies and also confusion matrix values are a little higher than those in the adjective case.

- In both adjective and phrase lexicons, the highest per-class accuracies (confusion matrix values) belong to the positive class and lowest accuracies belong to the negative class. Generally positive expressions are more clearly expressed by people when compared to the negative expressions.

7.5 Evaluation of Proposed Sentiment Analysis System

We evaluated the proposed approach in terms of its accuracy of classifying sentences, documents and aspects, in both binary and ternary classification scenarios, using 5-fold cross-validation on training data, and also confusion matrices for these classifications.

7.5.1 Dataset

We used a subset of Turkish movie reviews as dataset and manually labelled 1,000 randomly chosen documents from the dataset as positive, negative, or neutral.We also labelled 2,700 sentences appearing in these documents as positive, negative, or neutral. The distribution of [positive, neutral, and negative] sentences and documents are close: [50%, 30%, 20%] and [52%, 29%, 19%] respectively[2]. Finally, we also manually labelled all appeared aspects in the above mentioned sentences, which resulted in about 2,000 aspect mentions labelled as positive, negative or neutral.

We did not include the label "mixed" in our labelling; instead we chose the dominant sentiment in a mixed review and labelled it accordingly.

7.5.2 Dealing with Unbalanced Data

As mentioned above, our dataset is unbalanced in favour of positive reviews, which causes biased results for positive samples (sentences and documents) during the classification. To avoid this problem, we balanced the dataset by re-sampling under-represented classes. This technique increased per-class classification accuracies (Tables 7.13, and 7.14), while the overall accuracy over all classes did not change much.

7.5.3 Results

The accuracies obtained from binary and ternary classifications on sentence and document levels are presented in Tables 7.11 and 7.12. Using all features, we obtained 73.42% and 79.06% accuracies in binary sentence and document classification problems, respectively. For ternary classification, results are 60.33% and 73.01% for sentence and document levels. As expected, higher accuracies are achieved at document level (due to larger context) and binary classification problems (simpler problem).

We also performed an aspect-based sentiment analysis and achieved *70%* and *79%* accuracies in ternary and binary classifications, respectively. The method for aspect classification has been explained in Section 6.1.2.3 in Chapter 5.

Considering a simple classification system which uses only the positivity and negativity scores of words that would correspond to features $f_1 - f_2$ as a baseline, we could increase the classification accuracy over the baseline by about 4 percentage points, at document level (75.04 vs 79.06 and 69.30 vs 73.01%).

[2]This subset is available from Sentilab website at http://sentilab.sabanciuniv.edu/resources/TurkishMovieRevi

Table 7.11: **Sentence level binary and ternary classification accuracy (%) by Logistic Regression using 5-fold Cross Validation .**

Feature Subset	Binary	Ternary
f_1-f_2	59.73	59.33
f_3-f_4	59.00	58.74
f_5-f_6	63.24	59.61
f_7-f_8	51.79	49.20
f_9-f_{12}	51.50	59.20
f_{13}-f_{16}	57.99	59.07
f_1-f_4	59.73	60.00
f_1-f_6	70.05	60.12
f_1-f_8	70.40	60.08
f_1-f_{12}	72.28	60.14
all : f_1-f_{16}	**73.42**	**60.33**

Table 7.12: **Document level binary and ternary classification accuracy (%) by Logistic Regression using 5-fold Cross Validation.**

Feature Subset	Binary	Ternary
f_1-f_2	75.04	69.30
f_3-f_4	76.57	70.70
f_5-f_6	75.68	70.61
f_7-f_8	51.01	48.42
f_9-f_{12}	74.15	69.12
f_{13}-f_{16}	73.50	69.10
f_{17}-f_{20}	78.02	72.30
f_1-f_4	77.44	71.10
f_1-f_6	77.50	71.22
f_1-f_8	78.25	71.20
f_1-f_{12}	78.42	71.34
f_1-f_{16}	78.64	71.51
all : f_1-f_{20}	**79.06**	**73.01**

The confusion matrix for both binary and ternary classifications are given in Tables 7.13 and 7.14. Each value in these tables shows the per-class accuracy (diagonal values in matrix), separately for positive, negative, and neutral classes in ternary classification and for positive and negative classes in binary classification.

