Automatic Assessment of Dimensional Affective Content in Turkish Multi-party Chat Messages

Eda Aydın Oktay
Boğaziçi University
Computational Science and Engineering
Istanbul, Turkey
eda.aydin@boun.edu.tr

Koray Balcı
Boğaziçi University
Department of Computer Engineering
Istanbul, Turkey
balci@cmpe.boun.edu.tr

Albert Ali Salah
Boğaziçi University
Department of Computer Engineering
Istanbul, Turkey
salah@boun.edu.tr

ABSTRACT

This study presents a model for affective text analysis of online multi-party chat records in Turkish language. Online chats have challenges like non-standard word usage, grammatical irregularities, abbreviation usage, and spelling mistakes. We propose several pre-processing steps to deal with these. We adapt an affective word dictionary from English to Turkish, and by expanding it, obtain 15,222 words with annotations for valence, arousal, and dominance. We also employ a list of abbreviations, emoticons, interjections, modifiers (intensifiers and diminishers), and other linguistic indicators to capture the overall affective state at the sentence level. Lastly, we recruit and train annotators to obtain affective ground truth, and assess the accuracy of the proposed rule-based approach on a multi-party chat database collected from an online gaming environment.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems

Keywords

Chat; Affective computing; Natural language processing; Turkish; Computer games

1. INTRODUCTION

Computer mediated textual exchange has become an important channel of communication, whether it is performed through a stand-alone service, or as a part of a larger social ecosystem [10]. Particularly relevant are the online games, which typically integrate a chat platform for their users, whose numbers can easily reach millions. In this paper, we look at the problem of assessing affect in multi-party chat by automatic analysis of the chat text. Most languages do not have the rich text analysis tools developed for English language; we illustrate in this paper how some of the tools originally developed for English can be adapted to a different language (i.e. Turkish), and assess the accuracy of the ensuing system. The proposed approach could be used as a stand alone module, or for complementing multimodal approaches to analyze computer mediated communication.

Compared to other forms of affective content, text is believed to be relatively less challenging to analyze. However, unlike well-structured documents such as newspapers, books, and blogs, chat conversations are much more chaotic, and full of irregularities. On the other hand, they are also more loaded with affective content. Consequently, analysis of chat text for affect is far from trivial, but potentially rewarding.

We present here an affective analysis model for the Turkish language. The system was specifically designed for and tested with in-game multiparty chat logs, where the language is typically very flexible, irregular and emotive. We have collected and cleaned a large amount of chat data, and annotated them with syntactic and lexical semantic information. The second step was the adaptation of an affective lexicon from English to Turkish, and the creation of automated tools to analyze the chat corpus. Finally, we created an affect analysis model, which uses 120 emoticons, 98 abbreviations, 50 interjections, 72 modifiers, and the affective lexicon that includes more than 15,000 words. This model was used to cope with abbreviated and informal chat language and to capture a wide range of affective features.

The paper is organized as follows: Section 2 gives a brief explanation about the studies that have been proposed to investigate affective states and processes in online chat. Section 3 discusses the corpus, the annotation procedure, the assessment of annotation quality, and the creation of the affective lexicon for Turkish. Section 4 presents the affect analysis model. Section 5 describes the evaluation of the model, the results, and the limitations of the model. Section 6 concludes the paper.

2. RELATED WORK

Many researchers designed and implemented applications and intelligent user interfaces (such as email composers, online chat interfaces or instant messaging systems) with rich integrated emotion conveying engines. Sánchez et al. designed an instant messaging system called *Russkman* [20]. This system is 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 [25]. *Crys-*

ERM4CT'15, November 09-13, 2015, Seattle, WA, USA
 ISBN 978-1-4503-3988-9/15/11…$15.00
DOI: http://dx.doi.org/10.1145/2829966.2829972

talChat visualizes a user's social network by extracting user's chat log history and with the help of a graphical interface [23]. It presents patterns of 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 [30].

