

Emotionality of Turkish language and primary adaptation of affective English norms for Turkish

Mastaneh Torkamani-Azar, Sumeyra Demir Kanik, Ayse Tuba Vardan, Cagla Aydin & Mujdat Cetin

Current Psychology

A Journal for Diverse Perspectives on Diverse Psychological Issues

ISSN 1046-1310

Curr Psychol

DOI 10.1007/s12144-018-0119-x



Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Emotionality of Turkish language and primary adaptation of affective English norms for Turkish

Mastaneh Torkamani-Azar¹ · Sumeyra Demir Kanik¹ · Ayse Tuba Vardan^{1,2} · Cagla Aydin³ · Mujdat Cetin^{1,4}

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Emotional load assessment of the written words has gained considerable interest in psycholinguistics, semantics, and analysis of psychophysiological and electrophysiological correlates of emotional processing. Considering the lack of a publicly available database with affective ratings of contemporary verbal stimuli obtained from native Turkish speakers, we present the affective norms for two datasets of Turkish words carefully adapted from the Affective Norms for English Words (ANEW) database. The valence and arousal ratings are obtained from 61 college-aged participants for 127 highly arousing, emotionally-loaded words in the Adapted Turkish Affective List (ATAL). The ATAL ratings show a tendency of classifying fewer words as positive compared to the original list of stimuli, significantly higher arousal levels for positively rated Turkish stimuli compared to the negative and neutral words, and more congruence in arousal levels of positively exciting words. For the medium to high arousing 508 words in the Expanded Turkish Affective List (ETAL) that cover the whole 9-point spectrum of the valence dimension, 136 Turkish respondents from a wider age, education, and occupation background show higher excitability towards highly unpleasant words. Strong cross-linguistic correlations of +0.968 between the valence ratings of ANEW and ATAL and +0.878 for ANEW and ETAL demonstrate the ease of transferring and perceiving the valence levels across English and Turkish. The medium correlation of roughly +0.450 between the English and Turkish arousal ratings account for lower excitation levels perceived by the native Turkish speakers and indicate the arousal dimension is similar to familiarity and originality in exhibiting more variations between different cultures. These findings demonstrate that this expanded database of partial affective normative ratings can be used as the ground truth for emotional and neurocognitive assessments, and that the presented methodology can be utilized for developing a comprehensive Turkish affective lexicon. The utilized word selection criteria also enable a cross-cultural analysis of adapted words in Turkish and other languages. Detailed normative ratings of this Turkish adaptation are included in the supplementary materials.

Keywords Emotions · Valence · Arousal · ANEW · Affective norms · Affective ratings · Turkish language · SAM · Linguistic adaptation

Introduction

Assessment of emotional responses is an important research field that helps with understanding how emotions are

stimulated and perceived, how they are related to the underlying cognitive activities and subsequent behaviors, and how they can be intensified during interactions with multimedia stimuli. In fact, recording and characterizing subjective responses to the emotional loads of words and images have interesting applications in cognition and language development in children and adults (Bloom 1998; Sylvester et al. 2016) as well as in diagnosis and prognosis of neurodevelopmental disorders such as attention-deficit/hyperactivity disorder (ADHD) (Passarotti et al. 2010; Nigg and Casey 2005) and schizophrenia (Patrick et al. 2015). Recognition and assessment of mood states are also used in behavioral therapy and analysis (Lang 1980; Bailey and Chapman 2012; Trimmer et al. 2013; Wang et al. 2014), psycholinguistics and speech processing

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12144-018-0119-x>) contains supplementary material, which is available to authorized users.

- ✉ Mastaneh Torkamani-Azar
mastaneh@sabanciuniv.edu
- ✉ Mujdat Cetin
mccetin@sabanciuniv.edu

Extended author information available on the last page of the article.

