AFFECT ANALYSIS OF COMMUNICATION TEXTS IN TURKISH LANGUAGE

by

Eda Aydın Oktay

B.S., Physics, Boğaziçi University, 2012

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computational Science and Engineering
Boğaziçi University
2016

ACKNOWLEDGEMENTS

I would like to express my sincerest appreciation to my supervisor, Assist. Prof. Albert Ali Salah, whose supervision and scholastic acumen have greatly influenced not only this thesis work, but also my vision and academic perspective. I owe my deepest thanks for his friendly teaching style, for his invaluable patience throughout the study and for his availability whenever I needed. I am very thankful for having the chance of working with him.

Additionally, I would like to thank the rest of my thesis committee, Assist. Prof. Arzucan Özgür and Assist. Prof. Sibel Halfon for their valuable time, insightful remarks and criticisms.

This work is partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) project 114E481, ‘Player profiling and communication analysis for player complaints in social network based games’ and project 7141188, ‘Text analysis and information visualization based decision support system for student counseling in schools.’

Furthermore, I would like to thank to the 110 volunteer participants of affective data annotation for their kind interest and effort in our research.

I would also like to thank my friends Ahmet Emre Aladağ, Caitlin M. Miles, Seher Uğurcuklu, Ayşegül Şahin, Ayşe Demirel and Kamil Çöllü for their support and friendship.

I am very grateful and would like to thank my family for their encouragement, for unconditionally supporting me and for being the source of my motivation. Most importantly, I would like to thank my husband Ahmet Afşin Oktay, for never leaving me alone and helping me to keep my motivation high through these tough times.

ABSTRACT

AFFECT ANALYSIS OF COMMUNICATION TEXTS IN TURKISH LANGUAGE

As digital technologies expanded our range of online communication, textual affect analysis become the focus of considerable interest in several disciplines. While many studies have been conducted in English language, there are only a few tools specific for the Turkish language. One reason for that is the lack of lexical resources for affect analysis and annotated corpora for an accurate evaluation. In this thesis, we develop an approach for continuous and dimensional affect analysis of Turkish communications by combining several tools. We conduct experiments on various text corpora in Turkish including online multi-party chat records, psychotherapy records, Twitter data, movie reviews, and teachers' comments on high school students. Analyzing such texts brings challenges like non-standard word usage, grammatical irregularities, abbreviation usage, and spelling mistakes. We propose several pre-processing steps to deal with these. Then, we adapt an affective word dictionary from English to Turkish, and by expanding it with synsets, obtain 15,200 words with annotations for valence, arousal, and dominance. We also employ a list of frequently used abbreviations, emoticons, interjections, modifiers (intensifiers and diminishers), and other linguistic indicators to capture the overall affective state at the sentence level. We recruit and train annotators to obtain affective ground truth specifically for multi-party chat records. Our results show that the proposed system is useful, yet there is much room for improvement at different stages.

ÖZET

TÜRKÇE İLETİŞİM METİNLERİ ÜZERİNE DUYGULANIM ANALİZİ

Dijital teknolojilerin çevrimiçi iletişim yelpazemizi genişletmesiyle birlikte, metinsel duygulanım (afekt) analizi bir çok disiplinde önemli bir ilgi odağı haline gelmiştir. Bu alanda İngilizce dili üzerine birçok çalışma yapılmış olsa da, Türkçe diline yönelik oldukça kısıtlı sayıda çalışma bulunmaktadır. Bunun bir nedeni duygulanım analizi için gereken sözlüksel (lexical) kaynakların ve sistem değerlendirmesi için gerekli olan referans derlemlerin (corpora) yetersiz olmasıdır. Bu çalışmada, birçok aracı birleştirerek Türkçe iletişim metinlerinin sürekli ve boyutsal duygulanım analizini gerçekleştirecek bir yaklaşım sunuyoruz. Twitter verisi, film eleştirileri ve lise öğrencilerine yönelik öğretmen yorumları, çevrimiçi sohbet ve psikoterapi kayıtları gibi birçok Türkçe derlemler üzerinde bu modeli test ettik. Çevrimiçi sohbet kayıtlarını analiz etmek, dilbilgisel düzensizlikler, kısaltmalar, yazım yanlışları ve yapısal olmayan bir dil kullanımı gibi sorunları da beraberinde getirmektedir. Bu problemlerin üstesinden gelebilmek amacıyla birçok ön veri işleme tekniği gerçekleştirdik. Sonrasında, İngilizce için hazırlanmış bir duygulanımsal kelime sözlüğünü Türkçe'ye adapte ederek ve eş anlamlı kelime gruplarıyla genişleterek, memnuniyet (valence), uyarılma düzeyi (arousal) ve baskınlık (dominance) boyutları için ölçeklendirilmiş 15,200 kelimelik bir kaynak elde ettik. Aynı zamanda, bir cümlenin genel duygulanım durumunu yakalayabilmek için sık kullanılan kısaltmaların, emotikonların, özel isimlerin, pekiştirici kelimelerin birer listesini bazı dilsel sinyallerle birlikte kullandık. Sohbet kayıtları üzerine duygulanımsal bir değerlendirme refaransı oluşturabilmek için bir cümle işaretleme anketi düzenledik. Son olarak elde ettiğimiz sonuçlar kural ve kelime bazlı modelimizin faydalı olduğunu göstermekle birlikte birçok farklı aşama hala geliştirilmeye açıktır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

c	Judgement order
$d_{k_a k_b}$	Distance between each pair of distinct values
D_o	Observed disagreement of Krippendorff's alpha
D_o	Expected disagreement of Krippendorff's alpha
k	Unique category for coder's scores
k_a, k_b	Each pair of distinct values
n_{ik}	The number of coders that make judgement k
n_{ik_a}, n_{ik_b}	Pairs of judgements of item i
S_{affect}	Overall affective score of a sentence
$S_{polarity}$	Overall polarity of a sentence
α	Krippendorff's inter annotator agreement rate
$\alpha_{interval}$	Krippendorff's inter annotator agreement rate with interval metric
ω_{affect_i}	Affective score of i th unit
$\omega_{valence}$	Valence score of the unit
$\omega_{arousal}$	Arousal score of the unit
$\omega_{dominance}$	Dominance score of the unit

LIST OF ACRONYMS/ABBREVIATIONS

AffDic	Affective Dictionary
API	Application Programming Interface
ANEW	Affective Norms for English Words
CMC	Computer Mediated Communication
CNB	Complement Naive Bayes
DF	Document Frequency
F2F	Face to Face Communication
JSON	Java Script Object Notation
IR	Information Retrieval
KNN	K Nearest Neighbor
LIWC	Linguistic Inquiry and Word Count
ML	Machine Learning
mRMR	Minimum Redundancy Maximum Relevance
NB	Naive Bayes
NLP	Natural Language Processing
PANA	Positive Activation Negative Activation
POS	Part-of-Speech
SA	Sentiment Analysis
SAM	Self-Assessment Manikins
SVM	Support Vector Machine
TDK	Türk Dil Kurumu
TF-IDF	Term Frequency - Inverse Document Frequency
UTF-8	8-bit Unicode Transformation Format
VAD	Valence Arousal Dominance
VSM	Vector Space Model
WSD	Word Sense Disambiguation

1. INTRODUCTION

With the advent of new automated computer applications and improvements in human-computer interaction, the way people communicate is also changing. A few decades ago, computers were basically being used for science and business but in today's world, internet is also used for more social purposes. The prolific use of various communication technologies such as emails, text messages, or instant message chats has led to improved collaboration among different groups independent of time and space. This new age of communication fosters growing interest in research areas such as affective computing [1]. In this field, affect is an important key in understanding human behavior and can be applied in areas such as human interaction, communication, entertainment, design and health. Affective meaning can be conveyed and interpreted by numerous ways using physiological and behavioral signals [2], such as using facial expressions [3], body gestures [4], heart rate, text, speech, or cognitive and semantic cues [5]. There is also research to evaluate affective messages in online communication systems, such as social media and online chat. This highlights the need for further research on analysis of affect in online communication. Building automated systems which can recognize, interpret, and process textual affect will advance new technologies in online products and services, as well as enhancing the quality of online communications with more powerful and enjoyable computer interfaces. Textual analysis is comparatively less challenging in this domain and texts are emotionally rich information resources. Indeed, recent advances in sensing affect in text can strengthen understanding affect in other modes of communication as well, such as facial expressions and speech [2,6]. Therefore we find that text is a valuable tool for detecting and sensing affect for the improvement of various existing and potential applications.

In this study, we aim for assessing affect and sentiment in online communication texts by achieving automatic analysis in various domains. Our model is designed for the texts in Turkish language. This model was used to cope with abbreviated and informal chat language and to capture a wide range of affective features. One major need in analyzing Turkish texts is the lack of tools that are as advanced as existing tools for the English language. To remedy this, we adapted an affective lexicon from English to Turkish, and used it to analyze the chat corpus. We created an affect analysis model, which uses 120 emoticons, 98 abbreviations, 50 interjections, 71 modifiers (intensifiers & diminishers), and the adapted affective lexicon that includes more than 15,000 words. We also performed multiple annotation tasks to construct affective ground truth to assess the accuracy of the ensuing system. We have acquired a set of chat record for analysis, and a group of human annotators labeled them based on subjective affective scores of each sentence. The proposed affective analysis model can be used as a stand alone module, or can be integrated with multimodal approaches to analyze computer mediated communication or any textual data in Turkish.

The first application of our affect model was the analysis of affect in multi-party in-game chat logs. One advantage of using such texts is the emotionally rich characteristic of in-game chat records, however, there is no comprehensive study specific to this domain. Our model achieves an accuracy of 73% for course-grained affect estimation of chat records. Although the initial design of this system was targeting the in-game chat records [7], we came up with a system to provide more generalized solutions for analyzing affect and sentiment in various domains. We performed dimensional affective analysis of psychotherapy records with the proposed model. Psychotherapy records that we analyze contain the conversations between a psychotherapist and a patient during several anonymous therapy sessions. The results indicate that there are significant correlations between the predicted affect scores and the ground truth affect annotations on various factors. Our third application was assessing the sentiment analysis of Turkish movie reviews. With this data set, we observed the accuracy over 86% for binary (positive and negative) sentiment classification. This result outperforms the state-of-the art performance for sentiment analysis of movie reviews. Our fourth application is the opinion mining on anonymous and private teacher comments based on high school students evaluations. We observed 80% accuracy with this data set. The last application domain is Twitter, which is one of the most well known micro-blogging services. With Twitter data, our model achieves accuracy of 82% for the sentiment classification in two categories: positive and negative.

The rest of the thesis is organized as follows: In Chapter 2, we present the background and related work in the literature. In addition to the theoretical background information, we overview the existing methodologies and applications on textual affect analysis both for English and Turkish languages. In Chapter 3, we explain the textual data and resources that we have used, how the data annotation scheme performed and we report inter-annotator agreement for reliability of subjective annotation. In Chapter 4, affective model and architecture of the system is described in detail. In Chapter 5, we describe the experiments that we performed and discuss the results. Finally, in Chapter 6, we conclude the work with a summary and some future work.

2. BACKGROUND AND RELATED WORK

This chapter describes background and reviews the state of art research in textual affect analysis to gain more insight into the model and the experiments that we present in this thesis.

2.1. What is emotion?

Emotion creates a powerful motivational system that has an impact on perception, cognition, coping and creativity in important ways [8]. Emotion has been studied in several scientific disciplines such as psychology, sociology, cognitive sciences and neuroscience. Although emotion has been widely studied, there is still a lack of a generally accepted definition. Merriam-Webster Online Dictionary gives the definition of emotion as 'the affective aspect of consciousness; a state of feeling; a mindful mental reaction (as anger or fear) individually experienced as strong sensation usually directed toward a particular object and typically accompanied by physiological and behavioral changes in the body'.

From the perspective of a subjective or experiential position, emotion is described with affective and cognitive definitions; the cognitive aspect highlights perceptual/thinking, particularly appraisal, and/or labeling processes, whereas the affective aspect maintains feelings of excitement/depression and pleasure/displeasure [9]. According to cognitive theory proposed by Ortony, Clore, and Collins (OCC), emotions are considered as reactions to situational appraisals of events, actions, and objects [10]. While the *Jamesian* perspective of emotion interprets emotion as experience of physical and bodily responses [11], *social constructivist theory* links the emotions to social rules and culture [12].

There is a common confusion in the literature for the usage of terms 'affect', 'emotion' and 'sentiment', which are the terms that we frequently mention in this work. To clarify the usage of these subjectivity terms and clear the distinction between their meaning, let us discuss their relations briefly.

In Merriam-Webster Online Dictionary sentiment is defined as 'an attitude, thought, or judgement prompted by feeling; refined feeling; an idea colored by emotion.' E. Shouse states, 'an affect is a non-conscious experience of intensity; it is a moment of unformed and unstructured potential' [13]. We understand that affect is an unconscious, and more stable experience of emotion and sentiment is accepted as a more organized attitude produced by emotions.

There is an explanatory illustration (Figure 2.1) of Munezero et al. which characterizes these related terms. Affect is shown as a predecessor to feelings and emotions; feelings are person-centered, conscious phenomena, while emotions are pre-conscious social expressions of feelings and affect influenced by culture [14].

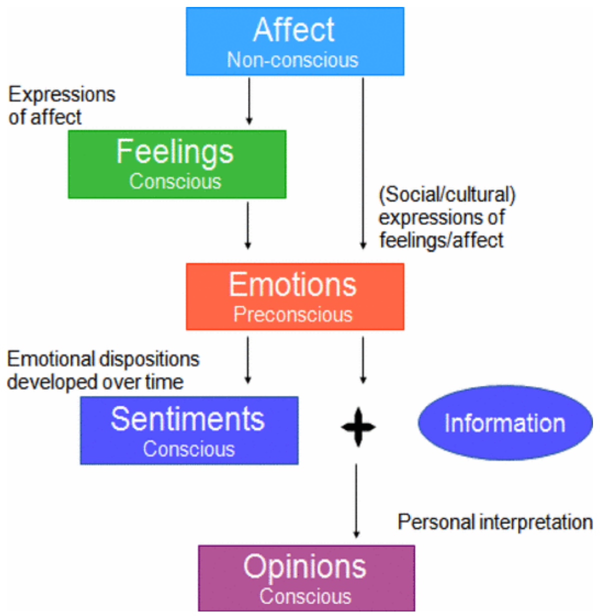


Figure 2.1. Differentiating factors between affect, feelings, emotions, sentiments and opinions.

Affect display refers to external expression of emotion, which is an important aspect of communication. Since we only deal with the observable form of affect in this work, what we mean by the term 'affect' is actually the displayed cues of affect.

Standing on this affective aspect of emotion, most of the time we used terms ‘affect’ and ‘emotion’ interchangeably.

2.2. Modeling Affect

As a result of multidisciplinary approaches, affect has been represented and modeled in various ways with different perspectives. Two main approaches are categorical and dimensional modeling [15, 16].

Categorical modeling (or discrete emotion model) is based on a taxonomy of emotions under discrete categories. One of the most influential studies towards describing emotion categories is conducted by Ekman [17] who identifies six basic universal emotions; happiness, surprise, sadness, fear, disgust, anger. A majority of affect recognition systems rely on Ekman’s six basic facial emotion categories in the literature. However, there are many approaches that describe alternate categories of emotions. For example, Izard stated that the basic emotions can be distinguished with their unique motivational properties and proposed ten primary emotions as; anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise [18]. On the other hand, Plutchik defined the basic emotions as; acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise [19].

In the **dimensional modeling**, the assumption is that emotions are related to each other and the affective state is investigated in continuous multidimensional space (generally two or three dimensions). A pioneering approach in dimensional modeling is proposed by Wundt who described emotions in three dimensions (pleasurable vs. unpleasurable, arousing vs. subduing and strain vs. relaxation) [20]. Osgood [21] proposed a three-dimensional evaluation, potency, activity (EPA) model based on the semantic differential technique. In this model, evaluation specifies the degree of pleasantness, potency specifies intensity level the word, and activity indicates active or passive state of the word. In 1980, Russell developed *the circumplex model of affect* [22] by utilizing both dimensional and categorical representations. This model places emotion categories on a two-dimensional circular space, where the horizontal axis scales with the valence dimension and the vertical axis scales with the arousal dimension. The origin of axes indicates neutral valence and neutral arousal. An example representation of the circumplex model of affect is given in Figure 2.2.

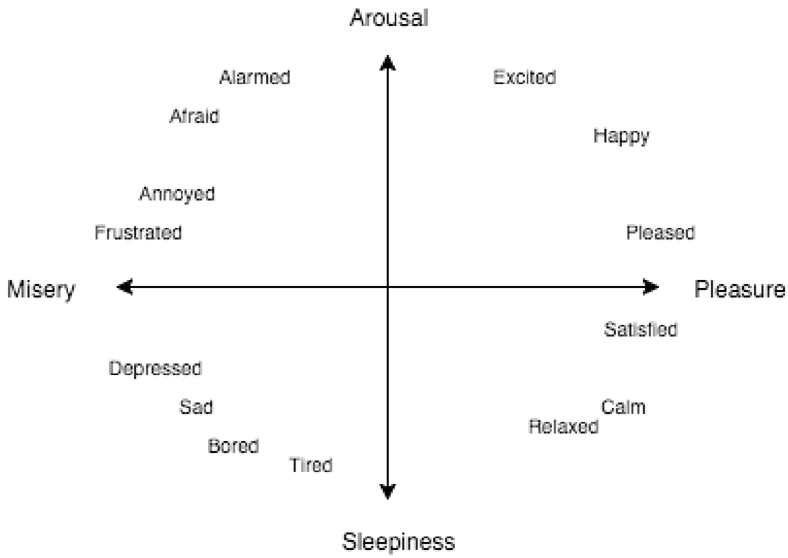


Figure 2.2. A graphical representation of the circumplex model of affect

Another dominant model is the ‘consensual’ positive activation - negative activation (PANA) model of emotion, which is proposed by Watson and Tellegan in 1985 [23]. According to this model, with 45-degree rotation of the axes in circumplex model, the vertical axis scales low to high positive affect and the horizontal axis scales low to high negative affect. In PANA positive activation axis is linked to mood terms like ‘elated’, ‘active’, ‘excited’ at high positive affect and ‘sleepy’, ‘dull’, and ‘sluggish’ at low positive affect. Similarly, negative activation axis is linked to ‘hostile’, ‘fearful’, ‘nervous’ and to ‘calm’, ‘placid’, ‘relaxed’.

One of the frequently used models is the two dimensional **vector model** that is proposed by Bradley et al. in 1992 [24]. According to this model, there is assumed to be underlying arousal and pleasantness (valence) dimensions which determine the direction of emotion. The circumplex model and the vector model are very similar but in contrast to circumplex model, high arousal is discriminated at positive and negative valences in the vector model. When detecting the affect in text, some studies incorporate Latent Semantic Analysis (LSA) [25] to measure the similarity between word usage in a particular text and defined affective categories [26–28].

On dimensional structure of emotion, Russell and Feldman Barrett represented the dimensions regarding a bipolar pleasure-displeasure and an orthogonal arousal dimension [29]. From another perspective, Thayer defined valence as a combination of energetic arousal and tense arousal in his multidimensional model of activation [30]. In addition to the models we have discussed, there exist several different approaches on dimensional structures [31–34]. However, there is still a lack of consensus on which dimensions are fundamental and which dimensions are a mixture of these basic dimensions [35, 36].

2.3. Affect Analysis in Online Communication

In the 1980s communication of emotions via computers is believed to be very difficult to express; in 1990s a lot of researchers have started to analyze the differences between computer mediated communication and face to face communication [37]. Differences between Computer Mediated Communication (CMC) and Face to Face Communication (F2F) are reviewed by Derks from the aspect of emotion communication under anonymity and reduced visibility. It is shown that despite some limitations of CMC such as absence of emotional embodiment, emotions can be communicated both explicitly and implicitly in an online platform very similar to face to face communications [37].

Current approaches for automatic emotional and affective content analysis from text generally include keyword spotting, lexical affinity, statistical natural language processing (NLP), learning based methods and commonsense-based approaches [38,39].

One of the most longstanding and simplistic methods is **keyword spotting**, which identifies a set of keywords to construct a look up table based on seed words and their corresponding emotional values. Although keyword spotting method is relatively easier to implement, basic limitation of this system is its incapability of dealing with negation and complex sentence structure. Therefore, most of the proposed keyword spotting systems incorporate rule-based methodologies to overcome these problems to some extent. Another point is that building a rich lexicon is a very expensive task. Moreover the emotion conveying words only form a small portion of a sentence, so this makes the system impractical, especially for complex domains.

Lexical affinity uses the mutual information of words based on their relationships in the document [39]. The aim is to link words that are relevant for certain affective dimensions and assign a probabilistic affinity value. Similar to keyword spotting, the disadvantage of lexical affinity is its inability to take the sentence level analysis into account, which makes it very limited in understanding complex and compound sentences.