These applications allow users to express emotions, but there are other studies that sense affect automatically, for instance by including a tactile emotional interface for instant messenger chat [22], or by using physiological data and animated text with the help of a physiological sensor attached to the body [27]. In the absence of multimodal input, just the text is analyzed to predict emotions [11].

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 [19].

The most straightforward approach is **keyword spotting**, which identifies a set of keywords to construct a look-up table that contains keywords and their affective values. The basic limitations of this approach is that it is incapable of dealing with negation and with complex sentence structure. Building a rich lexicon is a very expensive task, and affect-conveying words only form a small portion of a sentence. Ma et al. proposed an emotion estimation system for chat or other dialogue domains based on keyword spotting with sentence-level processing [15]. In a more recent approach, Dey et al. proposed a rule-based model for emotion extraction in a real-time chat messenger based on a lexicon [8]. Another powerful example of rule-based systems is the Affect Analysis Model, which analyzes affect specifically in informal online communication media using symbolic cue analysis, syntactic structure analysis, word-level, phrase-level, and sentence-level analysis [16].

Lexical affinity uses the mutual information of words based on their relationships in the document. 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.

Statistical NLP and **learning-based** approaches are popular, and they basically rely on automatic calculation of frequencies of some seed words, their co-occurrences, punctuations, abbreviations, and sometimes synonym and acronym information. Brooks et al. presented an automated affect classification system for chat logs exploiting NLP and machine learning techniques [5]. This system segments the chat data and uses an improved bag-of-words model (including non-verbal cues) to classify text into thirteen affect categories.

The **commonsense-based approach** was first proposed by Liu et al. for emotion classification [14]. They used three real-world commonsense databases, including David Lenat's famous Cyc [13]. Compared to other approaches, this model works robustly on the sentence level. In this approach, a number of emotion models corresponding to each emotion class compete with each other and the winning models are used to identify the affect label of the text segments.

The affect recognition literature from Turkish texts is quite sparse. Turkish is an agglutinative language, where words can take many suffixes that modify the meaning. Cakmak et al. analyzed emotions attributed to Turkish word roots and sentences, and found significant correlation between them based on valence, activation and dominance [6]. In their approach, they have used 197 Turkish emotion words. Boynukalın et al. analyzed emotion in Turkish texts by using machine learning methods [2]. 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 [29]. Vural et al. proposed a 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 [26]. Dehkharghani et al. [7] presented SentiTurkNet, which is the first comprehensive Turkish polarity lexicon and included positivity, negativity, and objectivity scores assigned to each synset in the Turkish WordNet (about 15,000 synsets).

Our study was conducted on the most extensive social media text corpus in Turkish, and it targets continuous and dimensional affect prediction on valence, arousal, dominance (VAD) dimensions.

3. DATA AND ANNOTATION SCHEME

We develop our approach for the online chat domain, which typically involves limited vocabulary, grammatical irregularities, and chat-specific expressions, emoticons, and abbreviations. 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 with 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 [1].

For manual annotation, we randomly selected about 1000 sentences from the 4 million sentences in the database. We selected the most expressive ones by excluding sentence fragments and meaningless utterances, obtaining 300 independent sentences (both emotive and non-emotive). Note that 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. Sentence-level annotations were performed online and anonymously, by native Turkish speaker annotators from different backgrounds and ages. Each annotator was given a set of 100 sentences and asked to complete the survey for one dimension at a time after reading a single page of instructions and sample questions. For initial training, we also provided a set of tagged words drawn from the Affective Norms for English Words (ANEW) corpus [4].

Annotators were instructed to evaluate each sentence on a 5-point Likert scale, using their first intuition about the sentence. For valence; a value of 5 indicated extremely happy, satisfied, hopeful, pleasant, and 1 indicated completely unhappy, dissatisfied or bored. 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, weak, or influenced. For all dimensions, 3 was selected if they felt neutral.

In order to clarify the emotional dimensions and easily assess affective stimulus, we provided several annotated sentences as a reference and incorporated the Self-Assessment Manikins (SAM) [3] when designing the annotation scheme. Manikins were presented at the top of the sentences.