(Degner et al. 2012; Opitz and Degner 2012; Schacht and Sommer 2009; Hinojosa et al. 2010; Citron 2012; Paulmann et al. 2013), and studying cultural implications of language emotionality (Kaviani et al. 2015). Recent interest towards neuromarketing is also built on advances in affective computing (Nicolaou et al. 2011) and assessment of psychophysiological and electrophysiological correlates of preference, happiness, and stress (Pietro et al. 2014; Chanel et al. 2006; Luu and Chau 2008; Kroupi et al. 2011; Yılmaz et al. 2014), with applications such as likeability analysis of advertisements and political campaigns (Vecchiato et al. 2014a; Vecchiato et al. 2014b). These responses usually rely on normative ratings along a set of affective dimensions such as valence (pleasantness or unpleasantness), arousal (mental excitation or calmness), and dominance (being in control or controlled by the stimulus) that form three axes of a continuous, multidimensional affective space, also known as the core affect space (Trimmer et al. 2013; Warriner et al. 2013).

From the wide pool of emotional stimuli, words have been a popular choice due to their ease of use in experiments and their ability to represent abstract concepts that images are not able to demonstrate (Soares et al. 2012). A common tool for obtaining affective ratings for verbal and pictorial stimuli is a set of pictorial self-assessment manikins (SAMs) (Lang 1980; Bradley and Lang 1994). SAMs enable any individual to provide nonverbal, subjective answers for emotion and semantic questions. For example, in affective normative studies, SAMs are used for collecting subjective ratings from groups of native speakers of the target language while a variety of quantitative measures are utilized for assessing psycholinguistic variables such as familiarity, originality, concreteness, and offensiveness.

Affective Norms for English Words (ANEW) is a well-known and widely used database published by Bradley and Lang (1999) with a list of 1,034 common English nouns, adjectives, and verbs, and later updated to contain 2,476 words in its 2010 release. Warriner et al. (2013) extended these efforts and developed a database with affective ratings for 13,915 English lemmas across a wide range of word categories. There has also been a growing interest in adapting ANEW to other languages. Due to variations in the cultural, social, and historical factors, this database has had to be carefully adapted – i.e., translated and back-translated – to the destination languages and rated by their native speakers before the associated ratings could be used as the ground truths for emotional assessment of verbal and auditory messages in their respective languages.

Several projects in various countries have developed affective norms for Indo-European, Uralic, and Chinese languages. These projects mainly use one of the following three methods: a. Collecting ratings for translated ANEW words as the main resource, as performed for the Spanish

(Redondo et al. 2007), British English and Finnish (Eilola and Havelka 2010), European Portuguese (Soares et al. 2012), Italian (Montefinese et al. 2014), and Polish ANPW (Imbir 2015) adaptations, b. Obtaining affective assessments from scratch using original literary resources and corpora as conducted for the Berlin Affective Word List (BAWL) and its updated version BAWL-R (Vò et al. 2009), the French Emotional Evaluation List (FEEL) (Gilet et al. 2012), the Dutch list (Moors et al. 2013), the Chinese Lexicon Project (Sze et al. 2014), and the Persian affective list (Mokhlesin et al. 2015), or c. Adapting non-English databases to destination languages as done by Polish researchers for developing the Nencki Affective Word List (NAWL) from BAWL-R (Riegel et al. 2015), ANPW-R (Imbir 2016), and the Dutch list (Moors et al. 2013).

In the context of the Turkish language, more research has been done on semantic analysis, natural language processing, and other word norming applications such as imageability and concreteness than on affective computing (Göz et al. 2017). Nonetheless, psycholinguistic studies on differences between emotional perception in the first and second/foreign languages have provided invaluable insights into perception of emotionality in Turkish. As an important contribution, Harris et al. (2003) developed a small English database with 48 positive, neutral, and negative words selected from the Handbook of Semantic Word Norms (Toglia and Battig 1978), as well as nine taboo words and seven reprimands used by parents towards their children. Words in this database were either translated or adapted to Turkish and tested on Turkish graduate students and professionals living in an English-speaking country to compare their autonomic responses to taboo words in their first (L1) and second (L2) languages. These authors also collected the unpleasantness ratings, and conducted tests of recall and recognition for the same database (Ayçiçeği and Harris 2004). A similar approach was later used in a study of British English and Finnish in rating valence, arousal, offensiveness, and familiarity of 210 taboos and words adapted from ANEW (Eilola and Havelka 2010).