One of the early applications for emotion detection which relies on keyword spotting was Text-to-Emotion Engine [40]. This system extracts emotions in real time Internet communications using lexical semantic analysis to identify both emotion category and emotion intensity. Apart from some grammatical rules and syntactical features, a special tag set is employed to classify the emotion. They perform static and dynamic tests to measure the performance of the engine and they have shown that the system gives a satisfactory result with some error caused by spelling mistakes. Ma et al. also proposed an emotion estimation system for chat or other dialogue domains, based on keyword spotting with sentence-level processing [41].

Recently, Dey et al. proposed a rule-based model for emotion extraction in a real-time chat messenger using an emotion lexicon [42]. After a POS Tagger is employed, the corresponding emotion is extracted based on the matches with the lexicon. Then their emotion analyzer executes word level analysis, phrase level analysis and sentence level analysis independently. The affective features of a word are represented as a vector of emotional state intensities. They define some rules to process a sentence for emotion analysis. For example, if there exist any verb and noun phrases with opposite valence, the verb is considered as the principal otherwise the maximum resultant intensity will be considered. They reported an overall precision of 56.37% for simple sentences, 42.71% for compound sentences and 27.68% for complex sentences.

A powerful example of rule-based system is the Affect Analysis Model, which analyzes affect specifically in informal online communication media [43]. The proposed algorithm has five main steps, namely; symbolic cue analysis, syntactic structure analysis, word-level, phrase-level and sentence-level analysis. The database they use includes more than 1600 words, 364 emoticons, abbreviations, interjections and modifiers to cope with abbreviated language and to capture a wide range of affective features. The system is able to process complex-compound sentences and shows promising results with averaged accuracy of 72% for classification of text into nine emotion categories. However, similar to statistical models, one weak point of the model is that it requires a rich lexicon for accurate results and the system is not able to detect and process misspelled words in text.

Another popular approach is **Statistical NLP** and **learning-based** recognition models, which basically rely on automatic assessments of frequencies of some seed words and their co-occurrences, punctuation, abbreviations and sometimes together with synonym and acronym information. Brooks et al. [44] presented an automated affect classification system in chat logs exploiting NLP and ML techniques. The system segments the chat data and makes use of an improved bag-of-words model including non-verbal cues to classify text into 13 affect categories. The basic drawback of these machine learning approaches is that they usually lack linguistic analysis, but mainly rely on statistical and syntactical features.

The **commonsense-based approach** was first proposed by Liu et al. for emotion classification [45]. They used three real-world commonsense databases. The first one is the Open Mind Common Sense (OMCS) [46], which includes 400,000 facts about everyday world, the second one is Cyc [47] with over 3 million assertions about the world and the last one is ThoughtTreasure [48] containing approximately 100,000 concepts and relations. Compared to other approaches such as keyword spotting and statistical techniques, this model works robustly on the sentence level; a bunch of emotion models corresponding to each emotion class compete with each other and the winning models identify the affect label of the segments. They also implemented an e-mail browser called EmpathyBuddy, which expresses Ekman's six basic emotions based on the affective state of the text being typed by user.

Another approach that is adapted to online chat conversations is transformation-based learning(TBL) [49] for the task of assigning tags (e.g. yes-no question, emphasis, emotion, reject, clarify) to postings. With this aim, Wu et al. describe seven model templates used for posting act tagging and they use an iterative error-driven contextual approach to assign the corresponding tag. They also extend previous in part-of-speech tagging methods and dialogue act tagging by adding regular expressions into their templates which improves the test accuracy.

Other than described methods, Danisman and Alpkocak [50] proposed a VSM approach for emotion recognition from text and measured the effect of stemming and emotional intensity on emotion classification in text. They found that Vector Space model based classification on short sentences performs as good as Naive Bayes and SVM classifiers and ConceptNet which is a commonsense knowledge base with a natural language processing toolkit.

In addition to emotion and affect recognition methods, several applications and intelligent user interfaces have been designed and implemented with rich integrated emotion conveying engines. Sánchez et al. [51] proposed an Instant messaging system which is called Russkman. The system has been enhanced with functionality that allows users to convey moods and emotions while interacting with other users. iFeel_IM! is another system that enables users to express emotions during online communication [52]. iFeel_IM! considerably enhances emotional experience of real-time messaging with haptic devices, such as HaptiHeart, HaptiHug, HaptiTickler, HaptiCooler, and HaptiWarmer. EmoHeart is another application which is developed based on Affect Analysis Model to convey emotions in 3D virtual world Second Life [53]. CrystalChat [54] visualizes the user social network by extracting the user's chat log history and with the help of a graphical interface, then presents the patterns about the conversation length and emotional tone based on the emoticons used. *EmotionChat* is designed as a chat platform between teachers and students in e-learners systems [55].

There are several studies that sense affect by utilizing non-textual features. For example, Wang [56] presented a chat system that expresses the affective state of a user by using physiological data and animated text corresponding to given emotional information. The system uses a physiological sensor attached to the user's body to capture the affective state real time, detects the emotion without user intervention and shows it to the chat partner. Fragopanagos and Taylor [57] have introduced ERMIS (emotionally rich man-machine intelligent system) which aims at recognizing emotions based on people's speech from aspects of prosodic and lexical content. They developed a neural network architecture and it's simulation for emotion recognition task from speech and face stimuli.

Up to now, we have discussed the current methods along with their different features and different granularity levels, but in the end it is difficult to rank or compare the performances of each approach. Depending on the structure and context, all methods have their strengths and weaknesses. Therefore, usage of one or combination of these methods requires understanding of all their respective challenges and potential advantages. Table 2.1 illustrates a summary of emotion studies in the chat domain and compares the methodologies, used features, tools and databases.

2.4. Lexical Databases for Affect Analysis

One of the early works for creating an affective resource is Whissell's Dictionary of Affect Language [61] which contains 8,742 words in English that are given in three emotional dimensions: evaluation, activation and imagery. Continuous affective scores are measured by human judges from 1 (unpleasant) to 3 (pleasant).

The Affective Norms for English Words (ANEW) [62] is another important resource which includes a set of normative emotional ratings for a large number English

Table 2.1. Emotion Studies in Chat Domain

Focus of the Study	Tools	Emotion Model	Emotion Categories	Databases
A methodology of emotion extraction from real time chat messenger [42]	POS tagger web API	Keyword spotting with rule based model	anger, joy, disgust, fear, guilt, interest, sad, shame, surprise 200 emoticons	SemEval Task 14 'Affective Text' test set
An approach to emotion estimation from chat and to employ animated agents [41]	Embodied Conversational Messenger	Keyword spotting with vector model	happiness, sadness, anger, fear, surprise, disgust	WordNet-Affect OMCS (Open Mind Common Sense)
Emotion extraction engine for real time Internet text communication [40]	Expressive real time Communication Interfaces	Syntactical + lexical analysis	happiness, sadness, anger, fear, surprise, disgust	A database designed containing 16400 words
A chatroom with emotion regulation between teachers and students in e-learning [55]	EmotionChat	Rule based model	love, joy, anger, frustration, neutral	Casebase (manually marked) user feedback store data
Recognition and interpretation of affect in text messaging [58]	-	Rule based model	anger, joy, disgust, fear, guilt, interest, sad, shame, surprise	Affect database with 1627 words, 364 emoticons, 337 acronyms and abbreviations
A novel way of textual affect sensing using real world knowledge [45]	EmpathyBuddy	Commonsense-based	happiness, fear, anger, sadness, surprise, disgust	OMCS, Cyc, ThoughtTreasure
An emotion recognition system using facial features, lexicon in speech [57]	ERMIS, ASSESS	Neural network model	happiness, fear, anger, sadness, surprise, disgust	HUMAINE databases
Classification of emotions in network-based chat conversations [59]	Microsoft Speech SDK, WEKA	Machine learning model	neutral, angry, sad, afraid, disgusted, ironic, happy, and surprise	A dataset is constructed from a set of conversations
To build automated classifiers of affect in chat logs [44]	WEKA	Statistical NLP and Machine learning model	interest, serenity, agreement, considering, confusion, acceptance, annoyance, amusement, supportive, surprise, frustration, apprehension, anticipation	485,045 chat messages from Nearby Supernova Factory
To construct a taxonomy of affect in collaborative online chat [60]	not specified	Adaptation of grounded theory	40 affect expressions based on intensity	Nearby Supernova Factory chat logs

words which is the commonly used set of 1,034 words. The purpose is to develop a set of verbal materials that have been rated in terms of pleasure, arousal, and dominance to support the emotion studies. Similarly, WordNet-Affect is a well known linguistic resource for extracting emotions from text [63]. The starting point of WordNet-Affect is to build an hierarchy of affective domain labels by labeling synsets (a set of one or more synonyms) that express affective concepts based on WordNet Domains [64].

Linguistic Inquiry and Word Count (LIWC) is one of the powerful text analysis software products with a comprehensive affective dictionary that analyzes text based on grammatical, psychological, and content word categorization. This dictionary allows to measure 74 different linguistic dimensions with more than 2,200 words and word stems. Affect sensing methods that based on LIWC, calculate the word counts of the input text depending on these linguistic dimensions [65–67].

SentiWordNet currently represents a good resource for opinion mining, however, it is comparatively noisy and limited. It basically contains polarity information for common sense knowledge; each WordNet synset has a polarity (positive/negative) and objectivity score ranged from 0.0 to 1.0. These scores obtained from a committee of eight ternary classifiers.

Another publicly available semantic resource is SenticNet which is a comprehensive database for opinion mining. SenticNet is created using common sense reasoning techniques, such as blending and spectral activation, combining with an emotion categorization model [68]. In opposition to SentiWordNet and WordNet-Affect, which provide only polarity and affective information, SenticNet 2 incorporates sentic computing to detect both cognitive and affective information by associating semantics and sentics to every common sense concept from the Open Mind corpus [69]. SentiSense is another affective lexicon that attaches emotional categories to WordNet synsets [70].

There have been some experiments enlarging databases in an attempt to build more comprehensive and enriched lexical resources by merging the existing ones. For example, Poria et. al [71] created a comprehensive publicly available semantic resource.

This database is created by automatically assigning an emotion to each concept in SenticNet which provides only polarity information for the concepts. With this purpose, they used WordNet Affect emotion list as a small dictionary and then merged the emotion labels and 2729 concepts from the SenticNet. Similarly, several other studies have been performed to develop or adapt WordNet Affect to other languages such as Bengali [72], Japanese [73], Russian and Romanian [74].

2.5. Related Studies in Turkish Language

Textual affect recognition has not been well studied in Turkish language. Majority of the related works in Turkish language have been based on only positive - neutral - negative sentiment classification. Sentiment analysis (also known as opinion mining) attempts to analyze emotions from textual information, and can be performed with three different granularity levels; document, sentence and aspect level. Conventional approaches on Sentiment Analysis (SA) generally rest on two main techniques; Machine Learning (ML) and lexicon based modeling. For example, Vural et al. proposed a lexicon based framework to classify the polarity of Turkish movie reviews. For this purpose, they translated the SentiStrength sentiment lexicon to Turkish and achieved 76% accuracy at word level and 75% accuracy at sentence level [75]. Dehkharghani et al. [76] presented SentiTurkNet, the first comprehensive Turkish polarity lexicon, which includes positivity, negativity, and objectivity scores assigned to each synset in the Turkish WordNet (about 15,000 synsets).

One of the early works in sentiment analysis was Erogul's thesis on Sentiment Analysis in Turkish [77]. The author achieved 85% accuracy using SVM classifier in their binary sentiment classification system on a database containing more than 10000 movie review documents. This study revealed that use of word roots as a feature is associated with a slightly better performance compared to use of entire word. Because Turkish is an agglutinative language, words can take many suffixes that modify the meaning. Thus, this feature of Turkish language makes the sentiment analysis even more challenging. Considering the same problem, Cakmak et al. analyzed emotions attributed to Turkish word roots and sentences. Their study included 197 Turkish emotion words and evaluated these words based on valence, activation and dominance. The study demonstrated that emotions attributed to sentences are significantly correlated with emotions attributed to word roots in the corresponding senteces. [78].

In another study, Boynukalın et al. analyzed emotion in Turkish text by using supervised machine learning methods [79,80]. In this thesis, Boynukalın used two different data resources. First one is ISEAR (International Survey on Emotion Antecedents and Reactions) that contains answers of student respondents to a questionnaire on seven major emotions; joy, fear, anger, sadness, disgust, shame, and guilt. The other data source is Turkish fairy tales labeled with four basic emotions; joy, sadness, fear and anger. The author reported the highest accuracies with each classifier as follow: 78.1% with NB, 81.3% with CNB and 75.9% with SVM.

Kaya et al. [81] used four ML algorithms; Naive Bayes, Maximum Entropy, SVM and the character based N-Gram Language Model for sentiment classification of Turkish political columns. Using different features, they reported accuracies between 65% and 77%. In their analysis the Maximum Entropy and N-Gram Language Model outperformed the SVM and Naive Bayes. Another ML approach was experimented by Akba et al. using Support Vector Machine (SVM) and Naive Bayes. They analyzed Turkish movie reviews and reported 83.9% accuracy in two category classification (positive and negative) and 63.3% accuracy with SVM in three category classification [82]. Some recent NLP tools have been examined by Yıldırım et al. on a manually normalized set of 13K tweets in Turkish, for positive and negative sentiment analysis [83]. Other than movie reviews and Twitter data, Başdemir studied emotion detection in Turkish novels using four seperate machine learning methods [84].

Ekşisözlük [85] is a very popular comment website in Turkey and similar to Reddit, it covers a very broad spectrum of topics of discussion. İşgüder Şahin et al. studied 'ekşisözlük' comments for polarity classification with Naive Bayes (NB), Support Vector Machines (SVM) and K nearest neighbor (KNN) classifiers. The authors specifically retrieved 'ekşisözlük' comments about 12 different technology companies and on their test set, they reported the best F-measure as 0,69 [86].

Owing to the insufficiencies of training data for learning systems and ground truth data for model evaluations, many researchers have tried to explore new useful features and training methodologies. For example, Gezici and colleagues examined different aspects of sentences, such as length, purity, irrealis content, subjectivity, and position within the opinionated text [87]. To evaluate the effect of these newly proposed features on polarity classification, they used TripAdvisor data and found small improvement in classification accuracy. In another work, to achieve Turkish sentiment analysis with fewer training data, Çetin and Amasyalı [88] studied active learning scheme by using different clustering algorithms. Their study revealed an accuracy value of 64% with full training set. A similar accuracy value was obtained with half of the data set, When an active learning algorithm was incorporated. In a cross-lingual work on sentiment analysis, Demirtaş and Pechenizkiy examined the effect of expanding training size in polarity detection task [89]. With this purpose, they used movie review and product review data sets in both English and Turkish and investigated the prospects of machine translation in polarity detection. They observed no accuracy improvement with cross-language co-training. However, they reported performance improvement with semi-supervised learning using unlabeled data coming from the same domain. For word polarity detection, Özsert and Özgür proposed a multilingual approach (English and Turkish) based on random walk model [90]. They created a word relatedness graph by using the relations in WordNet and extended it with Inter-Lingual-Index based on English WordNet. WordNet for Turkish [91] was constructed as a part of the BalkaNet Project [92]. This multilingual study demonstrated improvement in the performance for both English and Turkish languages.

None of the above mentioned works have examined fine-grained dimensional affect in Turkish text. In this respect, the model we propose is the first framework that aims to accurately capture affect in sentence level on Turkish texts.

3. DATA AND ANNOTATION SCHEME

The chapter is divided into following sections: (1) Features of textual data and resources used, (2) data annotation scheme design to construct a ground truth for evaluation, and (3) inter-annotator agreement results to assess the reliability of subjective annotation.

3.1. Data Acquisition

For our first application, we use a large database of multiparty chat records, collected from thousands of players of the 'Okey' game. Okey is a very popular board game in Turkey, played by four people, and having a strong social component. We analyzed a comprehensive chat database derived from in-game chat logs of an online version of Okey with more than 100,000 users [93].

Secondly, we use psychotherapy records for affect analysis. For textual analysis of psychotherapy records, we use transcriptions of audio recordings. All sessions are anonymous conversations between a psychotherapist and a patient. The data used for this study came from the Istanbul Bilgi University Psychotherapy Research Laboratory [94], established to study the psychotherapy processes conducted at Istanbul Bilgi University Psychological Center. In their study, a total of 22 therapists (all of the master level graduate students in clinical psychology) evaluated a total of 50 child patients using a psycho-dynamic play therapy. Each therapist worked with either one or two patients. The average number of sessions for a child was 27 with a maximum of 40 sessions. They used the Child Behavior Checklist (CBCL) [95] method to identify problematic behaviors in children. Several factors were assessed in play activity. But we specifically used the affective component which investigates some features (type, range, and regulation) of emotions brought by the child to the play. The affective component on affect modulation has a rating scale from 1 (low) to 5 (high). In addition, plays are coded with a scale from 1 (low) to 5 (high) under eight different affect categories: 'Pleasure', 'boredom', 'fear', 'anger', 'anxiety', 'sadness', 'shame' and 'guilt'. In reliability testing, they found a good agreement on the ‘Segmentation of the child’s activity’ (Kappa = 0.72) using three independent raters and eight play sessions. For the dimensional analysis, they observed the inter-rater reliability ranging from acceptable to excellent (ICC = 0.52 - 0.89) with the measure Interclass Correlation Coefficient for ordinal variables.

For the third application, we use a set of Turkish movie reviews from the work of Dehkharghani et al. [96]. In their work, movie reviews were randomly collected from www.beyazperde.com. The list contains 1042 reviews and 2700 sentences which have been labeled in three categories: positive, negative, and neutral. This full list is available online [97].

Fourth application is the opinion mining of teachers’ comments on their students. For this task, we used 300 different labeled comments. Some of the comments included more than one sentence but during the analysis we processed each unique comment as a sentence.

Lastly, we used approximately 2,500 tweets labeled in positive, negative, and neutral classes for sentiment level analysis. We used 1,900 tweets during the training of feature reduction and only 600 of them for evaluation. Similar to teachers’ comments, we treated the tweets with more than one sentence as a single input sentence when we classified them.

3.2. Annotation Scheme

We designed two different annotation systems for different tasks. First annotation was designed to create a ground truth to evaluate our affect analysis model’s performance in chat records. In order to construct a subset of chat dialogues for manual annotation, we randomly selected about 1000 sentences among 4 million sentences in the database. Secondly, we selected the most expressive sentences by eliminating sentence fragments and meaningless utterances. At the end, we gathered about 300 independent emotive and non-emotive sentences. Because of random selection, the final set is not necessarily a balanced distribution of emotive and non-emotive sentences. Then, we created nine pairs of surveys (each survey included 100 sentences) to evaluate for valence, arousal, and dominance dimensions.

In contrast to data annotation on a paper, online survey is easier to distribute and process. Therefore, all of our annotations were performed online by using Google forms. All surveys were completed anonymously, by native Turkish speaker annotators from a wide variety of age groups. Although literacy level makes no notable difference in affective judgement [98], we chose participant groups from different backgrounds. They varied from high school students to university professors.

It is well known that, detailed orientation to a questionnaire makes a significant impact on a participant's ability to achieve consistent and reliable labeling. For this reason, we provided a single page of instructions and sample questions specific to each dimension tested (See Appendix for Figure A.1, Figure A.3, Figure A.5). To clarify the emotional dimensions and easily assess affective stimulus, we presented several annotated sentences as a reference and incorporated the Self-Assessment Manikins (SAM) when designing the annotation scheme. The SAM is a well-validated pictorial scaling method developed by Bradley & Lang [99]. This method directly measures emotions such as pleasure, arousal, and dominance based on of a person's affective response. Manikins were presented at the top of the sentences. We also presented a set of tagged words drawn from the Affective Norms for English Words (ANEW) corpus [62]. As illustrated in Figures 3.1, 3.2, 3.3, each reference word was scaled on Manikin images (See Appendix for full scheme of annotation).