Misclassification of sentences/documents are due to different reasons such as lack of background knowledge. A sample misclassified sentence is provided below.

> 5 puan verdim, o da janistonun güzel yüzünün hatırına.
> (I gave 5 points, and that because of the lovely character of Janiston).

Table 7.13: **Confusion matrix for binary classification of sentences and documents.**

Document level		
True/Estimated	positive	negative
positive	0.86	0.14
negative	0.27	0.73
Sentence level		
True/Estimated	positive	negative
positive	0.92	0.08
negative	0.67	0.33

Table 7.14: **Confusion matrix for ternary classification of sentences and documents.**

Document level			
True/Estimated	positive	negative	neutral
positive	0.67	0.20	0.13
negative	0.15	0.81	0.04
neutral	0.18	0.17	0.75
Sentence level			
True/Estimated	positive	negative	neutral
positive	0.62	0.19	0.19
negative	0.09	0.86	0.05
neutral	0.30	0.41	0.29

In this example, our system cannot distinguish "5 points" (out of 10) as a low grade for a movie and therefore misclassifies this negative sentence as positive because of the positive phrase in it.

7.5.4 Discussion on Proposed Turkish Sentiment Analyser

As seen in Tables 7.11 and 7.12, the obtained accuracies in different cases range from 60% to 79%. Considering the results, we came up with the following conclusions.

- Document level sentiment analysis is more successful compared to sentence level, as expected. The intuition is that correctly classified sentences in a document compensate for misclassified sentences.
- The most effective group of features at binary sentence level task are $f_5 - f_6$ (number of positive and negative words in PWS). As PWS contains only

positive and negative words, these features are not very effective in ternary classification.

- The most effective features at document level in isolation are $f_{17} - f_{20}$ (polarity of the first and last sentences). This observation is in agreement with the assumption that the first and last sentences in a document are the best estimators of the document polarity. This was also cited in literature by a few researchers such as Meena and Prabhakar [2007] and Gezici et. al [2012]. In fact, the difference in classification accuracy between using only the polarity of the first and last sentence, and using all features (in document level) is less than one percentage point.

- The least effective feature set in isolation is $f_7 - f_8$ (emoticons) for both sentence and document level analyses.

- In almost all settings, each added feature subset improves the accuracy over the existing features. For example, adding $f_{17} - f_{20}$ to feature group $f_1 - f_{16}$, increases the accuracy by one percentage point.

- Generally, our system is more successful in classifying positive sentences and documents compared to negative or neutral ones.

- Our approach improves upon the simple baseline of using average word polarities (features $f_1 - f_2$) in the review by about four percentage points.

- As mentioned in Chapter 6, ITÜ parser may not parse all sentences correctly. This will affect our system in a negative way, but because the effect of adding parser-based features ($f_{13} - f16$) is only 1-2 percentage points, erroneous cases in the parser will not have much effect on our proposed methodology. Even getting rid of the parser in the proposed methodology will decrease the classification accuracy only by about 1.5 percentage points.

We could not apply other methods in the literature on our dataset because none of the previous work have released their detailed approach or dataset used for experiments. Moreover, related research report only binary classification results which neglects neutral reviews, while we consider both binary and ternary classifications.

Similar work to ours are [20] and [24], which have reported 76% and 85% accuracy in classifying Turkish movie reviews as positive and negative. The comparable accuracy (binary document classification) in our work is 79%; however these accuracies may not be directly comparable as the details of how they used the dataset are unknown. Moreover, previous work focus on document level sentiment analysis, while we consider aspect and sentence levels as well.