From the perspective of natural language processing, Çakmak et al. (2012) utilized fuzzy logic to analyze the emotional loads of Turkish words. They translated 197 English words to Turkish as the stimuli, and collected lower and higher values for subjective ratings of valence, activation, and dominance from each participant. This study provided an explanation for variations and intra-subject uncertainty in affective ratings of each word-scale. In contrast to their work, the current study focuses on collection of absolute ratings of the mentioned affective dimensions for comparing emotional perception across languages and for the side goal of developing a database whose ratings could be used as the ground truth for various neuropsychophysiological experiments.

Focusing on automatic polarity assessment of Turkish sentences acquired from a multi-party chat room data set, Aydın Oktay et al. (2015) selected subjective ratings for 300 independent emotive and non-emotive sentences. Subsequently, they collected over 13,000 ratings for sentence and modifier annotations using a 5-point scale for valence, arousal, and dominance dimensions. However, due to the difficulty of adapting a large affective English database for the native speakers of Turkish and the lack of a similarly comprehensive database of Turkish words, they translated 13,915 English lemmas of Warriner et al. (2013) to Turkish and directly transferred the 9-point scale ratings to a 5-point scale. They thus obtained an affective lexicon database with emotional ratings of 15,222 Turkish words and phrases while admittedly ignoring the role of cross-cultural variations in affect perception. This database was later utilized for assessing the affective states demonstrated and expressed by children in the psychodynamic play therapy sessions (Halfon et al. 2016).

Annotation of Turkish speech stimuli by native speakers was carried out by Ofıazođlu and Yıldırım (2013) who used acoustic features of 5,300 short utterances extracted from Turkish movies. Their work led to the TURKISH Emotional Speech database (TURES).¹ Finally, to date, the only effort on affective Turkish lexicons with a methodology similar to that of ANEW's developers has been implemented for TUDADEN (Gökçay and Smith 2008; Gökçay 2011). With mean ratings of arousal, valence, dominance, and concreteness collected for 1,240 words from 170 participants, TUDADEN is a useful database that considers the emotional reactions of native speakers to the verbal Turkish stimuli. However, it is not publicly available, and it is hard to use this database in comparative, cross-linguistic, or bilingualism studies as it is not clear whether the emotions represented by the Turkish words match identical emotions in the affective landscape of English speakers.

We noticed the inaccessibility of a controlled and publicly available database of affective ratings adapted to emotional perceptions of the Turkish population during an ongoing study on characterizing neural correlates of word processing.² This observation led us to develop a partial affective lexicon based on adapting 127 highly arousing words from ANEW 2010 to Turkish with ratings obtained from a group of college-aged, native Turkish speakers. Two main reasons motivated us to start adaptation by using

ANEW words with high arousal and most extreme valence ratings. First, arousal levels of emotions have been reported to exhibit cross-cultural differences. It has been shown that Chinese and Japanese participants associate happiness with low arousal, positive emotional states as demonstrated by the word "calm". This is while the Westerners conceptualize happiness with high arousal, positive emotional states as reflected by the word "excited", for instance (Lu and Gilmour 2004; Uchida and Kitayama 2009). Considering these dissimilarities, it might be the case that these conceptions induce differences in how rating scales are used. More importantly, these associations can shed light on which emotion is salient to the members of a particular culture. Therefore, in the first Study reported here, only the most extremely rated words from ANEW are used to cross check their rating patterns in Turkish. Second, it has been shown that valence and arousal may have differential effects on cognitive processes; a general pattern in this regard is that the valence effects are observed only when the arousal levels are high (Gomes et al. 2013). The word selection procedure utilized in Study 1 for the Adapted Turkish Affective List (ATAL) is designed to amplify this effect as well.