All participants were instructed to evaluate each sentence on a 5-point Likert scale. According to this scale, a value of 5 valence indicates extremely happy, satisfied, hopeful or pleasant mood and 1 indicates completely unhappy, dissatisfied or bored mood. For arousal, the annotation ranged from calm, inactive, and dull at the low end of the scale to highly aroused, excited and active at the high end. Similarly, for dominance, the highest value was given if the subject felt powerful, dominant, influential or controlling, and the lowest value if they felt controlled, unimportant,

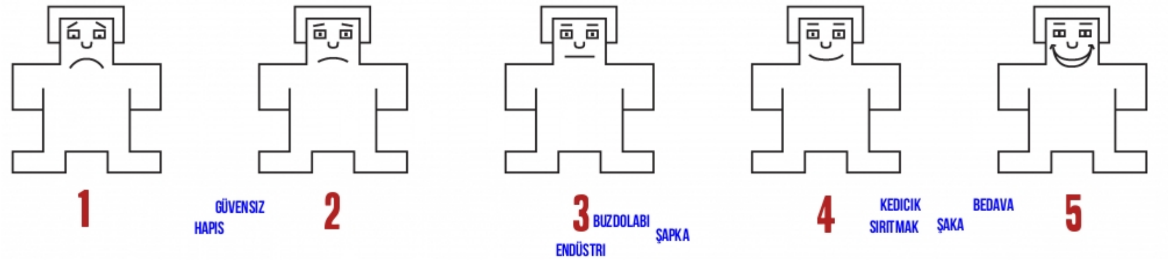


Figure 3.1. SAM images tagged with reference words for valence dimension

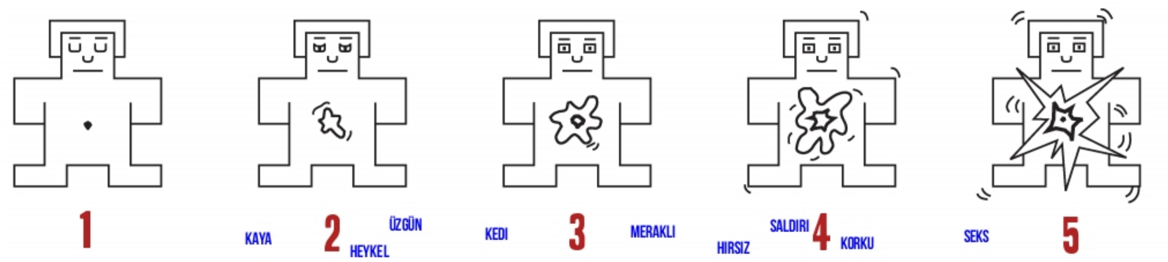


Figure 3.2. SAM images tagged with reference words for arousal dimension

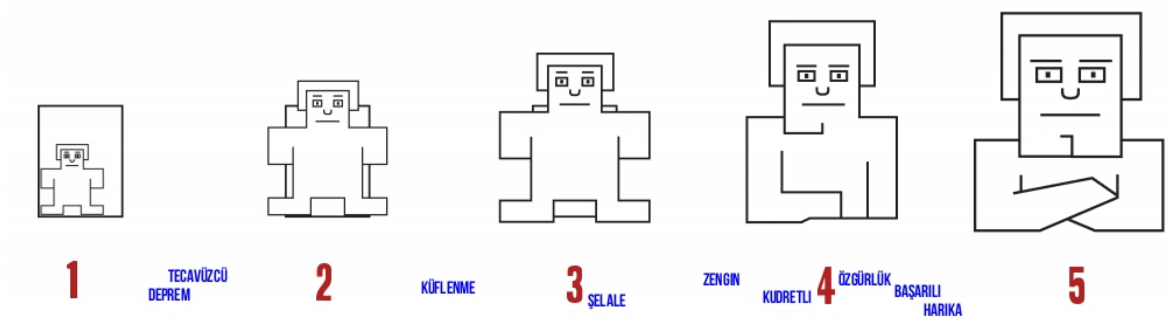


Figure 3.3. SAM images tagged with reference words for dominance dimension

weak, or influenced. For all dimensions, 3 was selected if they felt neutral. Each annotator was given a set of about 100 sentences and asked to complete the survey for one dimension. On average each sentence was annotated by 7 independent annotators for each domain. Individual responses were kept confidential and anonymous and at the end were reported in aggregate. Although there was no time limitation, annotators were notified to use their first intuition about a sentence. It took approximately 15 minutes for an annotator to complete one survey.

Second annotation procedure was designed to perform quantitative measurements for effect of modifiers (intensifiers & diminishers) and interjections. We attempt to assess the effect of each modifier in the context of a sentence, instead of annotating modifiers at the word level. This forms a second set of annotations, which we refer as 'modifier annotations.' Each modifier was investigated both in positive context and negative context considering that they may have different impact depending on the emotional polarity of sentence. We created two sets of 73 sentences. These two sets included the same wording with the only difference between corresponding sentences was the presence or absence of a modifier. For example, both 'ürünler fazlasıyla basit ve kalitesiz' and 'ürünler basit ve kalitesiz' sentences were annotated by independent annotators. In the end, mean values of subjective annotation was calculated for each sentences. Then the difference between the scores of those two sentences gave us the score for adverbial expression 'fazlasıyla' in negative context. Similarly, the same modifier was tested in a positive context which is illustrated in Table 3.1. (See Appendix 4.2. On average each modifier was annotated by 15 annotators for each domain and context.

Table 3.1. Sample sentence for modifier annotation and average VAD scores

Sentence	**Valence**	**Arousal**	**Dominance**
aslında kolay bir iş	3.40	2.75	4.13
aslında **fazlasıyla** kolay bir iş	4	2.5	4.28

Additionally, we tested a number of modified sentences to assess the effect of typing modifications on affective stimulus. For example, we compared the annotation VAD scores of ‘sağol Ali’ and ‘sağol Aliiii’ expressions by including both sentences in shuffled order to obtain the numerical effect of duplications. Similarly, we included both ‘O SEVIYOR’ and ‘o seviyor’ expressions to see the effect of capitalization; ‘sus artık’ and ‘sus artık!’ to see the effect of exclamation mark. Again, these features were investigated both in positive and negative contexts. These features are illustrated in detail (See Section 4.5.6).

In total, we collected 6,300 ratings for sentence annotations and 7,575 ratings for modifier annotations. 116 participants completed annotations for at least one dimension. The ground truth annotations for both sets were obtained by averaging the annotations for each sentence.

3.3. Inter-annotator Reliability

Computerized evaluation of psychological variables is challenging. This holds true for measurement of affective load of a sentence directly and precisely. One major challenge in affective data annotation is highly subjective nature of ground truth labeling. In addition to subjectivity, annotation is an error-prone process for several other reasons. Affective meaning can be ambiguous, or annotators might label the instances in a rapid manner. These factors can increase the number of noisy labels. Hence, it is crucial to measure the reliability of our approximate and subjective annotation in order to construct a ground truth.

Affective labeling is more reliable when multiple judges annotate the corpus. Therefore, at the first part of annotation, each sentence was labeled by seven different annotators with various educational and socioeconomic backgrounds in order to capture a broad consensus on affective judgement. Then, we examined the inter-annotator agreement among all annotators.

For the annotation of modifiers, we repeated the labeling with each participant for 15 times on average. Repeated labeling is especially useful for identifying noisy labels [100]. For both set of annotations, the values that were more than one standard deviation different from the mean were treated as outliers and eliminated. Means were recomputed afterwards.

Measures such as Fleiss' kappa and Cohen's kappa are considered convenient for nominal values, if the interval difference between all cases has the same meaning on reliability. However, we deal with an annotation scaling where the disagreement between valence score 2 and 4 is worse than the disagreement between 2 and 3. Therefore, we report the inter-annotator reliability for sentence and modifier annotations with Krippendorff's Alpha [101]. Besides, Krippendorff's Alpha is very suitable when there are two or more coders. Using the notations of [102], we illustrate the calculation of Krippendorff's Alpha as follows:

$$\alpha = 1 - \frac{D_o}{D_e} \tag{3.1}$$

Agreement coefficient α is obtained by subtracting the measured disagreements as in Equation 3.1. D_o^α is the observed disagreement of α where k is a unique category for the scores given by coders, n_{ik} is the number of coders that make judgement k. n_{ik_a}, n_{ik_b} are pairs of judgements of item i and for each pair of distinct values k_a, k_b and they are multiplied by the $d_{k_a k_b}$ which is the distance between them. Then the summation is divided by the number of ordered judgement pairs which is c(c − 1) as shown in Equation 3.2.

$$D_o^\alpha = \frac{1}{ic(c-1)} \sum_{i \in I} \sum_{j=1}^{k} \sum_{l=1}^{k} n_{ik_j} n_{ik_l} d_{k_j k_l} \tag{3.2}$$

Similarly, D_e^α is the expected disagreement of α between the judgement pairs and it is calculated as in Equation 3.3.

$$D_e^\alpha = \frac{1}{ic(ic-1)} \sum_{j=1}^{k} \sum_{l=1}^{k} n_{k_j} n_{k_l} d_{k_j k_l} \tag{3.3}$$

Krippendorff's alpha is applicable with different distance metrics such as ordinal, interval, ratio, etc. Because sentences are rated on a continuous scale in our annotation, we chose to use Krippendorff's interval coefficient for reliability test, however, the ordinal coefficient produces very similar results. α is given in a [0, 1] interval, where 0 indicates complete disagreement, and 1 indicates a perfect agreement. Apart from Krippendorff's alpha, we also calculated the average pairwise agreement between coders.

Table 3.2. Inter-annotator agreement for ground truth annotation

Measure	Valence	Arousal	Dominance
Krippendorff ($\alpha_{interval}$)	0.8	0.62	0.66
Average pairwise agreement	63.7%	55.3%	59.78%

Table 3.3. Inter-annotator agreement for modifier annotation

Measure	Valence	Arousal	Dominance
Krippendorff ($\alpha_{interval}$)	0.81	0.49	0.68
Average pairwise agreement	53.19%	41.40%	41.92%

There exist several different remarks on the substantial agreement interval and the poor level of agreement, when it comes to interpreting the α scores. In general, values greater than 0.7 are considered as good agreement and values more than 0.8 represent excellent agreement [103]. From another approach, Landis and Koch interpreted kappa interval [0.4 - 0.6] as moderate and [0.6 - 0.8] as substantial agreement [104]. Considering these interpretations, we affirm that the inter-annotator agreement in our annotation is substantial for valence dimension, but not as high for arousal and dominance. One potential explanation for this difference is as follow. Valence is generally more obvious and easier to annotate, whereas arousal and dominance are more abstract, subjective and therefore difficult to annotate.

4. AFFECT ANALYSIS MODEL

This chapter provides the tools and techniques utilized to perform affect recognition from textual data. It depicts how affective lexical choices are assessed, how dirty and unstructured input is preprocessed and what rules and features are employed with the affect investigation process.

4.1. Affective Lexicon

For Turkish language, there exist no comprehensive and widely-used lexicon of affective words with VAD annotation. Since it is very costly and time consuming to construct a dictionary from scratch, we have automatically translated a dictionary of English lemmas gathered by the study of Warriner et al. [105]. This study, which was based on the ANEW norms proposed by Bradley and Lang for 1,034 words [62], rated 13,915 English lemmas in a nine point scale (1-9). A total of 1,827 participants (through Mechanical Turk) contributed to their study. They provided mean values and standard deviations for valence, arousal, and dominance scores. They reported the variability in scores according to gender, age, and education level of the participants.

We linearly transformed the affect scores to a five point scale [1-5]. For translation, we initially used the Google Translation API. Two human translators manually checked each word independently, and corrected missing words and mis-translations. Some culture and language dependent words were omitted and the affective lexicon was finally expanded with synsets from TDK (Turkish Language Organization) dictionary [106]. As a result, we have developed a comprehensive affective lexicon for Turkish that includes valence, arousal and dominance scores for 15,222 different words and phrases (See Table 4.1 for sample words from dictionary). Although this resource has some limitations, our assessment shows that it is useful. Any future work on a proper Turkish affective lexicon would improve the system that we propose.

Table 4.1. Sample words and phrases from affective dictionary

Turkish	**English**	**POS**	**Valence**	**Arousal**	**Dominance**
açıkhava	outdoor	ADJ	4.17	2.28	3.1
açığa kavuşturmak	clarify	VB	3.5	1.93	3.52
adaletsizlik	injustice	NN	1.73	3.73	2.14
adam kaçırma	kidnapping	NN	1.53	3.18	1.74
mutlu	happy	ADJ	4.74	3.53	4.11
tatil	vacation	NN	4.77	3.11	4.06
yetenekli	talented	ADJ	4.48	2.78	3.57
yumuşak başlı	docile	ADJ	3.38	1.74	3.23
polis	cop	NN	2.75	2.95	1.92
akordiyon	accordion	NN	3.13	1.97	3.11
bebek bezi	nappy	NN	2.05	2.16	2.62
bebek karyolası	cot	NN	3.19	1.98	3.04
donuk	dull	ADJ	2.2	1.34	2.86
yatıştırıcı	soothing	ADJ	4.03	1.46	3.38
zelzele	earthquake	NN	2.03	3.88	1.57
lenfoma	lymphoma	NN	1.8	2.8	1.69
ümit	hope	NN	4.24	3.15	3.89
ağlamak	cry	VB	2.11	3.23	1.78
yusufçuk	dragonfly	NN	3.73	2.43	3.21
centilmence	gentlemanly	ADV	3.89	2.46	4

4.2. Preprocessing

For the preprocessing of 'okey' game chat records, we parsed JSON logs with UTF-8 format since we were dealing with Turkish text, thus Turkish characters. After we make required file conversions, we performed a three-step preprocessing to tokenize and clean the informal characteristics of the chat messages. Before normalizing the informal chat texts, we split the texts into sentences. Instead of using sentence terminators such as period, exclamation mark, quotation marks, question mark (.!"?), we evaluated entries line by line. For example, if a user writes his or her message and hits 'send', we accept the entire message as an input sentence no matter how long it is. We find this tokenization is more appropriate than using sentence boundary marks, because it is very likely that users press enter to send their message to the chat window when they complete their point. This method also help us avoid incorrect tokenizations when period is used in emoticon, date or abbreviation. Besides, we do not deal with a well structured text. To put another way, we observed that none of the players pays attention to spelling rules and the text is poorly punctuated. Similarly, we do not use word capitalization as a rule for tokenization. We segmented input sentences into words and each token was kept with part-of-speech (POS) information if it was present in affective dictionary. This was performed because we were only interested in POS tag of emotive words when calculating the overall affective score of a sentence.

We then retrieved the most frequently used 14000 words in the 'okey' chat records. For this statistics, we used gigantic chat logs consisting of 80 MB of text, 4.7 billion sentences and 11 billion words). Then, 2 human annotators manually checked and corrected the list consisting of the most frequently used words (14000 word). For the first step of normalization of a sentence, we checked the list of corrected words via direct look-up. Because of the presence of a lot of repetitive expressions and words in the chat domain, spelling mistakes are usually repeated as well. Therefore, keeping a list of the most frequent spelling mistakes is a very effective way to correct chat-specific spelling mistakes. Before correcting the misspelled input text, we recorded the intentional spelling mistakes such as duplications (e.g. 'selaaaam'), upper-case usage (e.g. 'HA-DII') or exclamation mark usage (e.g. '!!!'), since they serve as useful features, especially for high arousal patterns.

Lastly, we used a Turkish normalization tool proposed by Torunoglu and Eryigit [107]. In their work, normalization problem was examined under 7 seperate categories; letter case transformation, replacement rules & lexicon lookup, proper noun detection, deasciification, vowel restoration, accent normalization and spelling correction. They reported 40 percent improvement over a lexicon lookup baseline and nearly 50 percent improvement over available spelling correctors. This tool is available online [108]

4.3. Affect Prediction

Typically at the sentence level, affect analysis starts with calculating the affective values of small units in the sentence such as words and phrases. Then, the overall score is computed by summing the scores of these units. Similarly, we start by tokenizing the sentence into trigrams, bigrams and unigrams. In bag of words (BOW) model, a sentence is represented as an unordered combination of words, disregarding the word order. However, we have two-word and three-word phrases in our affective dictionary too, for example, ‘acı çektirmek’, ‘beş parasız’, ‘çocukluk çağı and ‘son dakika haberi’, ‘sokağa çıkma yasağı’, ‘yasadışı örgüt üyesi’. Therefore, we perform dictionary lookup first for trigrams, then for bigrams and lastly for unigrams with an N-gram approach. After we obtain word and phrase level scores as shown in Eq. 4.1, we calculate the sentence level VAD scores where AffDic stands for affective dictionary, ω_{affect} is the affective representation of a unit (word or phrase), $\omega_{valence}$ is valence score of unit, similarly $\omega_{arousal}$ is arousal score of unit, and $\omega_{dominance}$ is dominance score.

$$\omega_{affect} = AffDic(\omega) = (\omega_{valence}, \omega_{arousal}, \omega_{dominance}) \tag{4.1}$$

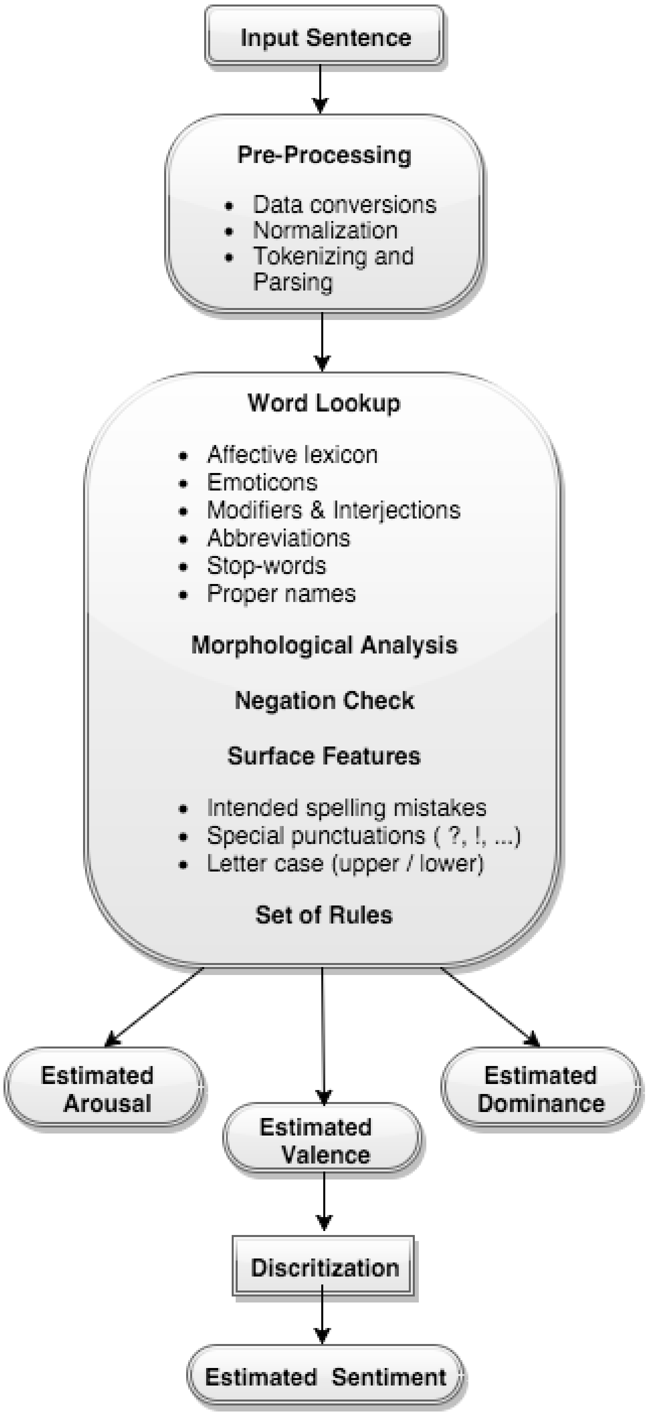


Figure 4.1. The Model of Affect Analysis

Sentence level VAD scores are calculated with two different techniques. The first technique involves calculation of overall value by averaging the word and phrase level affect scores as shown in Eq. 4.2 where S_{affect} stands for the overall VAD score of the sentence. Note that, only words with affective load are considered in the summation; stop and neutral words are excluded. This is because when we evaluate longer sentences consisting of neutral words mostly and a few emotive words, taking average almost neutralizes the affective score of the sentence. Therefore, we only take the emotive words with a score other 3 into consideration.

$$S_{affect} = \frac{\sum_{i=1}^{n} \omega_{affect_i}}{n} \tag{4.2}$$

The second technique is called *minmax* and it involves accepting the score of the most emotive unit as the overall score of a sentence. Thus, the absolute value of three minus each unit's valence score is calculated and the unit with the highest abdolute difference from 3 determines the overall VAD score of an entire sentence. In other words, the unit with the highest valence or the lowest valence gives us the overall value:

$$S_{affect} = max\,|3 - (\omega_{valence})_i| \tag{4.3}$$

4.4. Sentiment Prediction

Despite the fact that continuous affective scores are much more informative than polarity level emotion detection, we did not have affective ground truth to evaluate movie reviews and Twitter data. Since we aimed to test our affect model's performance in these domains too, we performed polarity detection by transforming continuous affective scores to three sentiment polarities. We only took the valence dimension into consideration, When assessing sentiment polarity (positive negative and neutral). Initially, we started with calculating the smallest units by using dimensional affective information as described in Section 4.3. After we obtained the continuous values (from 1 to 5) for each sentence, we mapped these scores to discrete levels as shown in Eq. 4.4 where $S_{polarity}$ denotes overall polarity of the sentence.

$$S_{polarity} = \begin{Bmatrix} -1, & S_{affect} < 3 \\ 0, & S_{affect} = 3 \\ 1, & S_{affect} > 3 \end{Bmatrix} \tag{4.4}$$

4.5. Set of Features and Rules

4.5.1. Emoticons

Verbal language is the main component of a conversation. However, most of the time, complex representations beyond verbal language is needed to emphasize and communicate emotions. Online chats lack face-to-face communication. Therefore, visual elements such as emoticons are utilized to communicate gestures and emotional facial displays during online chat. In addition to enriching and underlining the meaning of messages, **emoticons** can represent personality characteristics of individuals.

In the simplest form, emoticon exchange generally involves ASCII symbol combinations within the context of computer-mediated communication to display affective mood. Some examples for these emoticons are as follow; ':o' as 'surprised', ':)' as happy, ':(' as 'sad', ';)' as winking. Emoticons reinforce the emotional communication in a powerful way. For instance, emoticons play key role in communication of sarcastic expressions. Some sample sentences for sarcastic expressions are given below.