In order to perform a quantitative measurement for the effect of each modifier, we designed systematical annotation procedure for 72 intensifiers & diminishers, and 50 interjections that can shift the affective load[1]. This forms a second set of annotations, which we refer to as "modifier annotations". Rather than annotating modifiers at the word level, we assessed the effect of each modifier in context, considering that they may have different impact depending on the emotional polarity of sentence. Therefore, we created two sets of 73 sentences. These two sets included the same wording and the only difference between corresponding sentences was the presence or absence of a modifier.

In all, we collected 6,300 ratings for sentence annotations and 7,575 ratings for modifier annotations. 116 participants completed annotation for one or more dimensions. The ground truth annotations for both sets were obtained by averaging the annotations for each sentence.

3.1 Inter-annotator Reliability

Each sentence annotation was performed by seven different annotators with various educational and socioeconomic backgrounds in order to capture a broad consensus on affective judgment, which can be highly subjective. Then, we examined the inter-annotator agreement among all annotators.

For the annotation of modifiers, we repeated the labeling for 15 times on average with different participants. Repeated labeling is especially useful for noisy labels [21]. For both set of annotations, the values that were more than one standard deviation to the mean were treated as outliers and eliminated, and the means were recomputed.

We report the inter-annotator reliability for sentence and modifier annotations with Krippendorff's Alpha [12]. Because sentences are rated on a scale in our annotation, we preferred to use Krippendorff's interval coefficient for reliability test, 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.

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

Table 1: Inter-annotator agreement for sentence annotation

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

Table 2: Inter-annotator agreement for modifier annotation

We observe that the inter-annotator agreement is substantial in valence dimension, but not as high for arousal and dominance. Valence is more obvious and easier to annotate, whereas arousal is the most difficult.

3.2 Affective Lexicon

There is no comprehensive and widely-used lexicon of affective words with VAD annotation for Turkish. Since it is very costly and time consuming to construct a dictionary from scratch, we have automatically translated a dictionary of English lemmas. The study of Warriner et al. [28] is based on the ANEW norms proposed by Bradley and Lang for 1,034 words [4]. They have rated 13,915 English lemmas in a nine point scale (1-9) and provided mean values and standard deviations for valence, arousal, and dominance scores. A total of 1,827 participants (through Mechanical Turk) contributed to their study.

We linearly transformed these affect scores to a five point scale [1-5]. For translation, we initially used the Google Translation API. Two independent human translators manually checked each word, and corrected missing words and mis-translations. Some culture and language dependent words are omitted and the affective lexicon is finally expanded with synsets from TDK (Turkish Language Organization) dictionary[2]. As a result, we obtained a comprehensive affective lexicon for Turkish that includes valence, arousal and dominance values of 15,222 words and phrases. While this resource is not entirely reliable and well-formed, our assessment shows that it is useful. Any future work on a proper Turkish affective lexicon would improve the system that we propose.

Turkish	**English**	**POS**	**Val.**	**Aro.**	**Dom.**
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

Table 3: Some example words and phrases from our affective lexicon

[1] All resources are available online. https://github.com/verdeerosso/affective-turkish

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

4. MODEL OF AFFECT ANALYSIS

4.1 Preproccessing

We performed a three-step pre-processing method to normalize the informal characteristics of the chat messages. Before normalizing the noisy chat texts, the input sentences were segmented into words and each token was kept with part-of-speech (POS) information. We recorded the intentional spelling mistakes, duplications (e.g. 'selaaaam'), uppercase usage (e.g. 'HA-DII') or exclamation mark usage (e.g. '!!!'), since they serve as useful features, especially for high arousal patterns.