Later, in a subsequent study also reported in this paper, we decided to analyze whether the elimination of neutral stimuli affects the valence-arousal relationship and the way respondents evaluate the arousal levels of the emotionally loaded stimuli. Thus, we expanded the database to 508 carefully selected and back-translated Turkish words, and obtained subjective ratings for valence and arousal scales from a group of native Turkish speakers from diverse age and socioeconomic backgrounds. This adaptation and its accompanying emotional ratings are hereafter referred to as the Expanded Turkish Affective List (ETAL). The supplementary materials accompanying this article contain the utilized English words, the valence and arousal ratings for our ATAL and ETAL datasets, and information on their lexical categories and word frequencies obtained from a contemporary corpus of the Turkish language.

Our contributions are multifold, starting with a careful methodology for word selection and adaptation that has a great potential to be utilized for a full adaptation of English affective lists. Furthermore, realizing that the emotional loads of words and long texts depend on the language emotionality (Kloumann et al. 2012), age (Moors et al. 2013; Imbir 2016), and cultural and social stigma surrounding the concept and notion of each word (Grühn and Smith 2008; Gilet et al. 2012), affective ratings were obtained from a homogeneous group of Turkish students. Based on comprehensive statistical analyses, we observed a very high cross-linguistic correlation for the valence dimension of English and adapted Turkish words while the arousal levels exhibited higher variations across the two languages. Moreover, although the English affective

¹<http://www.turesdatabase.com>

²After completion of experiments and during the revision process, we were notified of a study on Turkish emotional word norms by Kapucu et al. (2018) who have obtained ratings for valence and arousal dimensions and five discrete emotions of 2,031 Turkish words. We compare our findings with theirs in the "Discussion" section.

databases demonstrate a bias towards higher number of positively rated words (Kloumann et al. 2012), the results obtained from ATAL and ETAL did not show such a bias in valence ratings. Finally, although positive words in the smaller dataset of ATAL were rated more arousing than negative words – similar to implications of the British English, Finnish, and Dutch affective word studies, the polarity completely changed towards high arousal norms for negative words when the number of stimuli was increased and neutral words were included in the ETAL. Thus, as the first partial but public adaptation of ANEW in Turkish, the present work provides an invaluable opportunity for psychologists, linguists, and the affective computing community to directly compare and contrast emotional perception between languages and different linguistic/cultural groups.

The rest of this paper is organized as follows. Section “Study 1” focuses on the acquisition of the initially adapted Turkish norms from 127 highly arousing, emotionally-loaded stimuli. Next, Section “Study 2” covers the development of the expanded Turkish affective list with 508 stimuli consisting of emotionally loaded and neutral words with medium to high arousal norms from the English database. The results sections of both studies include a thorough description of our findings and relationships between norms of valence and arousal dimensions in Turkish and English, congruence and reliability of collected ratings, and relationships among the linguistic-affective variables for the adapted Turkish words. Comparison of the obtained results with those from databases of other languages are presented under the “Discussion” section as are the explanations on changes in arousal perception and representation across different languages. Finally, the “Summary and Future Directions” section covers the implications of our results for future emotional studies as

well as the description of the accompanying supplementary materials.

Study 1: The 127-word Adapted Turkish Affective List

Material and Methods

Stimuli Selection

The *Original English List* used in our study consists of 128 words selected from the ANEW database (Bradley and Lang 2010). This database has 2,476 nouns, verbs, and adjectives rated by native speakers of American English in terms of subjectively perceived levels of valence, arousal, and dominance. In this study, however, we evaluate and analyze only the first two dimensions. The affective normative ratings of ANEW had been obtained non-verbally and through two sets of 9-point scale SAMs for pictorial assessment of affective reaction, as displayed in Fig. 1 (Bradley and Lang 1994). A rating of 1 in SAM’s valence scale corresponds to respondents feeling unhappy, annoyed, or despaired, while a rating of 9 reflects being very happy, satisfied, and hopeful. The arousal pictures symbolize how strong the emotional activations of the words are; a rating of 9 in this scale shows participants being highly aroused, excited, stimulated, or jittery, while a rating of 1 corresponds to being completely calm, relaxed, or dull.