(i) Aman ne güzel bir hava :(

(ii) Nasıl üzüldüm anlatamam :)

Considering the emoticons as dominant feature enables our model to identify these kind of sarcastic sentences to some extent. Therefore, we accepted emoticons to be the most informative qualities in our model when they are present.

In order to detect all the emoticons in our text, we constructed an emoticon table consisting of 120 popular emoticons used in Turkish multi-party chats (see Figure C.1 for the full list of emoticons.)

4.5.2. Modifiers and Interjections

Modifiers and interjections make up an important set of features. We collected a list of Turkish words (adverbials, adjectivals, and nominals) that can intensify or diminish the affective attribute of a sentence. We enriched this set with the synonyms of these words. In section 3.2, we explained in detail how these modifier scores are calculated with the help of human judges. The final modifier list includes 71 intensifiers and diminishers. We also created a list of 50 interjections. Interjections such as 'ah', 'of', 'eyvah' are important emotive qualities especially as an indicator of high arousal. Just like the modifiers, interjection scores are calculated by the same annotation scheme. Some example modifiers and interjections given in Table 4.2 with their corresponding VAD scores that are obtained by modifier annotation.

Current approaches for contextual valence shifters mostly rely on adding or subtracting approximate integers, as proposed by Polanyi and Zaenen [109]. In a similar vein, when there is a modifier, our model first determines the polarity of a sentence as positive or negative. Then the model updates the affective scores by adding the modifier score to the overall affective score of the sentence based on the polarity assignment.

Table 4.2. Example modifiers, interjections and corresponding scores in positive and negative contexts

		Valence		**Arousal**		**Dominance**	
Turkish	**English**	**POS context**	**NEG context**	**POS context**	**NEG context**	**POS context**	**NEG context**
fazlasıyla	greatly	+0.6	-0.5	-0.25	+0.17	+0.15	+0.67
inanılmaz	incredible	+0.91	-0.83	+0.5.	1.71	-0.19	+0.43
özellikle	specially	+1.04	-0.69	+1.25	-0.42	+0.04	+0.72
hafifçe	slightly	-0.32	-0.17	-0.3	-0.07	-0.84	+0.25
sadece	only	+0.9	-1.17	0.17	+1.4	-0.36	+0.09
of	ah	-0.02	-1.59	+0.27	-0.17	-0.36	-0.07
hey!	hey!	+0.42	-0.10	+1.13	+1.03	+0.29	+0.45

4.5.3. Morphological Analysis

Since Turkish is a morphologically rich agglutinative language, one can generate hundreds of legitimate words from a single root with derivational and inflectional morphemes. For example, when we look at the vocabulary size, with 1 million word corpus, Turkish has 106,547 distinct words while English has 33,398 distinct words, with 10 million word corpus, Turkish has 417,775 distinct words and English has 97,734 [110]. Therefore, Word-level analysis is a challenging task in Turkish language [111]. It has been shown that there is a strong link between word roots and the perceived emotion of the sentence in Turkish [78]. Therefore, we took **word roots** into consideration when we created our lexicon and analyzed the affect in sentences. We did not perform a detailed morphological analysis for different person forms and tenses, because we assumed that the affective load of a sentence is almost the same for different person forms and tenses in Turkish.

Compared to NLP tools in English, current Turkish POS taggers are less advanced mainly because of high potential for morphological ambiguity. In order to avoid these ambiguities in our model, two linguists manually tagged each word in the affective dictionary according to their part-of-speech. Since we only took affect words into consideration, POS tagging was not employed for the input words that are not found in affective dictionary. However since the POS tagging of dictionary elements was performed context free, word sense ambiguities could not be completely prevented.

Table 4.3. Morphophonemic operations with sample inflected word forms

Word	**Marked Word**	**Morphophonemic Change**	**Inflected Form**
romantik	romantikˆ	k → ğ	romantiğim
ızdırap	ızdırapˆ	p → b	ızdırabını
hissetmek	hissetˆmek	t → d	hissediyorum
katalog	katalogˆ	g → ğ	kataloğun
kıskanç	kıskançˆ	ç → c	kıskancım
istemek	iste>mek	e → i	istiyorum
kötülemek	kötüleˆmek	e → ü	kötülüyorlar
kısıtlamak	kısıtla>mak	a → ı	kısıtlıyor
kutlamak	kutlaˆmak	a → u	kutluyoruz
sabır	sab<ır	ı drop	sabrımı
kibir	kib<ir	i drop	kibrin
özür	öz<ür	ü drop	özrün
oğul	oğ<ul	u drop	oğlum

Morphological operations we processed included morphophonological changes such as vowel harmony, vowel drops and consonant changes [112], removing the infinitive suffix '-mek', '-mak', and most importantly, detection of negation that comes with suffixes besides the negation words. In order to capture the morphophonological changes, we marked the words of affective dictionary manually for all possible vowel and consonant changes since there is no general rule for specifying such words. In total, more than 2,500 words of affective dictionary are marked for morphophonemic alternation. It should be noted that monomorphonemic alternations were only searched for word roots. Because we only took word roots into consideration, the way suffixes are modified when they are attached to stem has no bearing on our affect prediction model. Morphophonemic changes that we process are shown in the Table 4.3 with sample inflected word forms.

4.5.4. Handling Negation

In Turkish, there are different negating markers such as 'değil' (not) and 'yok' (non-existent), or negative connective 'ne ... ne ... ' (neither ... nor ...). But majority of the time negation is formed morphologically with the negative marker -mA (can be -mı -mi -mu -mü based on the last vowel of the stem). Primarily, -mA suffix is used for negating verbal sentences, however, it can be be either on the main clause verb or on the verb of the subordinate clause [113]. For instance:

(i) Gerçeği asla bil-**me**-yeceksin. (You will never know the truth.)
(ii) Size toplantıya katıl-**ma**-yacağımı söyledim. (I told you that I will not attend the meeting.)

One problem is that apart from being a marker of negation, '-mA' suffix can also function on non-finite verbs as a verbal noun maker, for example 'anla**ma**sı zor bir kitap', 'sev**me**yi öğrenmeli'. Therefore, we do not check the existence of '-mA' suffix itself. Instead, we generated a list of all possible negation forms with verbal inflectional suffixes such as person makers and tense suffixes for the detection of '-mA' negative markers (e.g. '-madım', '-medim', '-meyeceğim', '-mayız', '-meyiz', '-meyeceğiz', '-mayacağız' etc.). Then, at dictionary word lookup step, our model checks the occurrence of each verb negation that originated from -mA suffix. If the negative maker is found, system negates the VAD score of the word by subtracting the VAD score of verb root from 6. Some sample verbs and their corresponding negated forms are given in Table 4.4.

Another important negative marker is 'değil' (not) which basically negates nominal sentences. However, in some cases 'değil' can be used to negate verbal sentences

Table 4.4. Comparative verb analyses for negation.

Sentence	**Valence**	**Arousal**	**Dominance**
izin veriyorum	3.7	1.98	3.56
izin ver**mi**yorum	2.3	4.02	2.44
nefret ederek	2.03	3.05	2.83
nefret et**me**den	3.97	2.95	3.17
tehlikeden korudu	3.45	2.38	3.43
tehlikeden koruya**ma**dı	2.55	3.62	2.57
sakinleşecek	3.45	2.31	3.28
sakinleş**me**yecek	2.55	3.69	2.72

as well. In our model ‘değil’ is assumed to negate the VAD score of previous word.

(iii) Evlat, bu yaptığın gerçekten doğru **değil**. (Son, what you are doing is not good.)
(iv) Hergün kitap okuyor **değil**-im. (It’s not the case that I read book everyday.)

Similarly, we detect negative existential term ‘yok’ and update the VAD score of the previous word by subtracting it from 6. Another suffix included in our model was -sHz (can be ‘-siz’, ‘-sız’, ‘-suz’, ‘-süz’ based on the last vowel of the stem). -sHz suffix is used in the meaning of ‘without’ or ‘-less’ when attached to nominal words (e.g. ‘kalpsiz’ [heartless], ‘ümitsiz’ [hopeless]. -sHz also negates verbal sentences in the form of ‘-meksizin’ or ‘maksızın’. Therefore words with -sHz suffix are considered to be negative marker in our model. -sHz suffix negates the VAD score of the word stem that it attaches.

(v) Gülü su**suz** seni aşk**sız** bırak**mam**. (I don’t let the rose be without water, I don’t let you be without love.)
(vi) Bugün hiç birşey yapasım **yok**. (I don’t feel like doing anything today.)
(vii) Bütün gün dur**maksızın** tezimi yazdım. (All day I wrote my thesis non-stop.)

As can be seen from Table 4.5, double negations such as ‘yok değil’, ‘etmeyecek değiliz’ reverse the affective value of the sentence twice.

Table 4.5. Comparative sentence analyses for negation

Sentence	Valence	Arousal	Dominance
Afşin’i davet edeceğiz.	3.34	3.05	3.98
Afşin’i davet edecek **değil**iz.	2.66	2.95	2.02
Afşin’i davet et**me**yeceğiz.	2.66	2.95	2.02
Afşin’i davet et**me**yecek **değil**iz.	3.34	3.05	3.98
Meryem’le sorunumuz var.	1.94	3.3	2.68
Meryem’le sorunumuz **yok**.	4.06	2.7	3.32
Meryem’le sorunumuz **yok değil**.	1.94	3.3	2.68
Meryem’le sorun**suz**uz.	4.06	2.7	3.32

One exceptional function of ‘değil’ is to emphasize the questions as used in the following sentence: ‘Bu söylediklerinde ciddisin değil mi?’. In these cases, model does not treat ‘değil’ as a negative marker, moreover, with this question form, sentence is accepted neutral in overall. Another marker that neutralizes the affective value is ‘ne ... ne ... ’ connective (neither ... nor ...) as illustrated below.

(viii) ‘Hava **ne** iyi **ne** de kötü.’ (The weather is neither good, nor bad.)

(ix) “ Memleket isterim
Ne zengin fakir, **ne** sen ben farkı olsun;
Kış günü herkesin evi barkı olsun. ”
— *Cahit Sıtkı Tarancı*

There also exist negation modifiers such as ‘asla’ (never), ‘katiyen’ (never) which always operate in the presence of other negative markersand fortify the negative meaning. (e.g. ‘**Asla** kimsenin kalbini kırmak iste**me**m.’). In our model, we process these words during modifier check.

4.5.5. Proper names and stop words

Most of the Turkish names such as 'Gizem', 'Çetin' and 'Barış' have high affective loads. Some application domains, especially the teachers' comments include many proper names. In our analysis, we attempted to detect the proper names and neutralize their affective value. For this purpose, a we compiled a list including more than 2 thousand Turkish proper names.

Stop words are usually a challenge for appliacation, as they return large amount of noisy information. Therefore, most of the applications filter out stop words before processing a piece of text. In this direction, we constructed a list of stop words to prevent the irrelevant stem matches with these frequently used words. There is no universal list of stop words. But our 300 item list includes conjunctions such as 'ama', 'hatta', 'dahi', some adjectives such as 'diğer', 'hepsi', 'falan', some certain adverbs such as 'daha', 'böylece', 'henüz', some pronouns (e.g. 'ben', 'onlar') and postpositions (e.g. 'ötürü', 'halinde', 'boyunca' etc.).

4.5.6. Other features

In order to have a comprehensive affect sensing, we examined textual messages for various affective attributes. Apart from the lexical features, we also benefit from some syntactic features such as upper case written texts, letter repetitions and the usage of some special marks. The way we obtained the quantitative scores for the effect of these features is explained in Section 3.2. As can be seen from the Table 4.6, letter repetitions are significantly reinforcing the arousal value in negative, neutral and positive contexts. Similarly, compared to lower case scores of the same sentence, upper case letter usage implies a higher level of arousal both in positive and negative context. On the other hand, upper case letter usage is indicating a lower level of valence in negative context. When we investigated the effect of punctuation marks, we found that exclamation mark expresses 1.4 higher arousal and question mark expresses 0.8 higher arousal when they are added to a negative context. We observed no significant effect of special punctuation marks on valence and dominance dimensions.

Table 4.6. Some important features and their quantitative effect on VAD that are measured by human judges

Sentence	**Sentence with Feature**	**Feature**	**Context Polarity**	**Effect on Val.**	**Effect on Aro.**	**Effect on Dom.**
ayşe	ayşeeeeeeeee	repetition	O	-0.2	+1.2	+0.8
okey gitti	okey gittiiiiiiiiiiii	repetition	N	-0.4	+2.3	+0.4
güzelim	güzeliiimmm	repetition	P	+0.2	+0.9	+0.4
yaş kaç güzelim	yaş kaç güzelim ?	?	P	+0.2	-0.1	0
nasılsınız bugün	nasılsınız bugün ?	?	O	+0.2	-1	-0.1
ne demek istiyorsun ki açık konuş	ne demek istiyorsun ki açık konuş ?	?	N	+0.1	+0.8	-0.6
sus artık amma konuştun	sus artık amma konuştun !	!	N	-0.2	+1.4	+0.1
küfür edenler çıksın salondan	KÜFÜR EDENLER ÇIKSIN SALONDAN	upper case	N	-0.9	+1.5	+0.6
herkese selamlar sevgiler	HERKESE SELAMLAR SEVGİLER	upper case	P	+0.3	+1.1	-0.2

4.5.7. Set of rules

In addition to the features we discussed so far, model is exploited with these linguistic rules to calculate the overall affective score of a sentence:

- The system first checks for the existence of any emoticons, then searchs for words and phrases and assigns the VAD score to each unit from affective dictionary.
- Secondly, the system checks for modifiers, surface features, negations and updates the VAD scores depending on these features.
- If there is any modifier connected to a verb or a noun as phrasal, the score of the word is updated based on the polarity of the sentence and on the particular coefficient of the modifier.
- Considering the transitive verbs in Turkish, for NN+VB structures, such as 'hayatını kaybetti', 'pislikleri temizledim', we take only the affective score of verb and then neutralize the noun.
- Considering the adjective clauses, if there is a NN+ADJ structure such as 'hava çok sıcak', 'kafam karışık', noun is neutralized and then only adjective is taken into consideration.
- If negation is detected, the VAD score is reversed by subtracting from 6 (e.g. 2.3 turns into 3.7).
- If there is an emoticon and a word with a conflicting score (mostly the case for sarcastic and ironic sentences), the emoticon is taken as a reference.
- Negation words 'yok' and 'değil' negate the VAD score of the previous word (again by subtracting each dimensions's unique score from 6).
- If there are no affect-carrying words, emoticons, or interjections in a sentence, then the sentence is considered neutral.

The affect analysis model is implemented with Python programming language. Its readable syntax and powerful libraries make Python an excellent choice for NLP related tasks. The code is properly commented for re-usable and easy-to-understand program. At the end, the project is shared open-source to encourage future contributions [114].

5. EXPERIMENTS AND RESULTS

In this chapter, we first examine the experiments on a continuous dimensional analysis model with the chat and psychotherapy records, then we discuss the results. Secondly, we examine the experiments on a binary (positive or negative) sentiment analysis model with three different data sets; movie reviews, Twitter data and teachers' comments. Lastly, we discuss our findings and the system's limitations with a failure analysis of mispredictions.

5.1. Experiments on Continuous Dimensional Analysis

5.1.1. Affect Analysis of In-game Chat Records

Based on our annotated affective data, we report both fine-grained dimensional and coarse-grained sentiment evaluation results with different metrics. For dimensional evaluation, model scores were scaled continuously between 1 and 5. In Table 5.1, the results for all three dimensions are reported in terms of mean squared error and accuracy. To compute fine-grained accuracy, we calculated the difference between the ground truth score (annotation score) and the predicted model score. If the difference was smaller than 0.5, the prediction was considered to be correct. Since the scale has five intervals, random agreement has %20 accuracy.

Table 5.1. The accuracy of the model for fine-grained dimensional affect estimation

Measure	**Valence**	**Arousal**	**Dominance**
Accuracy	50.0%	58.0%	54.0%
Mean Square Error	0.63	0.56	0.48
Correlation	0.64	0.27	0.42
Means	3.24	2.89	3.2

As can be seen from Table 5.1, the three dimensions studied had different annotation score ranges: Valence: 3.24(±0.88); arousal 2.89(±0.76); dominance 3.2(±0.73). The mean squared error was lower for dominance and arousal compared to valence.

We measured the strength of a relationship between the model predictions and the ground truth values by calculating the Pearson correlation coefficient. As illustrated in Table 5.1, while the accuracy was lowest for valence dimension, correlation between predicted scores and the ground truth annotation was highest for the Valence dimension. Moreover, we observed a very significant correlation among all three dimensions ($P < 0.001$).

Correlation coefficient results are generally interpreted as follow: The larger the correlation, the stronger the relationship; the smaller the p-level, the more significant the relationship. Generally, correlation between .40-.59 is considered as 'moderate', .60-.79 as 'strong' and .80-1.0 as 'very strong.' According to these scales, our findings demonstrated that the the proposed affect model provides accurate predictions on valence and dominance dimensions, relatively weak predictions on arousal dimension. However, it should be kept in mind that the distinctions between sentences were not easy to discern in arousal dimension even for the human annotators.

In the coarse-grained evaluation, model predictions in valence dimension were mapped to positive (>3) and negative (<3) classes to carry out the corresponding polarity detection. With this approach, we obtained %72.9 accuracy when all features were employed (See Table 5.2). In order to comparatively evaluate the performance of our affective lexicon, we tested our model with SentiTurkNet polarity lexicon [115]. In this experiment, all the model features were the same. But we utilized SentiTurkNet polarity lexicon instead of our affective dictionary and obtained %62.8 accuracy with this setup.

Compared to min/max approach, averaging technique gives slightly better results. We observed that our model performs well in the presence of emotive words but it mostly fails with compound sentences expressing different aspects of emotion such as positive and negative opinions at the same time. In contrast to the analysis of other application domains, online chat data analysis yielded the lowest level of accuracy. The major reason behind this low accuracy is the relatively more unstructured and abbreviated nature of the chat sentences. Obviously, there is some room for improvement in both types of scenarios.

Table 5.2. The accuracy of the model for course-grained affect estimation

	Accuracy (%)	
Measure	**Averaging**	**Min/Max**
All features	**72.9**	70.1
All features without normalization	68.3	70.3
All features without negation	68.7	66.7
All features without modifiers	71.5	69.5
All features without emoticons	70	67.6
All features without proper names	72.6	70.1
All features with SentiTurkNet	63.9	60.2

5.1.2. Affect Analysis of Psychotherapy Records

In this application, we used psychotherapy records collected by Istanbul Bilgi University Psychotherapy Research Laboratory (See Section 3.1). We analyzed psychotherapy transcriptions from all therapy sessions under three different categories: adult sentences only, child sentences only, and adult and child sentences combined. As a general rule, linguistic programs need to segment the transcript in equal sized units for comparison of the data while analyzing a text. The length of a scoring unit containing the minimum number of necessary words is determined by statistical procedures described before [116]. In psychotherapy research, an entry with minimum of 150 words is required by many linguistic programs such as the therapeutic Cycle Model and computer-assisted content analysis [117]. Therefore, for the grouping, we created

150-word chunks of sentences while paying attention to play segment borders. Then, in our affect analysis system, each 150 word block was processed as a single sentence.

As demonstrated in Table 5.3, we found significant positive correlation between Valence & Warmth and negative correlation between Dominance & Reason in the Father-Child sentence blocks. The analysis of Mother-Child sentence blocks revealed a significant negative correlation between Dominance & Anxiety.

Table 5.3. Father-Child and Mother-Child external validity

Measure	**Relation**	**Valence Correlation**	**P**	**Arousal Correlation**	**P**	**Dominance Correlation**	**P**
Warmth	Father-Child	**0.463**	0.01	0.29	0.12	**0.372**	0.043
Reason	Father-Child	0.283	0.129	0.339	0.069	**0.409**	0.025
Anxiety	Mother-Child	-0.66	0.694	-0.197	0.236	**-0.423**	0.008

When we evaluated the play factors under three different categories, strong positive correlation was detected for Valence & Isolated comparison and Dominance & Isolated comparison (Table 5.4). In the Father-Child segments, we observed significant positive correlation between Arousal & Play Investment - Affect regulation factors. On the other hand, we observed significant negative correlation between Arousal & Play Investment in the Mother-Child segments. Lastly, In the Therapist-Child segments, we observed significant positive correlation between Arousal & Complex factor, and significant negative correlation between Arousal & Isolated factor as expected. Although we did not detect a correlation among all factors in all dimensions, our results show that the direction of affect can be correctly predicted in all dimensions.

Dimensional relations of jointly analyzed therapist and child segment blocks is shown in Figure 5.1. Interestingly, a strong positive correlation was observed among all affect dimensions throughout the session segments.