Secondly, we checked this list of corrected words with a direct look-up. Because there are a lot of repetitive expressions and repetitive words in chat domain, spelling mistakes are usually repeated too. Therefore, keeping a list of the most frequent normalizations is very effective to correct chat-specific spelling mistakes. Lastly, we used a Turkish normalization tool proposed by Torunoglu and Eryigit [24]. This tool is available online.[3]

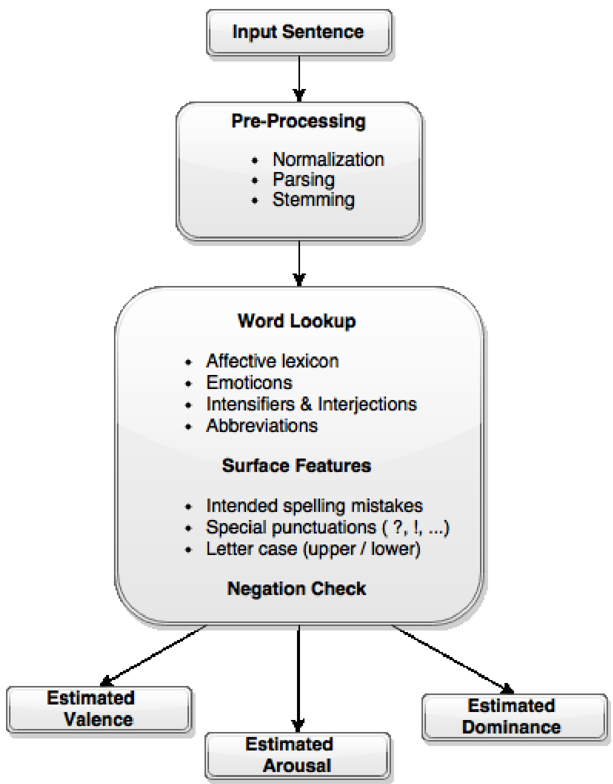

Figure 1: The Model of Affect Analysis.

4.2 Set of Features and Rules

People often use **emoticons** to enhance or underline the meaning of certain text elements.In the simplest form of emoticon exchange, special character combinations are used within the context of online chat to display affective mood. Frequently used examples include ':o' as 'surprised', ':(' as 'sad', and ';)' as winking. We constructed an emoticon table consisting of 120 popular emoticons used in Turkish multi-party chats.

[3]http://tools.nlp.itu.edu.tr/Normalization

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. The final modifier list includes 72 intensifiers and diminishers. We also created a list of 50 interjections. Some example modifiers and interjections given in Table 4 with their corresponding VAD scores.

Since Turkish is a morphologically rich agglutinative language, one can generate hundreds of legitimate words from a single root with derivational and inflectional morphemes. Subsequently, even word-level analysis is challenging [17]. On the other hand, there is a strong link between word roots and the perceived emotion of the sentence [6]. Therefore, we take the **word roots** into consideration when we create our lexicon and for analyzing the affect in sentences. Assuming that affective load of a sentence is almost the same for different person forms and tenses, we do not perform a detailed morphological analysis for these. Morphological operations that we do take into account include vowel harmony, such as consonant changes (e.g. ayak - ayağı), vowel drops (e.g. burun-burnum) [18], removing the infinitive suffix '-mek', '-mak', and most importantly, detection of negation that comes with the negation suffix '-me', '-ma'. We also checked other negation markers of Turkish, such as 'değil' (not), 'yok' (there is not), 'ne...ne' (neither .. nor), '-siz', '-sız', '-suz', 'süz' (without).

These features are prioritized in our system, which first checks for the existence of any emoticons, then looks for modifiers, surface features, negations, and finally, the corresponding VAD scores for each word unit.

Based on these features overall affective value is extrapolated as follows:

- 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. These coefficients are obtained from the modifier annotations.
- If there is a verb and a noun phrase with opposite scores, the verb is considered as dominant.
- If negation is detected, the VAD score is reversed by subtracting from 6 (e.g. 2.3 turns into 3.7). However, the 'ne ...ne' connector neutralizes the affective value of the sentence to 3.0 (e.g. "Sabah hava ne iyiydi ne de kötü." - "This morning the weather was neither good, nor bad.")
- 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.
- If there are no affect-carrying words, emoticons, or interjections, the sentence is considered as neutral.