English words selected in this study had average arousal ratings of higher than 6 in the ANEW list and were, consecutively, highly arousing. But their valence levels were either greater than 7 (positively valenced) or below 3 (negatively valenced). These criteria were utilized because valence effects are more strongly observed when arousal

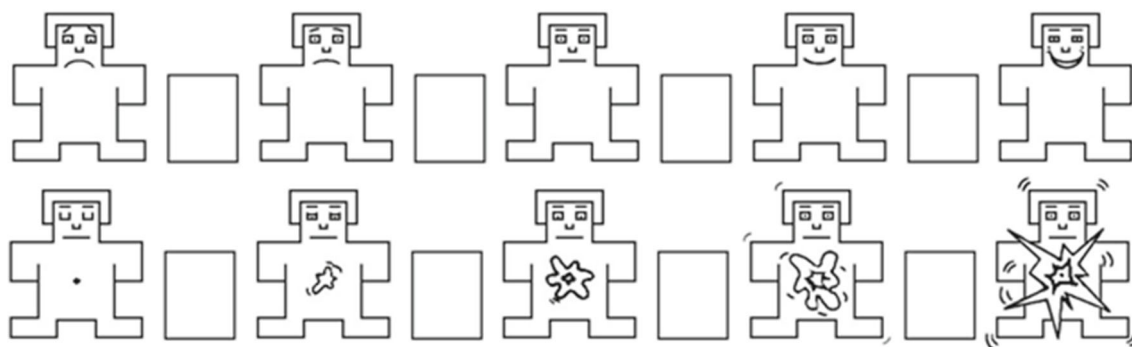


Fig. 1 The 9-point scale SAM pictures used for self-assessment of valence (top) and arousal (bottom) dimensions. (Top) The left-most SAM on the valence scale denotes being very unhappy, despaired, or unsatisfied, while the right-most SAM reflects being extremely happy, pleased, and satisfied. (Bottom) The left-most picture is selected to

show relaxed, calm, and unaroused states, whilst the right-most SAM is related to being highly aroused due to extreme excitement or jitteriness. Pictures adapted from the ANEW collection, the Center for the Study of Emotion & Attention, University of Florida

levels are high (Gomes et al. 2013). From the most recent ANEW list composed in 2010, 146 and 137 words satisfied these initial selection criteria for arousing-positive words and arousing-negative words, respectively. Next, English words were chosen in terms of their familiarity and ease of understanding for the majority of the Turkish population having acquired English as a second or foreign language. Furthermore, familiar words with slang or taboo content were excluded from the selection process. If two different entries of identical ANEW words from different lexical (sub)classes satisfied these conditions, as happened for the word “annoy”, the entry with the more extreme mean valence score was selected as the stimulus.

These criteria resulted in 65 negative words (50.78% of total words) with low valence and high arousal, and 63 positive words (49.22%) with high valence and high arousal. This collection is hereafter called the Original English List or OEL. Out of these 128 stimuli, 58 words (45.31% of the total) are exclusively used as nouns, 30 words (23.44%) as adjectives, and 10 words (7.81%) as verbs. Two words (1.56%) have dual usage as nouns and adjectives, and the remaining 28 words can be used as both nouns and verbs.

OEL Translation: All 128 words in the OEL were translated to Turkish by two bilingual Turkish-English experimenters, and the translations were contrasted and verified by a professional and bilingual linguist. They were asked to suggest translations closely matching the original meanings of words and with the highest frequency of use in the contemporary Turkish language. Special attention was given to the preservation of the lexical classes. If the two translators and the linguist were not in agreement or when more than one Turkish translation was offered for an English word, all the suggestions were considered for back-translation (Eilola and Havelka 2010; Montefinese et al. 2014). Cultural modifications were taken into account as well; for example, the word “Dollar” from the OEL was replaced by *Lira* to carry the same concept for people residing in Turkey. Finally, a number of distinct OEL words had identical translations in the Turkish language: “afraid/scared” (*korkmuş*), “discount/sale” (*indirim*), and “killer/murderer” (*katil*). Therefore, the words “afraid”, “discount”, and “murderer” were selected as the stimuli, and three new ANEW words satisfying the same valence and arousal criteria were added to the OEL before back-translation.