Table 5.4. VAD correlation with play factors

Measure	Relation	Val. Cor.	P	Aro. Cor.	P	Dom. Cor.	P
Play Investment	Mother-Child	-0.208	0.187	**-0.308**	0.011	-0.277	0.075
Isolated	Mother-Child	**0.4**	0.009	0.06	0.707	**0.547**	0.000
Affect Regulation	Father-Child	-0.069	0.708	**0.357**	0.045	0.181	0.321
Play Investment	Father-Child	-0.001	0.995	**0.374**	0.035	0.275	0.127
Isolated	Therapist-Child	-0.253	0.089	**-0.372**	0.011	-0.18	0.23
Complex	Therapist-Child	0.152	0.313	**0.295**	0.044	0.112	0.458

Val.:Valence, Aro: Arousal, Dom.:Dominance, Cor.:Correlation

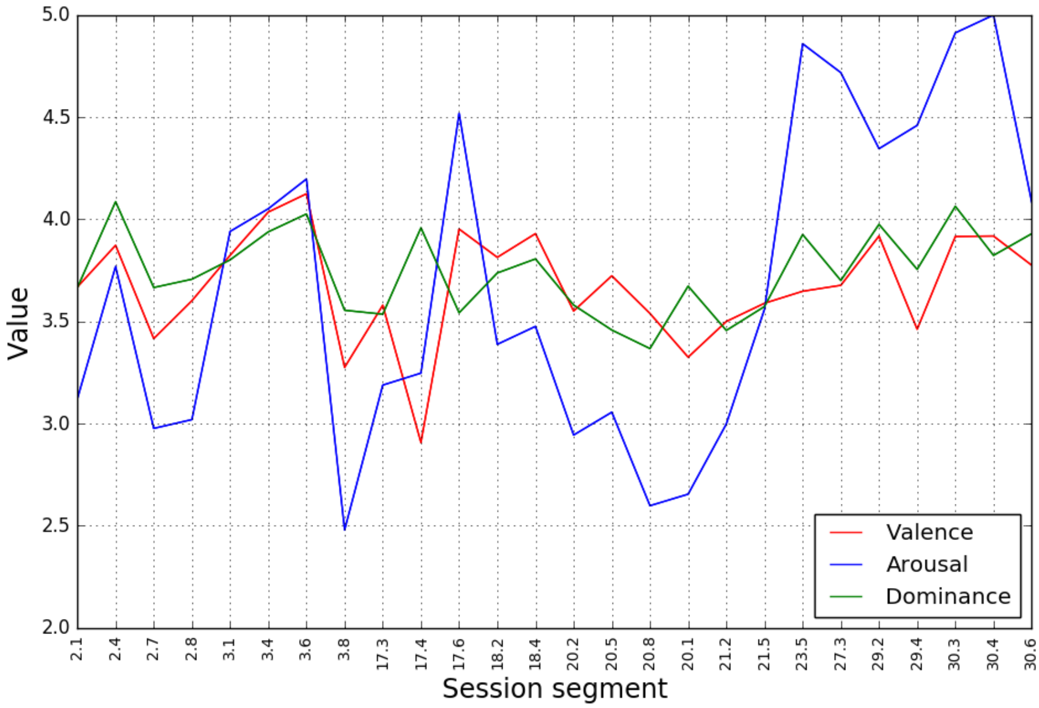


Figure 5.1. Dimensional relations of jointly analyzed therapist and child sentences

Additionally, Affective Regulation and Play Investment factors had significant positive correlations as illustrated in Figure 5.2.

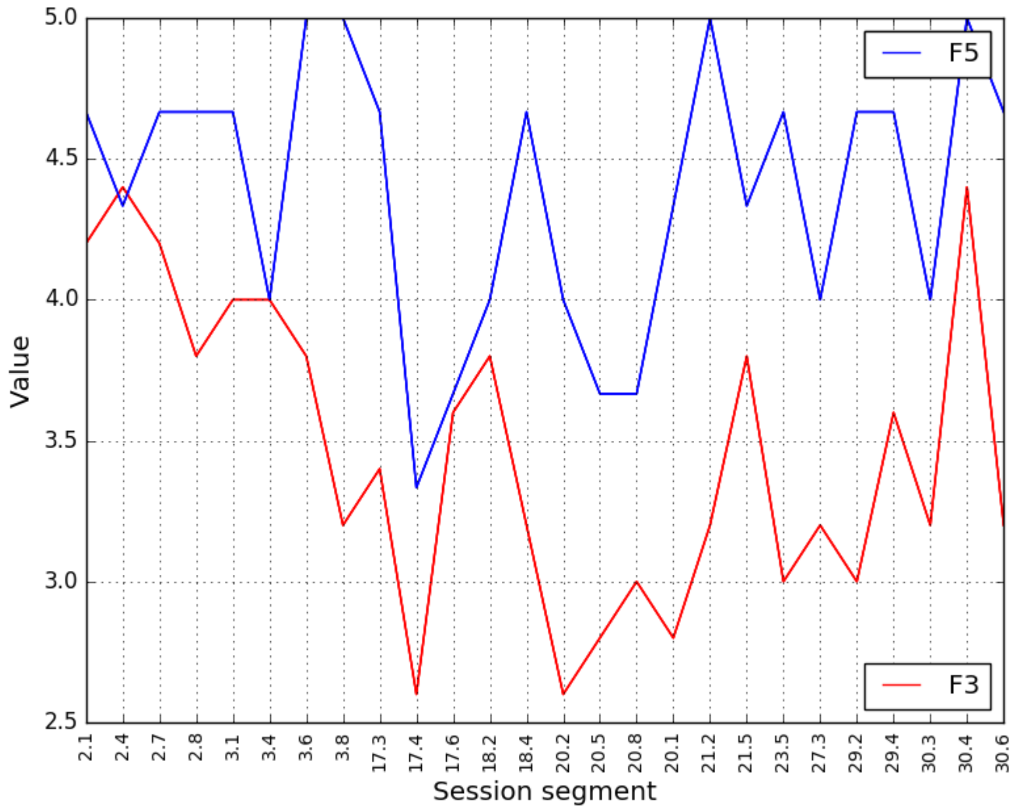


Figure 5.2. An example comparison of Affective Regulation (F3) and Play Investment (F5) of a child

When we look at the Valence, Arousal and Dominance trends for Therapist-Child segment blocks, we observed significant quadratic regression for affect regulation in all dimensions (See Figures 5.3, 5.4, 5.5).

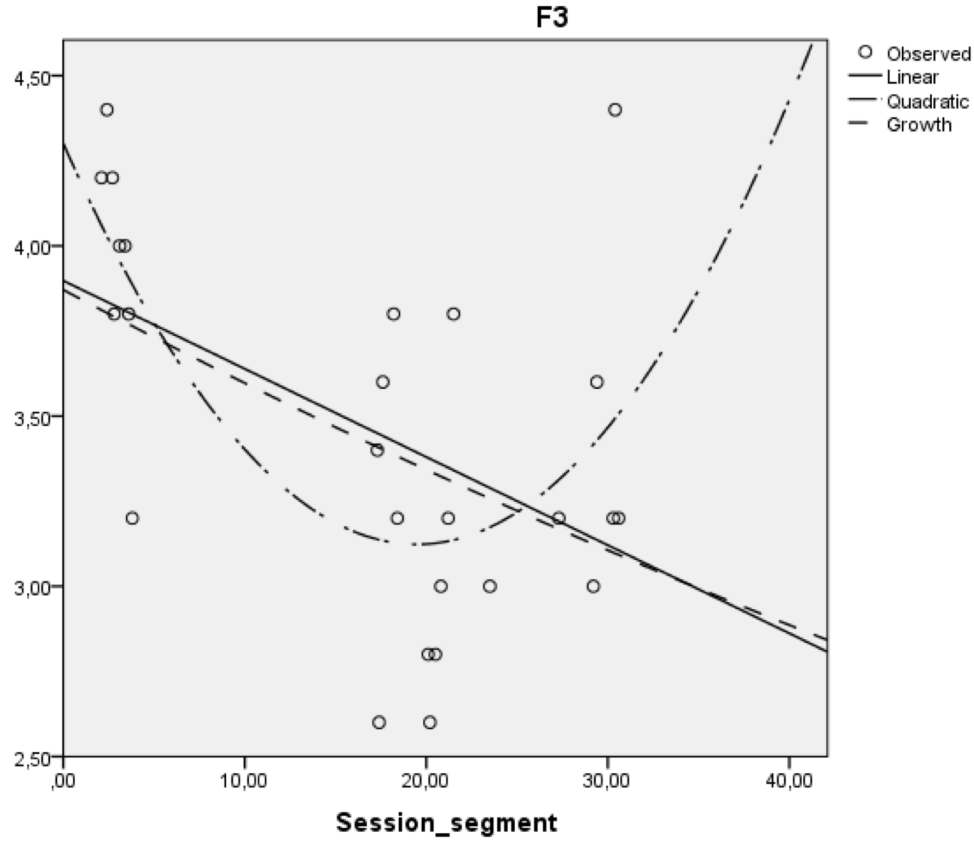


Figure 5.3. Valence trends of a therapist and child

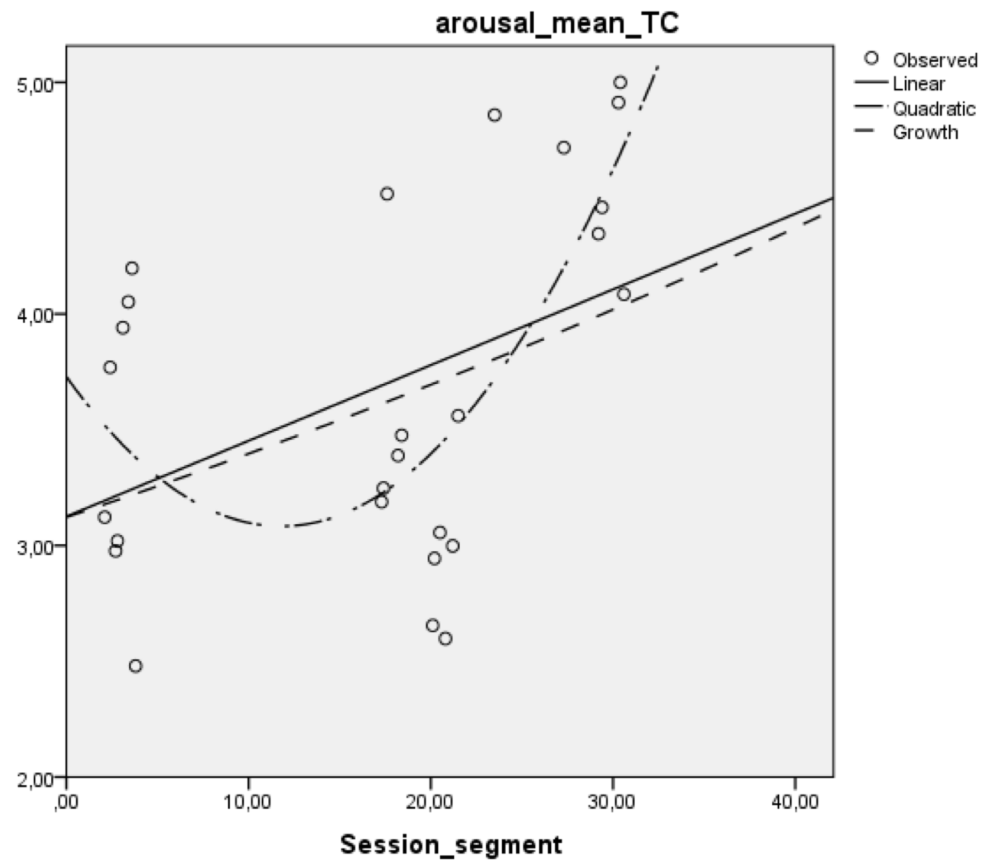


Figure 5.4. Arousal trends of a therapist and child

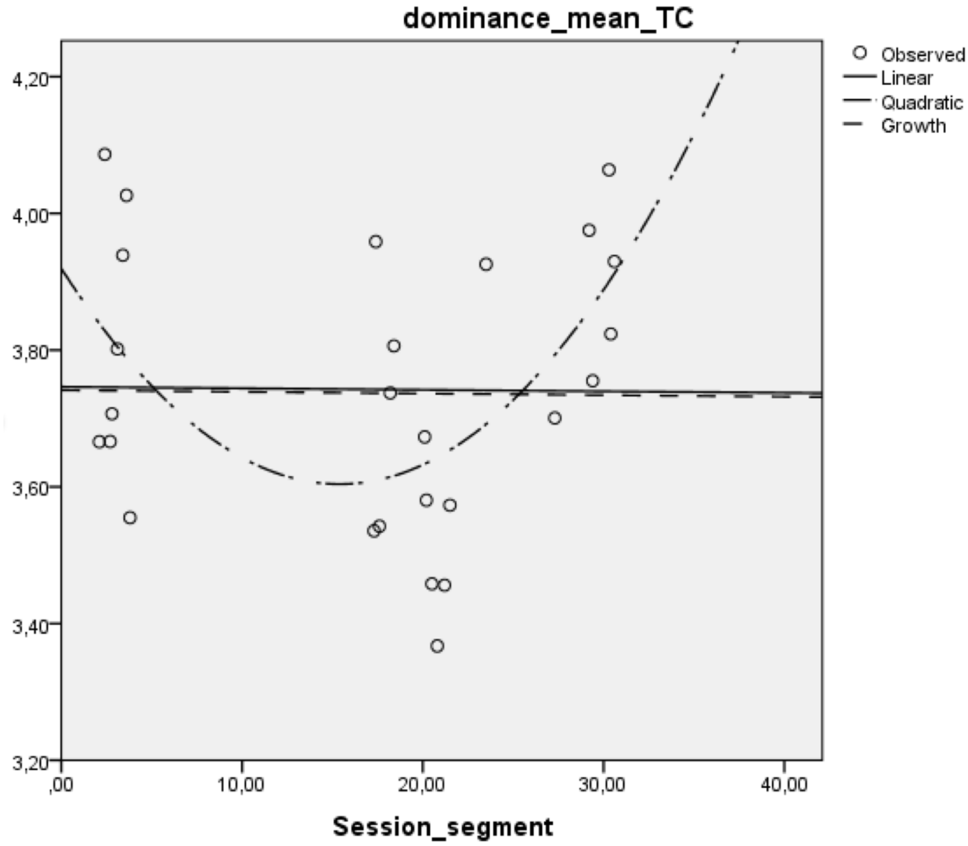


Figure 5.5. Dominance trends of a therapist and child

5.2. Experiments on Sentiment Analysis

To identify the most salient features belonging to the specific domains, we utilized several feature reduction techniques. Initially, we trained approximately 2000 polarity labeled tweets using Document Frequency (DF) method, a method based on ordering of features by their frequencies in the document frequencies. This method is one of the the simplistic measures for feature reduction with linear time complexity. Some examples of positive and negative features are given in the Table 5.5.

Table 5.5. Top sample Twitter words with their DF scores

Positive Features	DF-score	Negative Features	DF-score
mutluluk	105	gece	183
gece	97	bu	150
güzel	91	turkcell	132
turkcell	66	ne	84
ve	36	yok	65
iyi	63	ya	64
çok	61	mutluluk	63
bahar	56	ama	45
en	52	yine	33
var	36	mesaj	30
her	28	hiç	27
huzur	26	hiçbir	18
hayat	17	amk	16
teşekkürler	14	sabah	12
büyük	13	kötü	12

We utilized the Term Frequency - Inverse Document Frequency (TF-IDF) method for labeled tweets. This method involves a weighting scheme mostly used in information retrieval tasks and it aims to model each document into a vector space. This method ignores the order of the words in a document but recognizes the number of occurrences of each feature. Positive and negative feature samples are given in the Table 5.6.

Table 5.6. Top sample Twitter words with TF-IDF weighting

Positive Features	**Negative Features**
sayesinde	pardon
yelken	filan
şahane	sapığım
yağmurları	rezalet
bayram	şov
şarkı	yüksek
mis	gidilmez
canlar	gitmek
güzellik	herhalde
sıkmayan	eskiden
sıradışı	geber

The last feature selection method we used was the Minimum Redundancy Maximum Relevance (mRMR). This method includes an algorithm using mutual information to select top ranking features. This method first identifies a sublist of relevant features. Then the words that are relevant but redundant are removed from the sublist. Table 5.7 illustrates some of the top mRMR features that we extract.

When we implement the DF, TF-IDF and mRMR features, no improvement was observed with any of the features. Potential explanations for lack of improvement are the insufficiency of training data and extraction of labeled tweets with similar keywords. We believe that feeding the model with frequently used non-emotive word features (e.g. ‘ya’, ‘bu’, ‘her’, ‘turkcell’) shadows the effect of important features. Therefore, we did not utilize these features in our model. Compared to other frequency methods, mRMR algorithm achieves to eliminate the redundant features. However, this method does not produce better performance because of the small training size. Using the top DF features such as TF-IDF weighting, we filtered the top words and created a list of

Table 5.7. Top mRMR feature word stems

Name	**Score**	**Name**	**Score**	**Name**	**Score**
hamilelik	0.006	oyalama	0.002	manzara	0.002
adam	0.004	satış	0.002	bit	0.002
kere	0.004	şehr	0.002	cihaz	0.002
mevsim	0.003	yengeç	0.002	diren	0.002
öğren	0.003	sakla	0.002	türlü	0.002
oluştur	0.003	içer	0.002	yüksek	0.002
yedi	0.003	gel	0.002	sokak	0.002
durum	0.003	dizi	0.002	yaklaşık	0.002
taşıma	0.003	dolu	0.002	arkadaş	0.002
koy	0.002	süper	0.002	şen	0.002
ayrıl	0.002	insan	0.002	normal	0.002
hayali	0.002	kocaman	0.002	kavga	0.002

redundant features specific to Twitter domain. Because the performence of the model is reduced by low information, we eliminated the redundant features during the affect analysis by neutralizing their VAD scores if they were present in the input sentence.

Sentiment analysis part of this work is evaluated by four different metrics including Accuracy, F-measure, Precision and Recall. Precision represents the exactness of a classifier. Higher values of precision indicate less false positive (FP) results. Recall measures the number of items in a class is correctly predicted and also interpreted as the completeness, or sensitivity. F measure is the harmonic mean of precision and recall. Calculation of Precision (PRC), Recall (RCL) and F-measure (F) given in Eq. 5.1, 5.2 and 5.3 respectively.

$$PRC = \frac{TP}{TP + FP} \tag{5.1}$$

$$RCL = \frac{TP}{TP + FN} \tag{5.2}$$

$$F = 2 * \frac{PRC * RCL}{PRC + RCL} \tag{5.3}$$

5.2.1. Sentiment Analysis of Turkish Movie Reviews

In this section, we used movie reviews that are randomly collected from website www.beyazperde.com. This list contains 1042 reviews and 2700 sentences which have been labeled in three categories: positive, negative, and neutral [96]. However, we only used positive and negative instances for binary (positive and negative) sentiment classification of movie reviews.

As can be seen from Table 5.8, compared to the other application domains that sentiment classification is tested, we obtained best results with movie reviews. Movie reviews are much more emotive compared to online chat and Twitter data. When we examine the contribution of each individual feature, normalization takes the most significant role in the model's performance with approximately 4% of improvement. Second important feature is the negation handling and then the modifiers. We also observe that eliminating low information features increases performance a little. If we calculate the accuracy with min/max approach system gives 83% accuracy with the all features.

Table 5.8. Accuracy of the model for binary sentiment classification of movie reviews with averaging technique

Measure	**Accuracy (%)**
All features	85.3
All features without normalization	81.3
All features without negation	83.0
All features without modifiers	83.1
All features without conjunction analysis	83.3
Feature reduction	**86.6**
Feature reduction without negation	83.5
Feature reduction without modifiers	85.3
Feature reduction without proper names	86.3
Feature reduction without conjunction analysis	83.4

Table 5.9. Precision, Recall, F1-score with the feature reduction method on movie reviews

CLASS	**Precision**	**Recall**	**F1-score**	**# of instances**
-1	0.66	0.62	0.64	348
1	0.91	0.92	0.91	284
Average	0.86	0.86	0.86	632

When the all features are employed, our model outperforms the state-of-the-art performance for sentiment analysis of movie reviews. Using the same movie reviews dataset, Dehkharghani and colleagues obtained 73% accuracy for sentence level binary classification [96].

One component that we add to the model is the conjunction analysis when we investigate the sentiment analysis. A conjunction has a significant impact on the overall sentiment of a sentence. It is shown that conjunction analysis improves the sentiment classification of movie reviews in English by more than 25% [118]. For example, the sentence ‘oyuncular iyi ama kalitesiz bir film’ includes two conflicting opinions connected with conjunction ‘ama’. In this case, we eliminate the first part of the sentence that comes before conjunction ‘ama’ and take only the second part into consideration. We observed that adding conjunction analysis improves the model performance approximately 2%.

In order to eliminate redundant features from affective dictionary, we extracted a list of domain specific terms in movie reviews which contains 100 terms. Similar to stop words, occurrence of words and phrases from this list, is excluded from affective list. For example terms such as ‘oyuncu’, ‘senaryo’, ‘sinema’, ‘romantik komedi’ normally have high affective scores in dictionary, however, considering that these words function as a part of cinema terminology, we assumed all of these term neutral throughout the analysis in movie review context. We obtained the best accuracy with this feature reduction setup which is 86.6% in binary classification.

5.2.2. Sentiment Analysis of Twitter Data

Our another application domain is Twitter which is one of the most well known micro-blogging service. Again for the binary classification of Twitter data, we used 632 positive and negative polarity labeled tweets.