The quantitative effects of a modifier on the valence, arousal and dominance were calculated by comparing the mean annotation scores with the presence or absence of each modifier. Each modifier was tested both in a positive context and a negative context, as we initially assumed (and then showed) that some modifiers might intensify the valence when used in a sentence with positive polarity and diminish the valence in sentences with negative polarity (see Table 4

		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	increadible	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

Table 4: Example modifiers, interjections and corresponding scores in positive and negative contexts.

for examples). Therefore, when there exists a modifier, our model first determines the polarity (positive or negative) of a sentence, and then updates the affective scores based on the polarity assignment. We followed the same process for annotations of interjections and some other surface features, such as the effect of use of upper case typing, or intended spelling mistakes (e.g. 'helloooo') to boost the arousal score.

5. EVALUATION OF THE MODEL

Based on our annotated affective data, we report fine-grained (dimensional) and coarse-grained (sentiment level) evaluation results with different metrics. For dimensional evaluation, model scores were scaled continuously between [1-5]. The results were reported for all three dimensions in Table 5, in terms of mean squared error and accuracy. To compute fine-grained accuracy, we calculated 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.

Measure	Valence	Arousal	Dominance
All features (MSE)	0.62	0.58	0.47
All features (Acc)	%50	%54.2	%57

Table 5: The accuracy of the model for dimensional affect estimation.

As can be seen from Table 5, the mean squared error is lower for dominance and arousal compared to valence. However, the annotation score ranges are different for these three dimensions: Valence: 2.85($\pm$0.86); Arousal 3.06($\pm$0.92); Dominance 3.12($\pm$0.69). While the accuracy is lowest for valence, correlation tells a different story. The correlation between predicted scores and the ground truth annotation is highest for valence (0.62), lowest for arousal (0.25). Lower spread for scores makes prediction more difficult, as it implies that the distinctions between sentences are not easy to discern, even for the human annotators.

Measure	**Accuracy (%)**
All features	70.4
All features except modifiers	57.2
All features except emoticons	68.4
All features except normalization	64.9
All features with SentiTurkNet	62.8

Table 6: The accuracy of the model for course-grained affect estimation.

Secondly, for coarse-grained evaluation, model predictions in valence dimension are mapped to positive (>3) and negative (< 3) classes. This is the sentiment analysis scenario. We obtain %70.2 accuracy with this approach when all features are employed. In order to evaluate performance of our affective lexicon comparatively, we also tested the model with SentiTurkNet polarity lexicon [7] and obtained %62.8 accuracy with this setup. There is obviously room for improvement in both types of scenarios.

6. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we described a rule-based system for dimensional affect analysis in Turkish multi-party chat environment. We also presented a comprehensive Affective dictionary in Turkish language with Valence, Arousal, and Dominance scores in the [1-5] scale. The number of tools and prior work in analyzing affect from text is very limited for the Turkish language, and such a corpus will be beneficial. Partly due to the lack of such resources, there are no comparative results reported in our study. The affective lexicon and the other resources created in this study are made publicly available.

The application domain of the study is multi-party online social games. As indicated in [9], 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). We have not used any machine learning approaches, but directly modeled frequency distributions of words and modifiers. For supervised machine learning approaches, more extensive annotation of sentences is necessary. The existing work can help annotation efforts by focusing them on parts of the data space where predictions are poor.

7. ACKNOWLEDGMENTS

This work was supported in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under grant number 114E481.

8. REFERENCES

[1] K. Balci and A. A. Salah. Automatic analysis and identification of verbal aggression and abusive behaviors for online social games. *Computers in Human Behavior*, 53:517 – 526, 2015.

[2] Z. Boynukalin and P. Karagoz. Emotion analysis on Turkish texts. In E. Gelenbe and R. Lent, editors,

Information Sciences and Systems, volume 264 of *LNEE*, pages 159–168. 2013.

[3] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[4] M. M. Bradley and P. J. Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, C-1, The Center for Research in Psychophysiology, Univ. of Florida, 1999.