Chapter 8

CONCLUSIONS AND FUTURE WORK

Due to ever-increasing amount of information, it is a necessity to design (semi)automatic techniques for analysing them. One type of analysis is to extract the embedded sentiment from data (text in our case). SA is a technique for extracting the polarity from (mostly textual) data, which has been analysed in this dissertation.

Although a good deal of research has been accomplished on SA in recent decades, the current state is far from perfect. On the other hand, the request for SA systems is increasing in industry because almost all companies are interested to know their customers' ideas towards services or products.

In spite of fast growth of techniques, tools and resources for SA in English, most of other languages suffer from the shortage of research in this area. In order to fill this gap, we focused on the Turkish language. We comprehensively studied the SA problem in Turkish and highlighted sub-problems, specially those that need more attention. Although we were unable to solve all sub-problems in the current work, we suggested at least partial solutions for them.

In chapter 2, we had an overview of the problem and preliminary issues; Chapter 3 reported some related work in SA of Turkish, English, and other languages. Chapter 4 discussed polarity lexicon generation methods; Chapter 5 dealt with NLP issues and granularity levels; in Chapter 6, the SA problem in Turkish was comprehensively discussed which resulted in a SA tool for Turkish. Experimental evaluation is accomplished in Chapter 7 and finally Chapter 8 provides conclusions and suggest some future work.

Our proposed SA system can be employed by companies that require processing a large number of customer reviews in Turkish; it can also be employed to analyse Turkish tweets and comments regarding a political issue.

Suggested future work for this dissertation are listed below:

- *Idiom handling* which attempts to benefit from appeared idioms in a sentence/review to estimate the polarity of that sentence/review. Idioms are a group of words established by usage as having a meaning not deducible from those of the individual words (e.g. over the moon, see the light)
- *Sarcasm detection* which may be the most difficult problem in SA and NLP. The goal is to detect sarcastic sentences. For example the expression "harika olmuş" [it has been wonderful], at the end of a totally negative review is a sarcastic expression.
- *Spam detection* which aims to detect spam (fake) reviews. Fake reviews are written intentionally in favour of an entity or for underestimating it. For example too many negative reviews towards the products of a successful company may be spam reviews posted by its opponent companies or other spammers. There are even companies for distributing such reviews.
- *Investigating the sentence type* such as comparative, interrogative, or conditional sentences which can affect the sentiment included in the sentence. We did not touch comparative sentences but superficially considered other types (interrogative and conditional). This sub-problem lacks a good deal of research specially in Turkish.
- *Intention analysis* which attempts to extract the intention of the author from her review. For example the sentence "I need a new laptop, any suggestion?" implies that the author would like to buy a new laptop. This issue can be useful specially in marketing as suggesting new products/services to interested customers can increase the sale ratio of a company.

If you have any feedback on this dissertation, please contact the first author.[1]

[1]Please contact me via *rdehkharghani@sabanciuniv.edu*

Appendices

Appendix A

Polarity Lexicons

In the appendix, we list one table from each resource that we have generated namely, polar word list, polar adjectives, polar phrase lexicon, and SentiTurkNet. We have experimented our proposed approach in Chapter 5 on Turkish movie revews (beyazperde.com); here we provide a small subset of sentences which are manually labelled as positive, negative, and neutral (objective). The whole resources can be downloaded from http://sentilab.sabanciuniv.edu/resources/.

TABLE A.1: **A subset of polar (positive/negative) word list in Turkish.**

Negative	Positive
saçma	hakediyor
mutsuz	istekli
iştahsız	merhamet
abart	sabırlı
abartılı	memnun
lanet	sempati
üzgün	sevinc
kusurlu	gururlu
nasıl	heyecanlı
malesef	kahraman
karanlık	şampiyon
depresif	hayran
kalitesiz	mutlu
yazık	keyifli
ne yazk	kahkaha
aptal	korkusuz
klişe	kazan
lezzetsiz	başarılı
keyifsiz	hediye
terbiyesiz	ödül
hatalı	canlı
hata	ölümsüz
ilgisiz	suçsuz