Back-translation: Eight Turkish students with an average age of 21 years, enrolled in the senior level of undergraduate programs at Sabanci University, voluntarily assisted the experimenters with back-translation. These students had no prior knowledge of the original English stimuli or the suggested back-translations by others, and did not join in the subsequent rating of the adapted Turkish list. The

majority of votes were considered on a case-by-case basis, and when there was a complete discord, the original word from ANEW was selected.

Adapted Turkish Affective List: Since the translation suggested for “pretty”, *sevimli*, was found by back-translators to better match the word “cute”, “cute” replaced “pretty” in the list of English stimuli. Likewise, the suggested words for “toxic”, i.e., *toksik* and *zehirli*, were respectively back-translated to “toxic” and “poisonous”, even though the latter is not included in the full ANEW list. Thus, the pair of (“toxic”, *toksik*) was selected for the subsequent use. After adding three new words from ANEW, namely, “abuse”, “blackmail”, and “bribe”, a list with 127 English words, hereafter known as the Revised English List (REL), was obtained. The equivalent Turkish words used for adaptation are also known as the Adapted Turkish Affective List (ATAL). Table S1 in Supplementary Material A shows the classification of lexical categories for ATAL in terms of nouns, adjectives, verbs (ending in *-mak* or *-mek*), and words equally used as both nouns and adjectives.

The middle column of Table 1 shows the average valence and arousal values for three valence categories of the 127 REL words. This list contains 63 positive words with high valence and high arousal, 63 negative words with low valence and high arousal, and one neutral word, “bribe”, with mid-level valence and high arousal. In line with the selection criteria, the mean valence ratings for these English words significantly differed between positive and negative words (Cohen’s $d = 13.433$, Student’s $t = 75.387$, $p < 0.001$). However, there was no significant difference between the arousal levels of the positively and negatively valenced words (Cohen’s $d = 0.008$, $p > 0.9$). Finally, Table S2 in Supplementary Material A contains the classification of lexical categories for the 127 words of REL.

Table S3 in Supplementary Material B includes the full ATAL stimuli, their equivalent ANEW word numbers, and information on the number of letters, lexical categories, and frequency of use for the Turkish words. Word frequencies are reported as the number of case-insensitive instances per million obtained from TS Corpus V2³ (Sezer and Sezer 2013), a general purpose Turkish corpus based on the BOUN Web Corpus⁴ collected from various online resources such as newspapers, weblogs, and forums. This table does not contain the relative word frequency of English words because this information was partially mentioned in ANEW 1999 and was not included in the 2010 version from which our current stimuli have been extracted.

³<https://tscorpus.com/corpora/ts-corpus-v2>

⁴<http://79.123.177.209/hasim/langres/BounWebCorpus.tgz>

Table 1 Mean, standard deviation (SD), and standard error (SE) of affective ratings based on the valence categories of the 127-word English and Turkish Affective lists

	Revised English list	Adapted Turkish affective list
Range of Valence	1.25 – 8.72	1.15 – 8.44
Average Valence	5.03 ± 2.86 (SE = 0.25)	4.83 ± 2.70 (SE = 0.24)
Arousal Range	5.53 – 8.06	3.98 – 7.98
Average Arousal	6.67 ± 0.54 (SE = 0.05)	6.11 ± 0.80 (SE = 0.07)
Positive-Valence Category	N = 63 (49.61%)	N = 50 (39.37%)
Valence	7.86 ± 0.43 (SE = 0.05)	7.72 ± 0.36 (SE = 0.05)
Arousal	6.68 ± 0.56 (SE = 0.07)	6.42 ± 0.73 (SE = 0.10)
Neutral-Valence Category	N = 1 (0.78%)	N = 20 (15.75%)
Valence	4.14 ± 0.00 (SE = 0.00)	5.47 ± 1.46 (SE = 0.33)
Arousal	6.00 ± 0.00 (SE = 0.00)	5.59 ± 0.75 (SE = 0.17)
Negative-Valence Category	N = 63 (49.61%)	N = 57 (44.88%)
Valence	2.21 ± 0.42 (SE = 0.05)	2.07 ± 0.43 (SE = 0.06)
Arousal	6.67 ± 0.51 (SE = 0.06)	6.01 ± 0.77 (SE = 0.10)