For the analysis of Twitter data, we observe the best result with feature reduction method with averaging. If we calculate the accuracy with min/max approach, the system gives 70.4% accuracy when all features employed. To the best of our knowledge, the highest accuracy in the early works obtained by Eroglu with 85% accuracy using a SVM classifier [77]. Compared to this work, our 82% accuracy does not outperform the state-of-the-art performance. However, as the size and the content of the testing data would greatly effect the resultant accuracy, we believe that our model produces fairly good results for sentiment analysis of Twitter data.

Table 5.10. Accuracy of the model for sentiment prediction of Twitter data with averaging technique

Measure	**Accuracy %**
All features	**81.00**
All features without normalization	74.1
All features without negation	76.7
All features without modifiers	78.7
All features without proper names	80.3
All features without emoticons	77.3
All features without conjunction analysis	78.9
Feature reduction	**82.1**
Feature reduction without negation	78.4
Feature reduction without modifiers	79.6
Feature reduction without emoticons	78.2
Feature reduction without conjunction analysis	79.7

Table 5.11. Precision, Recall, F1-score with Twitter data when the all features are used

CLASS	**Precision**	**Recall**	**F1-score**	**# of instances**
-1	0.91	0.73	0.81	348
1	0.73	0.91	0.81	284
Average	0.83	0.81	0.81	632

5.2.3. Sentiment Analysis of Teachers' Comments

In this application we use the same sentiment analysis technique to automatically detect the teachers' opinions based on their short comments on students' situations.

Table 5.12. Accuracy of the model for sentiment prediction of teachers' comments with averaging technique

Measure	**Accuracy**
All features	74.2%
All features without normalization	55.2%
All features without negation	69.5%
All features without modifiers	73.6%
All features without proper names	73.2%
Feature reduction	**79.9%**
Feature reduction without negation	76.5%
Feature reduction without modifiers	79.2%
Feature reduction without proper names	78.9%

Similar to other applications we have discussed so far, we get the best accuracy which is %79.9 with feature reduction method. As the comments include a lot of personal proper names, eliminating student names increased performance 1%. Again normalization is the most effective step in improving the system results, then negation contributes approximate 3.5% improvement in this application.

5.3. Limitations and Failure Analysis

In this section, we discuss the limitation of systems and we carry out a failure analysis onexaples of false negative and positive classification to have a better understanding of the underlying problems. The unstructured nature of the texts and meaning ambiguities in the Turkish language are the major factors underlying classification errors.

Table 5.13. Precision, Recall, F1-score with the feature reduction method on teachers' comments

CLASS	Precision	Recall	F1-score	# of Instances
-1	0.85	0.81	0.83	181
1	0.73	0.79	0.76	118
Average	0.8	0.8	0.8	299

Affective values in dictionary were annotated by human judges in a context-free environment. On the other hand, the affective value of a word is highly context dependent. For example, when we say 'I am the boss!' dominance is expected to be high but when we say 'My boss is coming :/' dominance is expected to be lower. However, emotion carrying word is 'boss' for both cases and the system estimates the same score for both of the sentences. Therefore, one important drawback of the system is its inadequacy to generalize to different contexts. Similarly, many classification problems originate from sarcastic and idiomatic expressions, especially in false positive examples. Humans are intuitively capable of interpreting the sarcastic expressions such as 'En yakın arkadaşımın doğum gününü unuttum, harika!'. However, for a machine, it is a very challenging task to resolve the distinction between direct meaning and sarcastic meaning without the contextual understanding. In this particular example, word scores of 'doğum günü' and 'harika' dominate the negative score of 'unutmak'. However, if we could apply the negative contextual meaning of 'unutmak', model might classify sentence accurately. This situation demonstrates the need for the word-sense disambiguation (WSD) that may ameliorate the model for producing more general predictions.

Another limitation for the automatic affect analysis is that each level of a sentence, either words or phrases, may be loaded with different mood or emotion. Secondly, experiencing affect is an idiosyncratic process; affective meaning can be perceived in numerous different ways based on personal experience, characteristics or knowledge level. For instance, if one says ‘Angela is getting married to Sam’, Angela’s best friend would be happy; her mother who never knew that Angela had had such a relationship, would be surprised; Angela’s ex-boyfriend who still loves her might feel depressed and anyone who does not know Angela would be neutral.

Table 5.14. Positive sentences classified as negative. T.C. represents teachers’ comments, T.W. represents Twitter data and M.R. represents movie reviews

Error type	**Domain**	**Sentences**
Multiple opinions	T.C.	arkadaş tutumlarından kısmen **zorlanmakta**. haksızlığa tahammulu az bir öğrenci. **prensipli**.
Conflicting opinions	M.R.	iyi bir film keyifle izleyebilirsiniz ama öyle çok aman aman bir film değil
Conflicting emotions	M.R.	korkunç tüyler ürpertici ve **iğrenç** sahneleri olan **harika** bir yapım
Conflicting emoticon	M.R.	inanılmaz bir aşk hikayesi :(
Word sense	T.C.	keman **çalıyor**.
Ironic remark	M.R.	vaktiniz ve paranız çoksa seyredin derim
Normalization error	T.W.	son yıllrda izlediğm en **eğlncli** filmlerdn biriydi teşekkrler
Word order	M.R.	pek **sürükleyici** bir film **değil**
Sentence fragment	M.R.	Fransız sinemasına son
Morphologically ambiguous	M.R.	bence Korner yönetmenliği **denememeli** artık
Figurative meaning	T.W.	bu havaya **bayılıyorum** yaa ...
Contextual information	T.W.	CNBC-e **şiddetle** tavsiye edilir.
Indirect speech	M.R.	Peter pan’ın fok kids’teki çizgi filimi bile daha güzel
Figurative meaning	M.R.	bence bu filmi izlemeyenler çok şey **kaçırdı**
Negation reversal	M.R.	Dario argento’**suz** bir sinema **düşünemiyorum**

When it comes to teachers’ comments, main reason for the misclassification is the use of both highly positive or negative words when expressing multiple opinions in a single long sentence or complex-compound sentence. As these comments are the short summary notes based on the general level of the student, most of the time teachers give the positive and negative characteristics of the student in the same sentence. Therefore, these contrasting statements are the major source of errors in the sentiment prediction of teachers' comments.

Similarly, conflicting emotions are also a source of failure. For example, lets evaluate the sentence 'inanılmaz bir aşk hikayesi :(' that is shown in Table 5.14. We see that the movie is based on a love story but probably with the bittersweet side of the movie, commentator experiences sadness and uses a negative emoticon.

By averaging the word/phrase level sentiment scores, we lose significant amount of dependency information. Emotion is generally conveyed by connected units where the dependency of these units usually plays the key role to form the absolute sentiment of a sentence. From Table 5.15, we see the sentence 'filmin başyapıt olduğunu iddia edenlerin de maalesef samimi olduğuna inanamıyorum' which is misclassified due to such problem. The verb 'inanmıyorum' is dependent to the phrasal 'samimi olduğuna' but the model is not capable of detecting such a dependency. For this reason, we believe analyzing dependency structure of the sentence is crucially necessary to obtain a more complete system.

Another major source for failure is the insufficiency in handling negation. Our model captures the presence of negative words and suffixes, however, there is no general rule on which emotive word is modified. In our model, we assumed the basic word order SOV (subject-object-verb) is the most likely order for the words in a sentence. For example, an intensifier comes before the word that it scales or the markers 'değil' and 'yok' negate the previous word. However, word order is not mandatory in Turkish, so people are freely changing word order sometimes because of stylistic concerns, sometimes just because of laziness. This means that there can be other words between the affected word and the negative modifier. As illustrated in Table 5.15, in the sentence 'pek sürükleyici bir film değil' (it's not a fascinating movie), 'değil' negates the word 'sürükleyici', but our algorithm negates the previous word 'film'. Besides, some excep-

Table 5.15. Negative sentences classified as positive. T.C. represents teachers' comments, T.W. represents Twitter data and M.R. represents movie reviews

Error type	**Domain**	**Sentences**
Idiomatic speech	T.W.	Bugün için farklı planlarım vardı hepsi **suya düştü**.
Metaphorical expression	T.W.	Görmedin mi gözlerimde **yağmur** vardı ?
Multiple opinions	T.C.	aile problemleri, annesiyle **iyi anlaşıyor algısı zayıf**
Dependency	M.R.	filmin başyapıt olduğunu iddia edenlerin de maalesef **samimi olduğuna inanamıyorum**
Incomplete sentence	T.W.	Doğumgünlerinin telefonla kutlanmayıp facebook gibi sanal alemlerden kutlanması hiç **hoş**
Normalization error	T.W.	artık beklemek **ısıtmıyorum**
Idiomatic speech	M.R.	5 puan verdim o da Aniston'un **güzel yüzünün hatırına**
Word order	T.W.	rahat yok sanırım bu hayatta ... Bir ahize nasıl **acıtır canımı**
Lexicon	T.W.	Interstaller o kadar **ilgimi çekmiyor**
Indirect speech	M.R.	bu filme güzel diyenler eminim daha önce güzel film izlememişler
Negation error	M.R.	pek **sürükleyici** bir film **değil**
Conflicting emotions	T.W.	Başkanım inşallah **:)** Bizim dolandırıcı yönetici **kaçtı** ... Kaloriferler yanmıyor
Conflicting emoticon	T.W.	**:)** herkese günaydın.. Ben hala **hastayım**
Sarcastic expression	M.R.	saçma bir konuyu nasılda film yapmışlar **maşallah**
Lexicon	T.C.	çalışırken ciddi **sıkıntı yaşıyor**
Sentence fragment	T.C.	derslerinin iyiye gitmemesi

tional usages of these negative markers are inevitable. For instance, 'değil' can also function in means of *ellipsis* or in the meaning of 'şöyle dursun' and 'bir yana' [113] as in the following sentences:

(i) Adem döndü sonra Havva'ya; elma **değil** bu yediğimiz galiba, ayvaa!

(ii) **Değil** aileni, seni bile görmek istemiyorum! (I do not want to see you, let alone the rest of your family!)

In these cases, obviously it is not convenient to treat 'değil' as a negative marker. Unfortunately the system is incapable of differentiating the alternative meanings of such words, although we have tried to cover all forms of negations.

Another cause of error in classification is the insufficiency of the lexicon. When we consider the verb 'görmek' (to see) individually, it has a positive valence score of 3.64. In contrast, the phrasal verb 'hor görmek' (to humiliate) has a very low valence score; 2.11. Luckily this pharasal verb exists in our affective dictionary. Turkish languge is very rich of phrasal verbs and there exist thousands of such verb forms which are not included in our dictionary. For example, in the sentence 'Interstaller o kadar ilgimi çekmiyor', phrasal verb 'ilgimi çekmiyor' does not match any dictionary item. One limitation specific to Turkish is its agglutinatively structured nature which makes morphological analysis more challenging and error-prone. Development of more powerful NLP tools in the future might solve this problem to some extent.

6. CONCLUSION

6.1. Conclusions

In the scope of this thesis work, we investigated the affect analysis problem for communication texts in Turkish. In order to accomplish this task, we described a rule-based system for dimensional affect analysis using syntactic and lexical cues in Turkish. Two application domains for fine-grained dimensional affect analysis were multi-party in-game chat records and psychotherapy records.

There exist a limited number of tools and prior studies for affect analysis in the Turkish language. Although the immediate goal and the main contribution of this study was creating an affect analysis model for Turkish language, at a deeper level, we provide tools and comprehensive lexical resources that would hopefully ease the work of researchers for further research.

We developed our model specifically for communication texts, which typically include grammatical irregularities, and chat-specific expressions, emoticons, abbreviations and limited vocabulary. We presented a comprehensive affective dictionary with more than 15,000 Turkish words with valence, arousal, and dominance scores in the [1-5] scale. We believe that these resources can be very beneficial to construct more sophisticated NLP applications to perform tasks such as user profiling and social network analysis.

According to the evaluation results, our findings are promising for the fine-grained and coarse-grained affect analysis of game-in chat logs. In fine-grained analysis, our results show that there is a significant correlation between the predicted affect scores and the ground truth annotation scores in all dimensions. In course-grained analysis, we observed 73% positive/negative classification accuracy.

The proposed dimensional model also succeeds at sentiment analysis on a broad range of text domain. In support of this claim, we evaluated our model with Turkish Twitter data, teachers' comments for high school students and Turkish movie reviews for polarity level sentiment analysis. According to the results, our model has a very advanced performance for sentiment analysis of movie reviews with an accuracy level over 86%. With the Twitter data, the model performs 82% accuracy and with teachers' comments it performs 80% accuracy. Unfortunately, there is no comparative study for the opinion mining of teachers' comments in Turkish. Similarly, due to the lack of studies specific to dimensional affect analysis of psychotherapy records, we do not report any comparative results. However, the significant correlations between the factors such as the valence-warmth and dominance-anxiety bring an important dimension to our understanding of the relations between affect and these psychological variables.

6.2. Future work

Our initial efforts in creating affective resources for Turkish appears to be promising. The methodology we proposed for affect sensing includes several dictionary look up and the dictionary contains vast amount of relatively less meaningful words. With the help of error analysis on test sets, we observe that any feature selection method with an extensive training set would improve the accuracy of the system depending the specific text domain.

The initial application domain of the study was multi-party online social games. As indicated in [119], person-dependent models are more successful in affect recognition. A possible extension of the study is to learn person-dependent models by further employing contextual information (e.g. game data for multi-party chat during games). Secondly, social sharing studies have shown that men are much more likely to talk to women about their feelings, whereas women share their emotions with a wider range of persons [120]. Besides, women are more likely to thank, appreciate and apologize, whereas men seem less concerned with politeness and sometimes violate expected online conduct [121]. From this perspective, we believe that gender factor can be integrated with the model for more gender-specific analysis.

ML models can be incorporated when larger annotated affective data is available for Turkish language. For supervised machine learning approaches, more extensive annotation of sentences is required. The existing work can help with annotation efforts by focusing on parts of the data space where predictions are poor.

The proposed system can also be combined with dependency parser in order to achieve more sophisticated affect analysis systems. A more detailed morphological analysis and ambiguity resolution is also necessary for development of the model.

Another possible improvement is to recognize discrete emotion categories such as ‘anger’, ‘fear’, ‘joy’, ‘surprise’ etc. This can be accomplished by using categorical word dictionaries with their synsets.

The affective lexicon and all the other resources created in this study are made publicly available to promote future contributions to this field.

REFERENCES

1. Picard, R. W. and R. Picard, *Affective computing*, Vol. 252, MIT Press Cambridge, 1997.

2. Zeng, Z., M. Pantic, G. I. Roisman and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 31, No. 1, pp. 39–58, 2009.

3. Anderson, K. and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions", *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, Vol. 36, No. 1, pp. 96–105, 2006.

4. Castellano, G., S. D. Villalba and A. Camurri, "Recognising human emotions from body movement and gesture dynamics", *Affective computing and intelligent interaction*, pp. 71–82, Springer, 2007.

5. Zhang, L. and J. Barnden, "Affect sensing using linguistic, semantic and cognitive cues in multi-threaded improvisational dialogue", *Cognitive Computation*, Vol. 4, No. 4, pp. 436–459, 2012.

6. Chuang, Z.-J. and C.-H. Wu, "Multi-modal emotion recognition from speech and text", *Computational Linguistics and Chinese Language Processing*, Vol. 9, No. 2, pp. 45–62, 2004.

7. Aydın Oktay, E., K. Balcı and A. A. Salah, "Automatic Assessment of Dimensional Affective Content in Turkish Multi-party Chat Messages", *Proceedings of the International Workshop on Emotion Representations and Modelling for Companion Technologies*, pp. 19–24, ACM, 2015.

8. Izard, C. E., "Four systems for emotion activation: cognitive and noncognitive processes.", *Psychological review*, Vol. 100, No. 1, p. 68, 1993.

9. Kleinginna Jr, P. R. and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition", *Motivation and emotion*, Vol. 5, No. 4, pp. 345–379, 1981.

10. Ortony, A., *The cognitive structure of emotions*, Cambridge university press, 1990.

11. Cornelius, R. R., *The science of emotion: Research and tradition in the psychology of emotions.*, Prentice-Hall, Inc, 1996.

12. Gergen, K. J., "The social constructionist movement in modern psychology.", *American psychologist*, Vol. 40, No. 3, p. 266, 1985.

13. Shouse, E., "Feeling, emotion, affect", *M/c Journal*, Vol. 8, No. 6, p. 26, 2005.

14. Munezero, M., C. S. Montero, E. Sutinen and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text", *Affective Computing, IEEE Transactions on*, Vol. 5, No. 2, pp. 101–111, 2014.

15. Calvo, R. A. and S. Mac Kim, "Emotions in text: dimensional and categorical models", *Computational Intelligence*, Vol. 29, No. 3, pp. 527–543, 2013.

16. Gunes, H., B. Schuller, M. Pantic and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey", *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pp. 827–834, IEEE, 2011.

17. Ekman, P., W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion.", *Journal of personality and social psychology*, Vol. 53, No. 4, p. 712, 1987.

18. Izard, C. E., "Basic emotions, relations among emotions, and emotion-cognition relations.", *Psychological Review*, Vol. 99(3), pp. 561–565, 1992.

19. Plutchik, R., “A general psychoevolutionary theory of emotion”, *Theories of emotion*, Vol. 1, 1980.

20. Wundt, W., *Outlines of psychology*, Springer, 1980.

21. Osgood, C. E., “On the whys and wherefores of E, P, and A.”, *Journal of personality and social psychology*, Vol. 12, No. 3, p. 194, 1969.

22. Russell, J. A., “A circumplex model of affect”, *Journal of personality and social psychology*, Vol. 39, No. 6, p. 1161, 1980.

23. Watson, D. and A. Tellegen, “Toward a consensual structure of mood.”, *Psychological bulletin*, Vol. 98, No. 2, p. 219, 1985.

24. Bradley, M. M., M. K. Greenwald, M. C. Petry and P. J. Lang, “Remembering pictures: pleasure and arousal in memory.”, *Journal of experimental psychology: Learning, Memory, and Cognition*, Vol. 18, No. 2, p. 379, 1992.

25. Landauer, T. K., P. W. Foltz and D. Laham, “An introduction to latent semantic analysis”, *Discourse processes*, Vol. 25, No. 2-3, pp. 259–284, 1998.

26. Gill, A., R. French, D. Gergle and J. Oberlander, “Identifying emotional characteristics from short blog texts”, *Proc. 30th Ann. Conf. Cognitive Science Soc., BC Love, K. McRae, and VM Sloutsky, eds*, pp. 2237–2242, 2008.

27. Strapparava, C. and R. Mihalcea, “Learning to identify emotions in text”, *Proceedings of the 2008 ACM symposium on Applied computing*, pp. 1556–1560, ACM, 2008.

28. D’mello, S. K., S. D. Craig, A. Witherspoon, B. Mcdaniel and A. Graesser, “Automatic detection of learner’s affect from conversational cues”, *User modeling and user-adapted interaction*, Vol. 18, No. 1-2, pp. 45–80, 2008.

29. Russell, J. A. and L. F. Barrett, “Core affect, prototypical emotional episodes,

and other things called emotion: dissecting the elephant.", *Journal of personality and social psychology*, Vol. 76, No. 5, p. 805, 1999.

30. Thayer, R. E., *The biopsychology of mood and arousal*, Oxford University Press, 1989.

31. Mehrabian, A., "Basic Dimensions for a General Psychological Theory Implications for Personality, Social, Environmental, and Developmental Studies", , 1980.

32. Posner, J., J. A. Russell and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology", *Development and psychopathology*, Vol. 17, No. 03, pp. 715–734, 2005.

33. Plutchik, R., "The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice", *American Scientist*, Vol. 89, No. 4, pp. 344–350, 2001.

34. Lövheim, H., "A new three-dimensional model for emotions and monoamine neurotransmitters", *Medical hypotheses*, Vol. 78, No. 2, pp. 341–348, 2012.

35. Reisenzein, R., "Pleasure-arousal theory and the intensity of emotions.", *Journal of Personality and Social Psychology*, Vol. 67, No. 3, p. 525, 1994.

36. Frijda, N. H., "Moods, emotion episodes, and emotions.", , 1993.

37. Derks, D., A. H. Fischer and A. E. Bos, "The role of emotion in computer-mediated communication: A review", *Computers in Human Behavior*, Vol. 24, No. 3, pp. 766–785, 2008.

38. Pang, B. and L. Lee, "Opinion mining and sentiment analysis", *Foundations and trends in information retrieval*, Vol. 2, No. 1-2, pp. 1–135, 2008.

39. Cambria, E., B. Schuller, Y. Xia and C. Havasi, "New avenues in opinion mining

and sentiment analysis", *IEEE Intelligent Systems*, , No. 2, pp. 15–21, 2013.

40. Zhe, X. and A. Boucouvalas, "Text-to-emotion engine for real time internet communication", *Proceedings of International Symposium on Communication Systems, Networks and DSPs*, pp. 164–168, 2002.

41. Ma, C., H. Prendinger and M. Ishizuka, "Emotion estimation and reasoning based on affective textual interaction", *Affective computing and intelligent interaction*, pp. 622–628, Springer, 2005.