TABLE A.2: **A subset of polar (positive/negative) adjective list in Turkish.**

Positive	Negative
atletik	alaycı
uğurlu	alaylı
aydın	alak
aydınlatıcı	alçakça
aydınlık	aldatıcı
gelişmiş	alengirli
azami	aleyhtar
azat	alk
azatlık	çalımlı
azimli	alıngan
aziz	alışılmadık
soylu	alışılmam
babayiğit	alışmış
bağdaık	alkollü
baarılı	allahsız
başlı	allak
becerikli	bullak
bedelsiz	pullu
beğenir	sarısı
belirgin	amansız
bereketli	andavallı
berrak	ankastre

TABLE A.3: **A subset of polar (positive/negative) phrase list in Turkish.**

Positive	Negative
üstüne yok doğrusu	daha güzel olabilirmiş
özellikle sakin kafa	acı bir yön
çok başarılı bir performans	sonucu belirli olmadı
ekstra güzellik katmış	bir anlam veremiyorum
güzel sahnesi çok var	hep olumsuz şeyler
harika bir yapım	Sıradan bir eser
etkileyecek bir konu	abartılacak bir şey yok
güzel bir yapım izlenebilir	çoğunlukla kötü geldi
kötü de değil	bir kara haber
çok iyi olmuş	olmamş en kötü
bence gerçekten olmuş	çaresiz bir durum
güzel bir konuya sahip	sıkıntılı saatler başlar
çok güzel anlatıyor	güzel bir iş diyemem
verdiği baygınlıktan olacak	büyük bir ayıp
başarılı bir şekilde	daha iyi olabilirdi
iyi seyirler dilerim	tavsiye edemem yalnız
kesin kaçırmayın bu filmi	sıradan bir eser
gerçekten tavsiye ederim	basit bir hikaye
eğlenceli birşey arıyorsanız	sert tepkiler gelebilir
yapılmış en iyi	çok saçma olmuş
en iyi sistem	bir anlamı kalmaz
farklı bir senaryo	yapacak bir şey yok

TABLE A.4: **A subset Turkish synsets in SentiWordNet.**

Synonyms	Neg.	Obj.	Pos.
kaçış , kaçma , firar	0,575	0,357	0,068
gösteri , numara	0,06	0,872	0,068
tesadüfen	0,06	0,872	0,068
bitiştirmek , yanaştırmak	0,06	0,872	0,068
süslenip püslenmek	0,06	0,06	0,88
güzel giyinmek , şık giyinmek , şık şık giyinmek	0,06	0,06	0,88
aktif , etkin , faal	0	0	1
maalesef , ne yazık ki	0,49	0,442	0,068
giyinmek	0,06	0,872	0,068
ulaşma , varma , vusul	0,06	0,872	0,068
varış	0,06	0,872	0,068
sönmüş	0	1	0
doğuş , ortaya çıkma , zuhur	0,06	0,462	0,478
giriş , duhul	0,06	0,872	0,068
kilo vermek , incelmek , zayıflamak	0	0,083	0,917
kilo almak , şişmanlamak	0,731	0,208	0,062
giyinmek	0,06	0,872	0,068
kayıt	0,06	0,872	0,068
mevcut , fiili	0,06	0,012	0,928