Data Acquisition

A web-based survey was prepared using the Qualtrics Research Suite⁵ to be completed under the supervision of experimenters in a computer laboratory. This survey consisted of 127 stimuli divided alphabetically into three blocks of 32 words and one block of 31 words. Words in each block were randomly sorted for each participant. The corresponding SAM scales (five pictures and four white bars) of Fig. 1 were displayed on the top of the screen while each word was followed by nine radio buttons. The task was to click on the appropriate radio buttons to rate the valence of all the words in a block before proceeding to the corresponding arousal evaluation screen and subsequent word blocks.

Questionnaire completion for rating the Turkish words was conducted in four sessions, with a different group of participants in each session. After participants' admission, experimenters provided a short introduction in English about the subjective evaluation of affective loads based on the two-dimensional valence-arousal plane. The pictorial emoticons corresponding to 9-point ratings were simultaneously projected on a screen, and a few examples were presented to familiarize students with picture selection based on their instantaneous feelings towards those stimuli. However, unlike a number of earlier adaptation efforts, such as Montefinese et al. (2014), participants were not explicitly told to choose level five if they were unsure of a word's

valence or arousal level. This instruction was withheld to avoid any bias towards impulsive selection of five before contemplating the words' emotionality. Finally, participants answered basic demographic questions before proceeding to work on the web-based survey at their own pace. The experiment was not timed, but the majority of participants completed the task in under 30 minutes. The data collection protocol and practices were approved by the Research Ethics Council of Sabanci University.

Participants

A total of 61 undergraduate students (49 females and 12 males) enrolled in two undergraduate psychology courses at Sabanci University evaluated the affective loads of ATAL in return for research credit. Student ages ranged from 19 to 24 years old ($M = 21.48$, $SD = 1.15$), and 57 students (93.44%) identified themselves as right-handed. All participants had normal or corrected-to-normal vision, were native Turkish ($N = 60$) or Azeri speakers, and were professionally fluent in English as the medium of instruction in this university is English.

Results

The reported mean (M) and standard deviation (SD) values for ATAL are calculated from 61 sets of responses – votes – to the 9-point scale valence and arousal ratings. The mean and standard deviation of valence and arousal for the REL words are obtained from the ANEW 2010 dataset. The descriptive statistics for the affective normative ratings of REL and ATAL as well as the Student's t -test, Cohen's d ,

⁵<https://www.qualtrics.com/research-core/>

and Pearson's r statistics are computed in Excel and verified with the Statistics Toolbox of MATLAB 2016a. Due to the inaccessibility of the individual raw data from the Qualtrics Research Suite, the currently available ratings could not be analyzed in terms of gender bias and reliability assessment with the split-half method. In order to calculate the required number of samples for observing a specific statistical power in advance, the test power, $1 - \beta$, was selected to be at least equal to 0.80 and β was at least 4 times larger than the significance level α (Cohen 1988). Thus, with an anticipated effect size (Beck 2013) of 0.36, a desired statistical power of 0.80, and a statistical significance level of 0.05, each of the two groups in a null hypothesis test should have at least 123 samples for a two-sided test while there are currently 127 samples in the REL and ATAL datasets.

Descriptive Statistics

Columns H to K of Table S3 from Supplementary Material B contain the mean and standard deviations of valence and arousal norms from 61 submitted answers. The descriptive statistics for valence and arousal dimensions of the entire ATAL are calculated separately for each valence category and displayed in the right column of Table 1. The mean valence of 50 ATAL words, i.e., 39.37% of the total stimuli, was higher than seven. 58 words or 44.88% of the total stimuli were rated highly negative with a mean valence lower than 3, and the remaining 20 words had a mean valence between 3 and 7.