42. Dey, L., M.-U. Asad, N. Afroz and R. P. D. Nath, "Emotion extraction from real time chat messenger", *Proc. ICIEV*, pp. 1–5, IEEE, 2014.

43. Neviarouskaya, A., H. Prendinger and M. Ishizuka, "Affect analysis model: novel rule-based approach to affect sensing from text", *Natural Language Engineering*, Vol. 17, No. 01, pp. 95–135, 2011.

44. Brooks, M., K. Kuksenok, M. K. Torkildson, D. Perry, J. J. Robinson, T. J. Scott, O. Anicello, A. Zukowski, P. Harris and C. R. Aragon, "Statistical affect detection in collaborative chat", *Proc. CSCW*, pp. 317–328, ACM, 2013.

45. Liu, H., H. Lieberman and T. Selker, "A model of textual affect sensing using real-world knowledge", *Proc. IUI*, pp. 125–132, ACM, 2003.

46. Singh, P., T. Lin, E. T. Mueller, G. Lim, T. Perkins and W. L. Zhu, "Open Mind Common Sense: Knowledge acquisition from the general public", *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, pp. 1223–1237, Springer, 2002.

47. Lenat, D. B., "CYC: A large-scale investment in knowledge infrastructure", *Communications of the ACM*, Vol. 38, No. 11, pp. 33–38, 1995.

48. Mueller, E. T., *ThoughtTreasure: A natural language/commonsense platform*, New York: Signiform, 1998.

49. Wu, T., F. M. Khan, T. A. Fisher, L. A. Shuler and W. M. Pottenger, "Posting act tagging using transformation-based learning", *Foundations of Data Mining and knowledge Discovery*, pp. 319–331, Springer, 2005.

50. Danışman, T. and A. Alpkoçak, "Feeler: Emotion classification of text using vector space model", .

51. Sánchez, J. A., N. P. Hernández, J. C. Penagos and Y. Ostróvskaya, "Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states", *Proceedings of VII Brazilian symposium on Human factors in computing systems*, pp. 66–72, ACM, 2006.

52. Tsetserukou, D., A. Neviarouskaya, H. Prendinger, N. Kawakami, M. Ishizuka and S. Tachi, "iFeel_IM! Emotion enhancing garment for communication in affect sensitive instant messenger", *Human Interface and the Management of Information. Designing Information Environments*, pp. 628–637, 2009.

53. Neviarouskaya, A., H. Prendinger and M. Ishizuka, "EmoHeart: conveying emotions in second life based on affect sensing from text", *Advances in Human-Computer Interaction*, Vol. 2010, p. 1, 2010.

54. Tat, A. and S. Carpendale, "CrystalChat: Visualizing personal chat history", *Proc. 39th Annual Hawaii Int. Conf. on System Sciences*, Vol. 3, p. 58c, IEEE, 2006.

55. Zheng, D., F. Tian, J. Liu, Q. Zheng and J. Qin, "Emotion Chat: A Web Chatroom with Emotion Regulation for E-Learners", *Physics Procedia*, Vol. 25, pp. 763–770, 2012.

56. Wang, H., H. Prendinger and T. Igarashi, "Communicating emotions in online chat using physiological sensors and animated text", *CHI'04 extended abstracts on Human factors in computing systems*, pp. 1171–1174, ACM, 2004.

57. Fragopanagos, N. and J. G. Taylor, "Emotion recognition in human–computer interaction", *Neural Networks*, Vol. 18, No. 4, pp. 389–405, 2005.

58. Neviarouskaya, A., H. Prendinger and M. Ishizuka, "Textual affect sensing for sociable and expressive online communication", *Affective Computing and Intelligent Interaction*, pp. 218–229, Springer, 2007.

59. Holzman, L. E. and W. M. Pottenger, "Classification of emotions in internet chat: An application of machine learning using speech phonemes", *Retrieved November*, Vol. 27, p. 2011, 2003.

60. Scott, T. J., K. Kuksenok, D. Perry, M. Brooks, O. Anicello and C. Aragon, "Adapting grounded theory to construct a taxonomy of affect in collaborative online chat", *Proceedings of the 30th ACM international conference on Design of communication*, pp. 197–204, ACM, 2012.

61. Whissell, C., "The dictionary of affect in language", *Emotion: Theory, research, and experience*, Vol. 4, No. 113-131, p. 94, 1989.

62. Bradley, M. M. and P. J. Lang, *Affective norms for English words (ANEW): Instruction manual and affective ratings*, Tech. rep., C-1, The Center for Research in Psychophysiology, Univ. of Florida, 1999.

63. Strapparava, C. and A. Valitutti, "WordNet Affect: an Affective Extension of WordNet.", *LREC*, Vol. 4, pp. 1083–1086, 2004.

64. Magnini, B. and G. Cavaglia, "Integrating Subject Field Codes into WordNet.", *LREC*, 2000.

65. Kahn, J. H., R. M. Tobin, A. E. Massey and J. A. Anderson, "Measuring emotional expression with the Linguistic Inquiry and Word Count", *The American journal of psychology*, pp. 263–286, 2007.

66. Hancock, J. T., C. Landrigan and C. Silver, "Expressing emotion in text-based

communication", *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 929–932, ACM, 2007.

67. Pennebaker, J. W., M. R. Mehl and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves", *Annual review of psychology*, Vol. 54, No. 1, pp. 547–577, 2003.

68. Cambria, E., R. Speer, C. Havasi and A. Hussain, "SenticNet: A Publicly Available Semantic Resource for Opinion Mining.", *AAAI Fall Symposium: Commonsense Knowledge*, Vol. 10, p. 02, 2010.

69. Cambria, E., C. Havasi and A. Hussain, "SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis.", *FLAIRS Conference*, pp. 202–207, 2012.

70. de Albornoz, J. C., L. Plaza and P. Gervás, "SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis.", *LREC*, pp. 3562–3567, 2012.

71. Poria, S., A. Gelbukh, E. Cambria, P. Yang, A. Hussain and T. Durrani, "Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis", *Signal Processing (ICSP), 2012 IEEE 11th International Conference on*, Vol. 2, pp. 1251–1255, IEEE, 2012.

72. Das, D. and S. Bandyopadhyay, "Developing Bengali WordNet Affect for Analyzing Emotion", *International Conference on the Computer Processing of Oriental Languages*, pp. 35–40, 2010.

73. Torii, Y., D. Das, S. Bandyopadhyay and M. Okumura, "Developing Japanese WordNet affect for analyzing emotions", *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 80–86, Association for Computational Linguistics, 2011.

74. Bobicev, V., V. Maxim, T. Prodan, N. Burciu and V. Angheluş, "Emotions in words: Developing a multilingual wordnet-affect", *Computational Linguistics and Intelligent Text Processing*, pp. 375–384, Springer, 2010.

75. Vural, A. G., B. B. Cambazoglu, P. Senkul and Z. O. Tokgoz, "A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish", *Computer and Information Sciences III*, pp. 437–445, Springer, 2013.

76. Dehkharghani, Y. Saygın, B. Yanıkoğlu and K. Oflazer, "SentiTurkNet: a Turkish polarity lexicon for sentiment analysis", *Language Resources and Evaluation*, pp. 1–19, 2015.

77. Eroğul, U., "Sentiment analysis in Turkish", *Middle East Technical University, Ms Thesis, Computer Engineering*, 2009.

78. Çakmak, O., A. Kazemzadeh, D. Can, S. Yildirim and S. Narayanan, "Root-word analysis of Turkish emotional language", *Corpora for Research on Emotion Sentiment & Social Signals*, 2012.

79. Boynukalın, Z. and P. Karagoz, "Emotion Analysis on Turkish Texts", E. Gelenbe and R. Lent (Editors), *Information Sciences and Systems*, Vol. 264 of *LNEE*, pp. 159–168, 2013.

80. Boynukalın, Z., "Emotion analysis of Turkish texts by using machine learning methods", *Middle East Technical University*, 2012.

81. Kaya, M., G. Fidan and I. H. Toroslu, "Sentiment analysis of Turkish political news", *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, pp. 174–180, IEEE Computer Society, 2012.

82. Akba, F., A. Uçan, E. A. Sezer and H. Sever, "Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews", *8th European Conference*

on Data Mining 2014, 2014.

83. Yıldırım, E., F. S. Çetin, G. Eryiğit and T. Temel, "The Impact of NLP on Turkish Sentiment Analysis", *Proc. of the Int. Conf. on Turkic Language Processing*, pp. 7–13, 2014.

84. Başdemir, B., *Emotion detection in novels*, Ph.D. Thesis, 2015.

85. "Ekşi sözlük", `https://eksisozluk.com`, accessed at February 2016.

86. Isgüder-Şahin, G. G., H. R. Zafer and E. Adah, "Polarity detection of Turkish comments on technology companies", *Asian Language Processing (IALP), 2014 International Conference on*, pp. 136–139, IEEE, 2014.

87. Gezici, G., B. Yanıkoğlu, D. Tapucu and Y. Saygın, "New features for sentiment analysis: Do sentences matter", *SDAD 2012 The 1st international workshop on sentiment discovery from affective data*, p. 5, 2012.

88. Çetin, M. and M. Fatih Amasyalı, "Active learning for Turkish sentiment analysis", *Innovations in Intelligent Systems and Applications (INISTA), 2013 IEEE International Symposium on*, pp. 1–4, June 2013.

89. Demirtaş, E. and M. Pechenizkiy, "Cross-lingual polarity detection with machine translation", *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, p. 9, ACM, 2013.

90. Özsert, C. M. and A. Özgür, "Word polarity detection using a multilingual approach", *Computational Linguistics and Intelligent Text Processing*, pp. 75–82, Springer, 2013.

91. Bilgin, O., Ö. Çetinoğlu and K. Oflazer, "Building a wordnet for Turkish", *Romanian Journal of Information Science and Technology*, Vol. 7, No. 1-2, pp. 163–172, 2004.

92. Tufis, D., D. Cristea and S. Stamou, “BalkaNet: Aims, methods, results and perspectives. a general overview”, *Romanian Journal of Information science and technology*, Vol. 7, No. 1-2, pp. 9–43, 2004.

93. Balcı, K. and A. A. Salah, “Automatic analysis and identification of verbal aggression and abusive behaviors for online social games”, *Computers in Human Behavior*, Vol. 53, pp. 517 – 526, 2015.

94. “Istanbul Bilgi University Psychotherapy Research Laboratory”, `http://www.bilgipsikoterapiarastirmalari.com`, accessed at December 2015.

95. Achenbach, T. M., *Manual for the Child Behavior Checklist/4-18 and 1991 profile*, Department of Psychiatry, University of Vermont Burlington, VT, 1991.

96. Dehkharghani, R., B. Yanıkoğlu, D. Tapucu and Y. Saygın, “Adaptation and use of subjectivity lexicons for domain dependent sentiment classification”, *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference*, pp. 669–673, IEEE, 2012.

97. “SentiLab”, `http://sentilab.sabanciuniv.edu/resources`, accessed at February 2016.

98. Ekman, P., W. V. Friesen and P. Ellsworth, *Emotion in the human face: Guidelines for research and an integration of findings*, Elsevier, 2013.

99. Bradley, M. M. and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential”, *Journal of behavior therapy and experimental psychiatry*, Vol. 25, No. 1, pp. 49–59, 1994.

100. Sheng, V. S., F. Provost and P. G. Ipeirotis, “Get another label? Improving data quality and data mining using multiple, noisy labelers”, *Proc. 14th ACM SIGKDD*, pp. 614–622, 2008.

101. Krippendorff, K., “Computing Krippendorff’s alpha reliability”, *Departmental Papers (ASC)*, p. 43, 2007.

102. Artstein, R. and M. Poesio, “Inter-coder agreement for computational linguistics”, *Computational Linguistics*, Vol. 34, No. 4, pp. 555–596, 2008.

103. Krippendorff, K., *Content analysis: An introduction to its methodology*, Sage, 2012.

104. Landis, J. R. and G. G. Koch, “The measurement of observer agreement for categorical data”, *biometrics*, pp. 159–174, 1977.

105. Warriner, A. B., V. Kuperman and M. Brysbaert, “Norms of valence, arousal, and dominance for 13,915 English lemmas”, *Behavior research methods*, Vol. 45, No. 4, pp. 1191–1207, 2013.

106. “TDK”, `http://www.tdk.gov.tr`, accessed at May 2015.

107. Torunoğlu, D. and G. Eryiğit, “A Cascaded Approach for Social Media Text Normalization of Turkish”, *5th Workshop on Language Analysis for Social Media (LASM) at EACL*, pp. 62–70, Association for Computational Linguistics, Gothenburg, Sweden, April 2014.

108. “ITU Turkish NLP pipeline”, http://tools.nlp.itu.edu.tr/Normalization.

109. Polanyi, L. and A. Zaenen, “Contextual valence shifters”, *Computing attitude and affect in text: Theory and applications*, pp. 1–10, Springer, 2006.

110. Dilek Z. Hakkani-Tür, G. T., Kemal Oflazer, “Statistical Morphological Disambiguation for Agglutinative Languages”, *Computers and the Humanities*, Vol. 36, No. 4, pp. 381–410, 2002.

111. Oflazer, K., “Two-level description of Turkish morphology”, *Literary and linguistic computing*, Vol. 9, No. 2, pp. 137–148, 1994.

112. Oflazer, K., E. Göçmen and C. Bozşahin, "An Outline of Turkish Morphology", *Report on Turkish Natural Language Processing Initiative Project*, 1994.

113. Göksel, A. and C. Kerslake, *Turkish: A comprehensive grammar*, Routledge, 2004.

114. "Affect resources and the tools of the Turkish affect analysis model", https://github.com/verdeerosso/affective-turkish.

115. Dehkharghani, R., Y. Saygin, B. Yanikoglu and K. Oflazer, "SentiTurkNet: a Turkish polarity lexicon for sentiment analysis", *Language Resources and Evaluation*, pp. 1–19, 2015, http://dx.doi.org/10.1007/s10579-015-9307-6.

116. Mergenthaler, E., "Textbank systems", *Computer science applied in the field of psychoanalysis*, 1985.

117. Mergenthaler, E., "Computer-assisted content analysis", *Nachtrichten Spezial*, pp. 3–32, 1996.

118. Meena, A. and T. Prabhakar, *Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis*, Springer, 2007.

119. D'Mello, S. K. and J. Kory, "A Review and Meta-Analysis of Multimodal Affect Detection Systems", *ACM Computing Surveys (CSUR)*, Vol. 47, No. 3, p. Article 43, 2015.

120. Rime, B., B. Mesquita, S. Boca and P. Philippot, "Beyond the emotional event: Six studies on the social sharing of emotion", *Cognition & Emotion*, Vol. 5, No. 5-6, pp. 435–465, 1991.

121. Herring, S. C., "Gender differences in CMC: Findings and implications", *Computer Professionals for Social Responsibility Journal*, Vol. 18, No. 1, 2000.

APPENDIX A: ANNOTATION DESIGN

Figure A.1. Information page of valence annotation

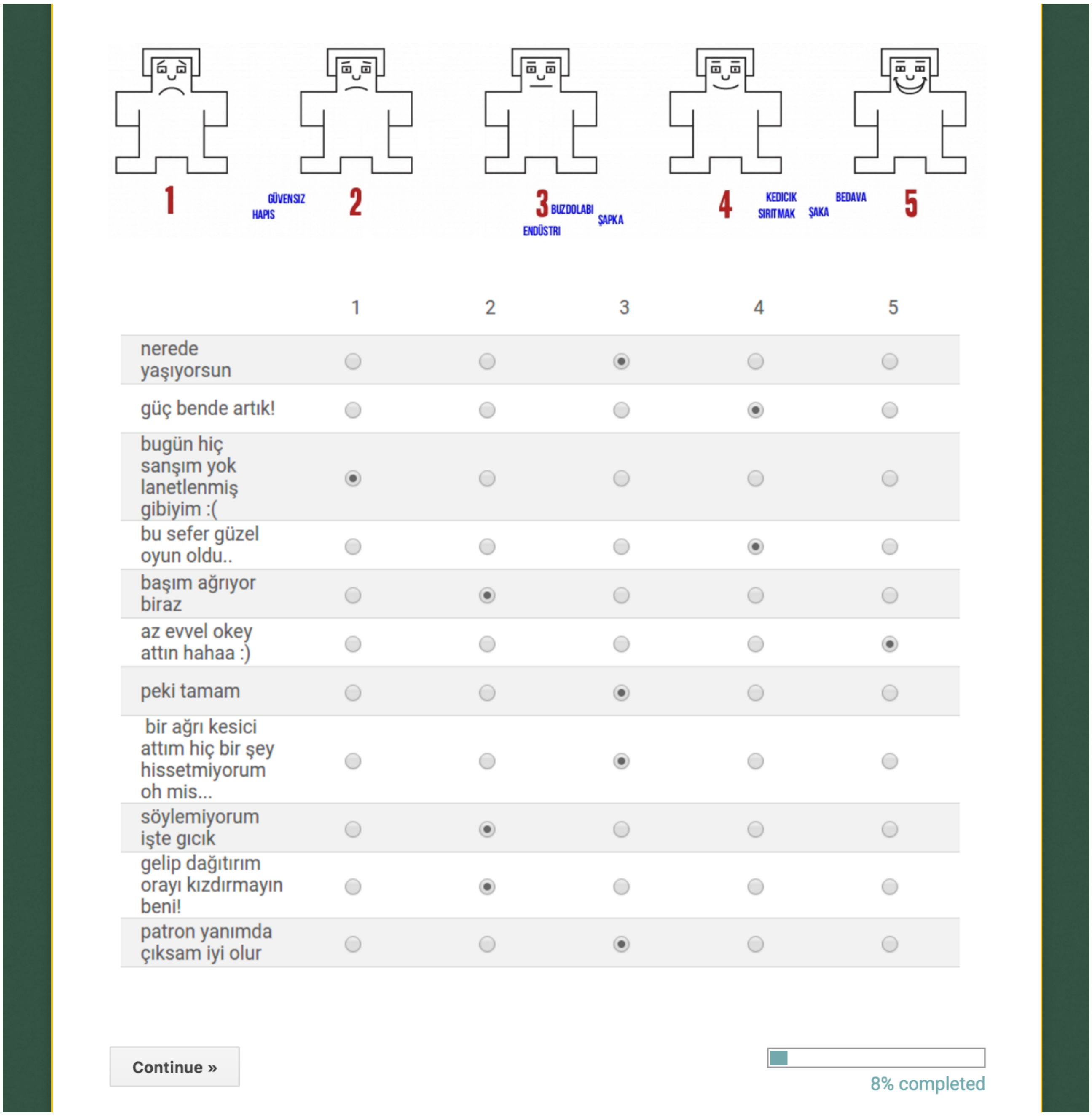
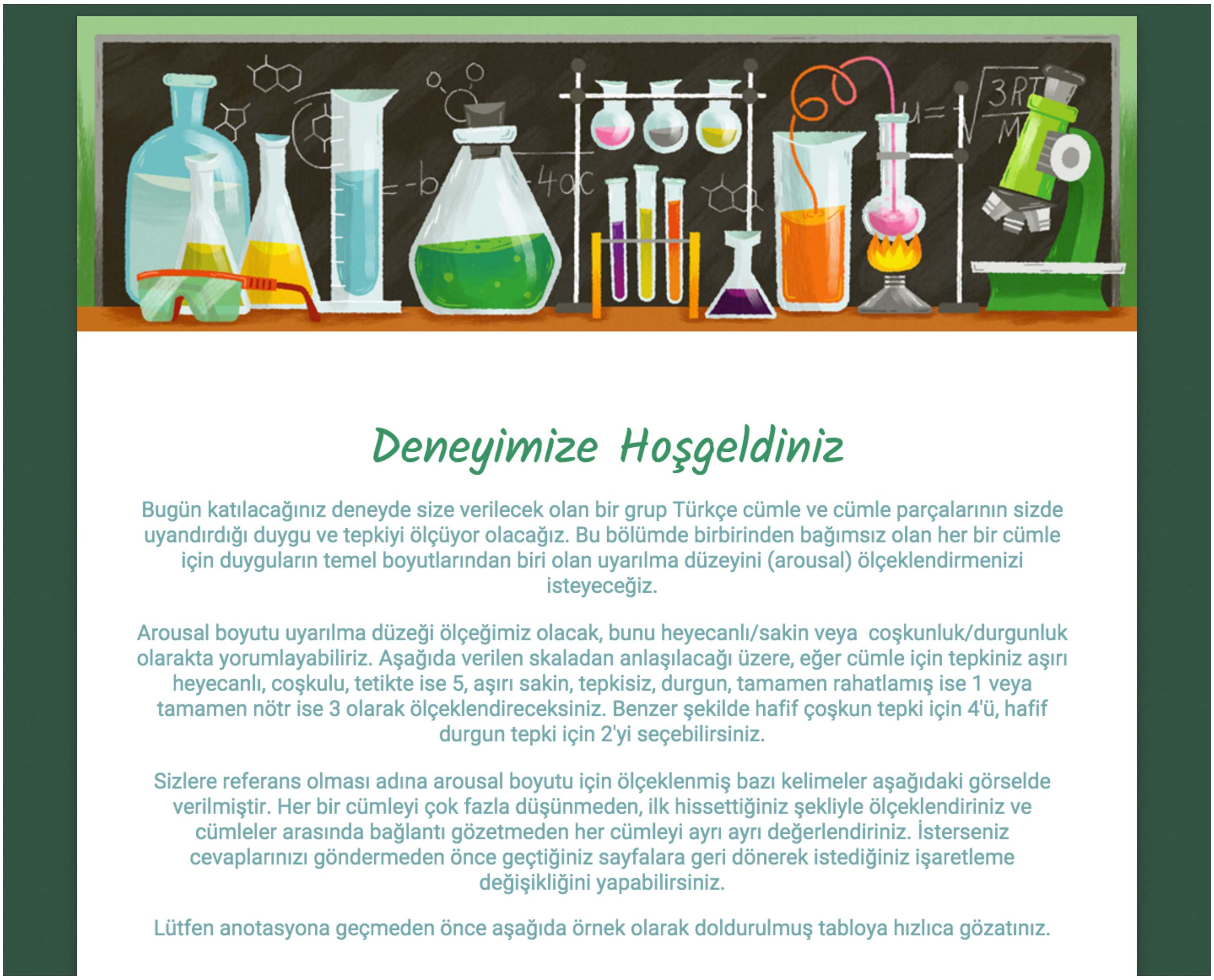


Figure A.2. Pre-filled reference annotation for valence dimension



Deneyimize Hoşgeldiniz

Bugün katılacağınız deneyde size verilecek olan bir grup Türkçe cümle ve cümle parçalarının sizde uyandırdığı duygu ve tepkiyi ölçüyor olacağız. Bu bölümde birbirinden bağımsız olan her bir cümle için duyguların temel boyutlarından biri olan uyarılma düzeyini (arousal) ölçeklendirmenizi isteyeceğiz.