TABLE A.5: **A subset of labelled sentences in Turkish movie reviews.**

Sentence	Label
aniston nasıl böyle saçma bir sahneli film de rol aldı anlayamadım	n
5puan verdim o da anistonun güzel yüzünün hatırına	n
son derece sıkıcı bir filim olduğunu söyleyebilirim	n
..saçma bir konuyu nasılda filim yapmışlar maşallah	n
bence hoş vakit geçirmek için seyredilebilir	p
hoş ve sevimli bir film	p
itici bir film değildi sonuçta	o
seyrederken bu kadar sinirlendiğim film hatırlamıyorum	n
J.Aniston ın hiç mi umut yok diye sorduğu sahnede kıracaktım televizyonu!	o
kimse yazmamış ben yazıyım:)	o
güzel bi pazar günü şirin bi film izlemek isteyenler için çok güzel	p
ama daha fazlası yok	n
biraz da durum orjinal işte	o
film tam benim genç kızlık dönemimde geçiyor	p
o yılların müziklerini filmde duymak çok hoşuma gitti	p
bana nostalji yaptırdı bu film	o
o yılların saflığı sevgisi çok güzeldi	p
sevdim bu filmi çok sıcak ve çok sevimli hoş bir filmdi	p
güzel izlenebilir bi film	p
Akıcı klsik bi tarzda kurgusu var iyi idare eder	p
80 lerin saçları bomba bunu bi kez daha hatırladım	o
Filmde Drew ile bir numara oluyor	p
zledi]ugim en iyi üç romantik komediden biri	p
zaman kaybındam başka bişi değil	n

Bibliography

[1] Bing Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[2] Bing Liu. *Sentiment Analysis*. Cambridge University Press, NY, USA, 2015.

[3] Kemal Oflazer and H Cem Bozşahin. Turkish natural language processing initiative: An overview. In *Middle East Technical University*. Citeseer, 1994.

[4] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.

[5] George A Miller. Wordnet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

[6] Erik Cambria, Robert Speer, Catherine Havasi, and Amir Hussain. Senticnet: A publicly available semantic resource for opinion mining. *Artificial Intelligence*, pages 14–18, 2010.

[7] Saif M Mohammad and Peter D Turney. Crowd sourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013.

[8] Veselin Stoyanov, Claire Cardie, and Janyce Wiebe. Multi-perspective question answering using the OpQA corpus. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 923–930. Association for Computational Linguistics, 2005.

[9] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! In *ICWSM*, pages 538–541, 2011.

[10] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12, 2009.

[11] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, pages 30–38. Association for Computational Linguistics, 2011.

[12] Rahim Dehkharghani and Cemal Yilmaz. Automatically identifying a software product's quality attributes through sentiment analysis of tweets. In *Natural Language Analysis in Software Engineering (NaturaLiSE), 2013 1st International Workshop on*, pages 25–30. IEEE, 2013.

[13] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79–86, 2002.

[14] Arun Meena and TV Prabhakar. Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. pages 573–580, 2007.

[15] Xiaowen Ding, Bing Liu, and Philip S. Yu. A holistic lexicon-based approach to opinion mining. In Marc Najork, Andrei Z. Broder, and Soumen Chakrabarti, editors, *WSDM*, pages 231–240. ACM, 2008.

[16] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics, 2005.

[17] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, pages 399–433, 2009.

[18] Lingjia Deng and Janyce Wiebe. Sentiment propagation via implicature constraints. In *Proceedings of EACL*, 2014.

[19] Ezgi Yıldırım, Fatih Samet Çetin, Gülşen Eryiğit, and Tanel Temel. The Impact of NLP on Turkish Sentiment Analysis. *Türkiye Bilişim Vakfi Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 8(8), 2015.

[20] A. Gural Vural, Berkant Barla Cambazoğlu, Pınar Şenkul, and Z. Özge Tokgöz. A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish. In Erol Gelenbe and Ricardo Lent, editors, *ISCIS*, pages 437–445. Springer, 2012.

[21] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

[22] Mesut Kaya, Guven Fidan, and Ismail H Toroslu. Sentiment analysis of Turkish political news. In *Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 174–180. IEEE Computer Society, 2012.

[23] Çiğdem Aytekin. An opinion mining task in Turkish language: A model for assigning opinions in Turkish blogs to the polarities. *Journalism and Mass Communication*, 9(3(3)):179–198, 2013.

[24] Umut Eroğul. Sentiment analysis in Turkish. *MSc thesis, Middle East University, Turkey*, 2009.

[25] Ahmet Afşin Akın and Mehmet Dündar Akın. Zemberek, an open source NLP framework for Turkic languages. *Structure*, 10, 2007.