[5] M. Brooks, 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. In *Proc. CSCW*, pages 317–328. ACM, 2013.

[6] O. Cakmak, 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.

[7] R. Dehkharghani, Y. Saygin, B. Yanikoglu, and K. Oflazer. Sentiturknet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, pages 1–19, 2015.

[8] L. Dey, M.-U. Asad, N. Afroz, and R. P. D. Nath. Emotion extraction from real time chat messenger. In *Proc. ICIEV*, pages 1–5. IEEE, 2014.

[9] S. K. D'Mello and J. Kory. A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)*, 47(3):Article 43, 2015.

[10] J. F. Grafsgaard, R. M. Fulton, K. E. Boyer, E. N. Wiebe, and J. C. Lester. Multimodal analysis of the implicit affective channel in computer-mediated textual communication. In *Proc. 14th ICMI*, pages 145–152. ACM, 2012.

[11] B. Kolz, J. M. Garrido, and Y. Laplaza. Automatic prediction of emotions from text in Spanish for expressive speech synthesis in the chat domain. *Procesamiento del Lenguaje Natural*, 52:61–68, 2014.

[12] K. Krippendorff. Computing Krippendorff's alpha reliability. *Departmental Papers (ASC)*, page 43, 2007.

[13] D. B. Lenat. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

[14] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proc. IUI*, pages 125–132. ACM, 2003.

[15] C. Ma, H. Prendinger, and M. Ishizuka. Emotion estimation and reasoning based on affective textual interaction. In *Affective computing and intelligent interaction*, pages 622–628. Springer, 2005.

[16] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Affect analysis model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*, 17(01):95–135, 2011.

[17] K. Oflazer. Two-level description of Turkish morphology. *Literary and linguistic computing*, 9(2):137–148, 1994.

[18] K. Oflazer, E. Göçmen, and C. Bozsahin. An outline of Turkish morphology. *Report on Turkish Natural Language Processing Initiative Project*, 1994.

[19] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[20] J. A. Sánchez, 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. In *Proceedings of VII Brazilian symposium on Human factors in computing systems*, pages 66–72. ACM, 2006.

[21] V. S. Sheng, F. Provost, and P. G. Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proc. 14th ACM SIGKDD*, pages 614–622, 2008.

[22] H. Shin, J. Lee, J. Park, Y. Kim, H. Oh, and T. Lee. A tactile emotional interface for instant messenger chat. In *Human Interface and the Management of Information. Interacting in Information Environments*, pages 166–175. Springer, 2007.

[23] A. Tat and S. Carpendale. CrystalChat: Visualizing personal chat history. In *Proc. 39th Annual Hawaii Int. Conf. on System Sciences*, volume 3, page 58c. IEEE, 2006.

[24] D. Torunoğlu and G. Eryiğit. A cascaded approach for social media text normalization of Turkish. In *5th Workshop on Language Analysis for Social Media (LASM) at EACL*, pages 62–70, Gothenburg, Sweden, April 2014. Association for Computational Linguistics.

[25] D. Tsetserukou, A. Neviarouskaya, H. Prendinger, N. Kawakami, M. Ishizuka, and S. Tachi. iFeel_IM! emotion enhancing garment for communication in affect sensitive instant messenger. In *Human Interface and the Management of Information. Designing Information Environments*, pages 628–637. 2009.

[26] A. G. Vural, 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. In *Computer and Information Sciences III*, pages 437–445. Springer, 2013.

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

[28] A. B. Warriner, V. Kuperman, and M. Brysbaert. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4):1191–1207, 2013.

[29] E. Yıldırım, F. S. Çetin, G. Eryiğit, and T. Temel. The impact of NLP on Turkish sentiment analysis. In *Proc. of the Int. Conf. on Turkic Language Processing*, pages 7–13, 2014.

[30] D. Zheng, F. Tian, J. Liu, Q. Zheng, and J. Qin. Emotion chat: A web chatroom with emotion regulation for e-learners. *Physics Procedia*, 25:763–770, 2012.