Arousal boyutu uyarılma düzeği ölçeğimiz olacak, bunu heyecanlı/sakin veya coşkunluk/durgunluk olarakta yorumlayabiliriz. Aşağıda verilen skaladan anlaşılacağı üzere, eğer cümle için tepkiniz aşırı heyecanlı, coşkulu, tetikte ise 5, aşırı sakin, tepkisiz, durgun, tamamen rahatlamış ise 1 veya tamamen nötr ise 3 olarak ölçeklendireceksiniz. Benzer şekilde hafif çoşkun tepki için 4'ü, hafif durgun tepki için 2'yi seçebilirsiniz.

Sizlere referans olması adına arousal boyutu için ölçeklenmiş bazı kelimeler aşağıdaki görselde verilmiştir. Her bir cümleyi çok fazla düşünmeden, ilk hissettiğiniz şekliyle ölçeklendiriniz ve cümleler arasında bağlantı gözetmeden her cümleyi ayrı ayrı değerlendiriniz. İsterseniz cevaplarınızı göndermeden önce geçtiğiniz sayfalara geri dönerek istediğiniz işaretleme değişikliğini yapabilirsiniz.

Lütfen anotasyona geçmeden önce aşağıda örnek olarak doldurulmuş tabloya hızlıca gözatınız.

Figure A.3. Information page of arousal annotation

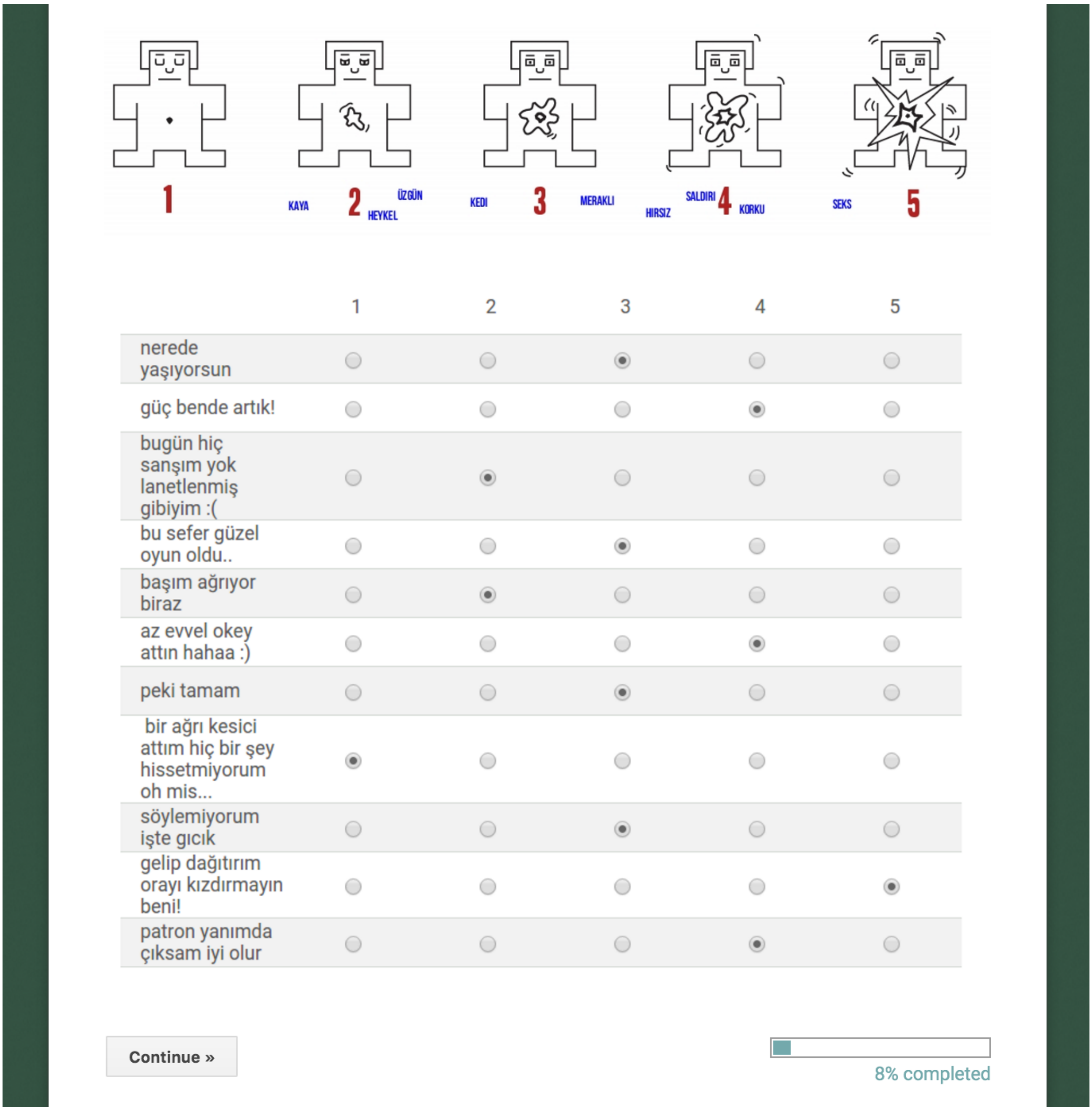
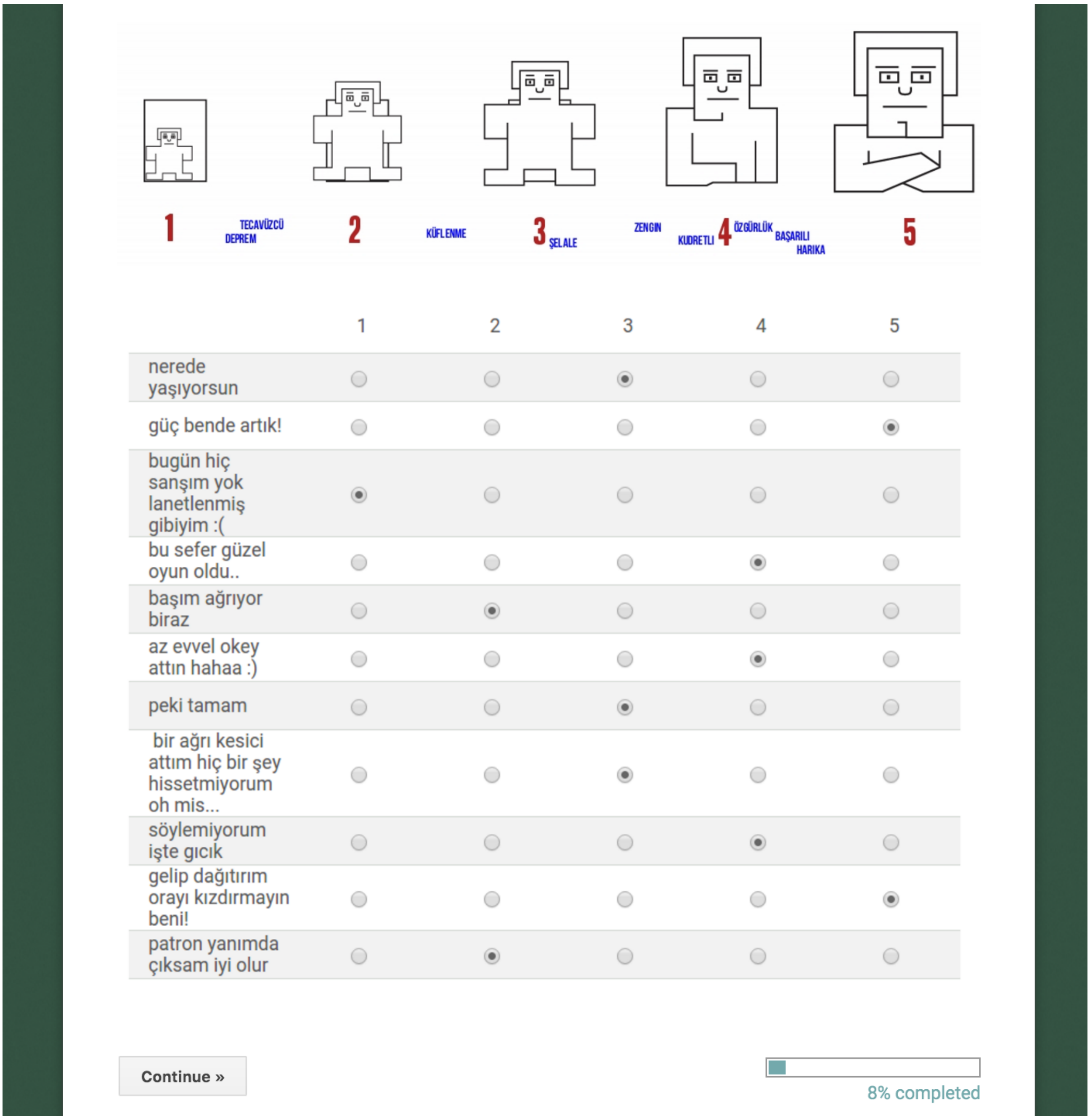


Figure A.4. Pre-filled reference annotation for arousal dimension

Figure A.5. Information page of dominance annotation



	1	2	3	4	5
nerede yaşıyorsun	○	○	●	○	○
güç bende artık!	○	○	○	○	●
bugün hiç sanşım yok lanetlenmiş gibiyim :(	●	○	○	○	○
bu sefer güzel oyun oldu..	○	○	○	●	○
başım ağrıyor biraz	○	●	○	○	○
az evvel okey attın hahaa :)	○	○	○	●	○
peki tamam	○	○	●	○	○
bir ağrı kesici attım hiç bir şey hissetmiyorum oh mis...	○	○	●	○	○
söylemiyorum işte gıcık	○	○	○	●	○
gelip dağıtırım orayı kızdırmayın beni!	○	○	○	○	●
patron yanımda çıksam iyi olur	○	●	○	○	○

Figure A.6. Pre-filled reference annotation for dominance dimension

	1	2	3	4	5
başarılarının devamını dilerim canı gönülden	○	○	○	○	○
şeytanın bol kılıcın keskin olsun	○	○	○	○	○
hayırlı sabahlar hoşgeldiniz gününüz aydın şansınız bol olsun	○	○	○	○	○
işlere taş gelmedi çok şanslısınız	○	○	○	○	○
kısmet degilmiş :)	○	○	○	○	○
tüh	○	○	○	○	○
üzülme ya	○	○	○	○	○
sormaz olaydım	○	○	○	○	○
evet :(	○	○	○	○	○
sen ne okuyosun	○	○	○	○	○

« Back   Continue »

16% completed

Figure A.7. Example sentences

APPENDIX B: GROUND TRUTH

Table B.1. A sample set of average ground truth annotation scores

SENTENCES	Valence	Arousal	Dominance
[beni karıştırdınız galiba]	2.67	3.17	2.17
[sağol Ali]	4.00	2.75	3.20
[sağol Aliiii]	4.50	4.20	3.75
[sen anlat biraz]	3.40	2.60	2.25
[arkadaşlar kendime kahve yapmıştım sizlere de afiyet olsun]	3.80	2.33	3.50
[bilgisayar takıldı hazıra basamadım kalktım]	2.50	2.40	2.33
[biraz kavga edelim diye yazdım :)]	3.83	3.75	4.40
[küstün mü]	1.83	2.75	2.25
[kumarda kaybeden aşkta kazanır kardeş]	3.67	2.33	3.33
[bazen yazılarımda gitmiyor çıldıracağım]	1.17	3.75	2.67
[yok komşu sen değil]	3.00	2.83	3.20
[zaten arkadaşız ne oldu ki]	2.83	3.20	3.00
[konu bu canım]	2.60	2.60	3.20
[gitmen lazımsa git]	2.67	2.67	3.60
[arkadaşlar bazıları gelip buraya terbiyesizlik yapıyor]	2.00	3.83	3.50
[ben senin gibi acizlerle muhatap olamıyorum]	2.00	3.75	4.00
[afedersiniz bekletmek istemezdim]	2.33	2.33	2.00
[taşla taşla kafasını yarmasın]	2.67	3.67	3.25
[kırgınlık bitsin :)]	3.50	3.33	3.50
[ben epey gurur duydum]	4.33	3.43	4.50
[yenmekte var yenilmekte]	2.83	2.50	3.25
[eski sisteme göre daha hoş olmuş]	3.83	2.80	3.40
[çok istiyorum ama olmuyor ona hep özlem duyuyorum]	2.00	2.50	1.25
[oley!]	4.83	4.33	4.40
[ayakkabı alcağım para yok]	1.60	2.33	2.40
[kızmak yok arkadaşlar burda eğleniyoruz sadece]	2.60	3.33	3.00
[iyi akşamlar]	3.33	2.83	3.20
[aşırı gerginim]	1.33	4.33	2.00

APPENDIX C: LEXICAL RESOURCES

Emoticon	Score	Emoticon	Score	Emoticon	Score	Emoticon	Score
=(	-1	8)	1	:p	1	(^,^)	1
=)	1	(8	1	:s	-1	(^-^)	1
=[	-1	8c	-1	s:	-1	(^.^)	1
=]	1	8-)	1	:S	-1	(^_^)	1
/=(	-1	8/	-1	:|	0	(^o^)	1
|=)	1	8\	-1	;^)	1	^_^	1
(-:	1	:#	-1	:^)	1	^-^	1
(:	1	:*(	-1	;o)	0	^.^	1
:'(	-1	:,(	-1	<3	1	^_^	1
:'-(	-1	:@	-1	<B	1	^o^	1
:(	-1	:-(	-1	<:}	0	^y^	1
:)	1	)-:	-1	>/	-1	:X	1
:Þ	1	(-:	1	>:(	-1	X(	-1
;)	1	:-)	1	>:)	1	X-(	-1
(;	1	:-*	1	>=D	1	XD	1
:o	-1	:-/	-1	}:)	1	OX	-1
:O	-1	:-D	1	{:)	1	XO	-1
:0	-1	:-O	0	>:D	1	XP	-1
:P	1	:-P	1	>:O	-1	xD	1
:p	1	:-S	-1	>[	-1	xP	-1
(o:	1	:-\	-1	>\	-1	|D	1
(o;	0	:-|	-1	B(	-1	|)	1
)-:	-1	:-}	1	B-P	1	:E	-1
):	-1	:/	-1	B)	1	:F	-1
)o:	-1	:3	1	B-S	-1	S:	-1
*)	0	:9	1	B-(	-1	:[	-1
-^--@	1	:D	1	B-)	1	[:	1
@}->-	1	:d	1	Bc	-1	:\	-1
0:)	1	:_(	-1	D:	-1	:]	1
:l	0	:o(	-1	:o)	1	]:	-1

Figure C.1. List of emoticons and their corresponding sentiment scores

Table C.1. A sample set of modifiers and their scores

	VALENCE		AROUSAL		DOMINANCE	
WORD	POS context	NEG context	POS context	NEG context	POS context	NEG context
tamamen	0.25	-0.3	-0.95	-2.04	-0.13	-0.17
bütünüyle	0.25	-0.3	-0.95	-2.04	-0.13	-0.17
kesinlikle	0.18	-0.16	0.25	0.07	-0.29	0.75
mutlaka	0.18	-0.16	0.25	0.07	-0.29	0.75
tabi	0.33	0.33	0.93	-0.55	-0.2	-0.2
tabiki	0.33	0.33	0.93	-0.55	-0.2	-0.2
özellikle	1.04	-0.69	1.25	-0.42	0.04	0.72
mahsus	1.04	-0.69	1.25	-0.42	0.04	0.72
aşırı	-0.8	-0.69	0.2	1.48	0	0.05
inanılmaz	0.91	-0.83	0.5	1.71	-0.19	0.43
olağanüstü	0.91	-0.83	0.5	1.71	-0.19	0.43
sadece	0.9	-1.17	0.17	1.4	-0.36	0.09
yalnızca	0.9	-1.17	0.17	1.4	-0.36	0.09
çok	0.6	-0.8	0	0.1	0.14	-0.8
epey	0.6	-0.8	0	0.1	0.14	-0.8
hayli	0.6	-0.8	0	0.1	0.14	-0.8
bayağı	0.6	-0.8	0	0.1	0.14	-0.8
gerçekten	0.2	-0.65	-0.3	0.05	-0.11	0.11
hakikaten	0.2	-0.65	-0.3	0.05	-0.11	0.11
sahiden	0.2	-0.65	-0.3	0.05	-0.11	0.11
bazı	0.5	-1.44	0.15	-0.08	-1	0.12
birtakım	0.5	-1.44	0.15	-0.08	-1	0.12
hafifçe	-0.32	-0.17	-0.3	-0.07	-0.84	0.25
az	-0.32	-0.17	-0.3	-0.07	-0.84	0.25
biraz	-0.32	-0.17	-0.3	-0.07	-0.84	0.25
aksine	0.33	0.61	0.17	0.27	-0.49	0.47
bilakis	0.33	0.61	0.17	0.27	-0.49	0.47
tersine	0.33	0.61	0.17	0.27	-0.49	0.47
fazlasıyla	0.6	-0.5	-0.25	0.17	0.15	0.67
pek	0.6	-0.5	-0.25	0.17	0.15	0.67
maşallah	-0.28	0	-0.25	0	0.44	0
asla	0	-0.11	0	-0.17	0	0.69
katiyen	0	-0.11	0	-0.17	0	0.69

Table C.2. A sample set of interjections and their scores

	VALENCE		**AROUSAL**		**DOMINANCE**	
WORD	POS context	NEG context	POS context	NEG context	POS context	NEG context
aa	0.2	-0.1	0.75	-1.56	0	0.08
oo	0.2	-0.1	0.75	-1.56	0	0.08
ah	0.14	-0.97	0.545	0.55	0.145	0.1
oh	0.14	-0.97	0.545	0.55	0.145	0.1
aman	-0.14	-0.97	0.545	0.55	0.145	0.1
sakın	-0.14	-0.97	0.545	0.55	0.145	0.1
ay	0.14	-0.97	0.545	0.55	0.145	0.1
oy	0.14	-0.97	0.545	0.55	0.145	0.1
vay	0.14	-0.97	0.545	0.55	0.145	0.1
bravo	0.63	0	1.2	0	-0.11	0
eyvah	-1.55	-1.55	0	-0.8	0	-0.39
tüh	-1.55	-1.55	0	-0.8	0	-0.39
hoppa	0	0.05	0	0.17	0	0.63
hayda	0	0.05	0	0.17	0	0.63
of	-0.02	-1.59	0.27	-0.17	-0.36	-0.07
üf	-0.02	-1.59	0.27	-0.17	-0.36	-0.07
ya	-0.3	-0.78	0	-0.25	0	-0.6
be	-0.3	-0.78	0	-0.25	0	-0.6
vah	-0.29	-0.29	0	0.5	0	0.69
yo	0.16	-1.01	-0.5	1	-0.61	1.04
yoo	0.16	-1.01	-0.5	1	-0.61	1.04
hah	0	0.44	0	-0.6	0.33	0.88
yuh	-0.86	-0.86	0	0.89	0	0.43
oha	-0.86	-0.86	0	0.89	0	0.43
lan	-0.86	-0.86	0	0.89	0	0.43
bre	0.5	0	1.02	0	0.44	0
abe	0.5	0	1.02	0	0.44	0
hey	0.42	-0.105	1.13	1.035	0.29	0.45
alo	0.42	-0.105	1.13	1.035	0.29	0.45
hoop	0	-0.14	0	0.5	0	0.85
yahu	-0.11	0	-1	0	0.15	0
hımm	-0.69	-0.36	-0.33	-0.67	-0.92	-0.05