[26] Zeynep Boynukalın. Emotion analysis of Turkish texts by using machine learning methods. *MSc thesis, Middle East Technical University, Turkey*, 2012.

[27] Hung-Yu Kao and Zi-Yu Lin. A categorized sentiment analysis of chinese reviews by mining dependency in product features and opinions from blogs. In Jimmy Xiangji Huang, Irwin King, Vijay V. Raghavan, and Stefan Rueger, editors, *Web Intelligence*, pages 456–459. IEEE, 2010.

[28] Amitava Das and Sivaji Bandyopadhyay. Sentiwordnet for indian languages. *Asian Federation for Natural Language Processing, China*, pages 56–63, 2010.

[29] Julian Brooke, Milan Tofiloski, and Maite Taboada. Cross-linguistic sentiment analysis: From English to Spanish. In *RANLP*, pages 50–54, 2009.

[30] Robert Remus, Uwe Quasthoff, and Gerhard Heyer. Sentiws-a publicly available german-language resource for sentiment analysis. *LREC*, 2010.

[31] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics, 2002.

[32] Vasileios Hatzivassiloglou and Kathleen R McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 174–181, 1997.

[33] Justin Martineau and Tim Finin. Delta tfidf: An improved feature space for sentiment analysis. In *ICWSM*, 2009.

[34] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM, 2004.

[35] Rahim Dehkharghani, Yucel Saygin, Berrin Yanikoglu, and Kemal Oflazer. SentiTurkNet: a Turkish polarity lexicon for sntiment analysis. *Language Resources and Evaluation (to appear)*, 2015.

[36] Soo min Kim. Determining the sentiment of opinions. In *Proceedings of COLING*, pages 1367–1373, 2004.

[37] Hiroshi Kanayama and Tetsuya Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 355–363. Association for Computational Linguistics, 2006.

[38] Yunfang Wu and Miaomiao Wen. Disambiguating dynamic sentiment ambiguous adjectives. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 1191–1199. Association for Computational Linguistics, 2010.

[39] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.

[40] Yejin Choi and Claire Cardie. Adapting a polarity lexicon using integer linear programming for domain specific sentiment classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 590–598, 2009.

[41] Vasileios Hatzivassiloglou and Kathleen R. Mckeown. Predicting the semantic orientation of adjectives. In *Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics*, pages 174–181, 1997.

[42] Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A high-coverage lexical resource for opinion mining. Technical report, 2007-TR-02, Istituto di Scienza e Tecnologie dellInformazione, Consiglio Nazionale delle Ricerche, Pisa, IT, 2007.

[43] Orhan Bilgin, Özlem Çetinoğlu, and Kemal Oflazer. Building a wordnet for Turkish. *Romanian Journal of Information Science and Technology*, 7(1-2): 163–172, 2004.

[44] David W Hosmer Jr and Stanley Lemeshow. *Applied logistic regression*. John Wiley & Sons, New Jersey, USA, 2004.

[45] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall Upper Saddle River, NJ, USA, 1994.

[46] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[47] Geoffrey Holmes, Andrew Donkin, and Ian H Witten. Weka: A machine learning workbench. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, pages 357–361. IEEE, 1994.

[48] Tom M. Mitchell. *Machine learning*. McGraw Hill series in computer science. New York, USA, 1997. ISBN 978-0-07-042807-2.

[49] Seslisözlük-group. Sesli sözluk, bilingual english-turkish-english dictionary, 2014. URL http://www.seslisozluk.net.

[50] Christopher D Manning and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, Cambridge, MA, USA, 1999.

[51] Gülşen Eryiğit. ITU Turkish NLP web service. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Sweden, April . Association for Computational Linguistics.

[52] Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, 19(2):313–330, 1993.

[53] Gizem Gezici, Berrin Yanıkoğlu, Dilek Tapucu, and Yücel Saygın. New features for sentiment analysis: Do sentences matter? In *SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data*, page 5, 2012.