To test whether there was any significant difference among the arousal ratings of three valence categories, a one-way ANOVA was conducted that showed a main effect of valence on the arousal ratings, $F(2, 124) = 9.58$ and $p < 0.001$. From two-sample Student's t -tests, the largest

difference in arousal values was observed between those of the positively and neutrally rated words with Cohen's $d_{pos-neut} = 1.127$.

As shown in the right column of Table 1 and the right plot of Fig. 2, a low congruence was observed among participants in evaluating the pleasantness of the 20 words not belonging to the extreme valence categories. This is while their equivalent English words, as shown in the left plot of Fig. 2, were distinctly positive, neutral, and negative. To better compare and contrast the distributions of obtained and ground truth ratings, box plots in Figs. 2 and 3 demonstrate the 25th, 50th, and 75th percentiles of the average valence and arousal ratings for each REL and ATAL valence category. It can be seen that the perception of pleasantness or unpleasantness of Turkish words covered the whole 9-point spectrum and was not limited to either extreme. Furthermore, the activation or arousal that the native Turkish speakers felt by these written stimuli was higher for the two extreme valence categories than for words they rated to have medium valence levels.

Homogeneity Plots

The linear fits to the $M \times SD$ distributions of the ATAL valence and arousal dimensions only accounted for -0.64% and 9.81% of their total variations, respectively. A reverse U-shape curve fitted to the SD versus M plot of Turkish valence ratings explained 46.17 % of the model variations and depicted more congruence in rating the valence dimension of highly emotional words. A similarly reversed quadratic curve fitted to the homogeneity plot of the arousal dimension resulted in the adjusted $R^2_{Arousal, poly2} = 0.217$. Similar shapes were observed for the majority of other affective lexicons.

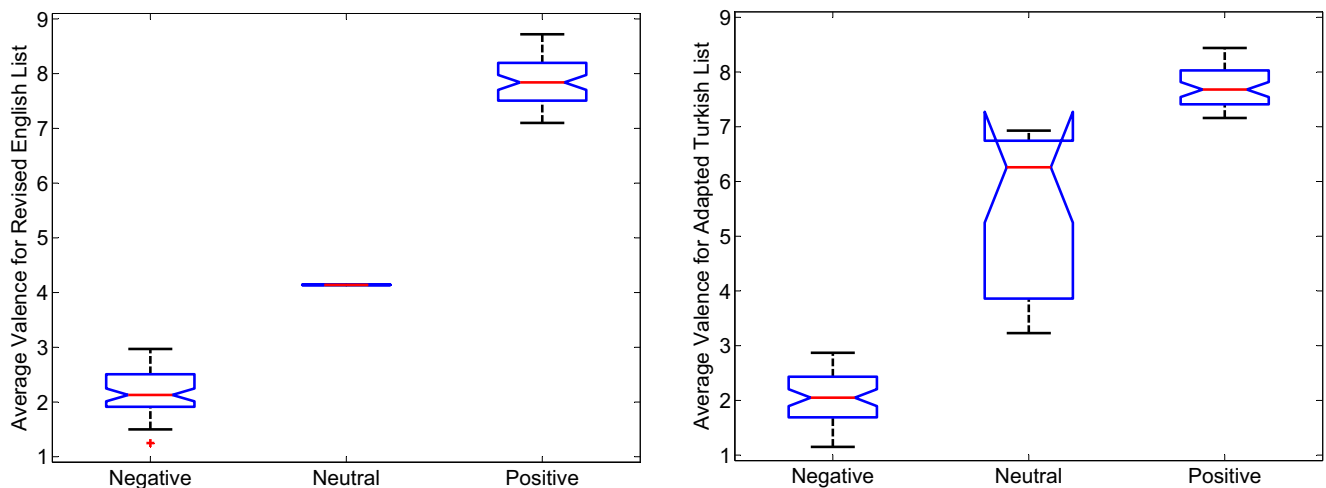


Fig. 2 The box plots for sample mean distributions of valence ratings for the 127 words of (left) REL and (right) ATAL

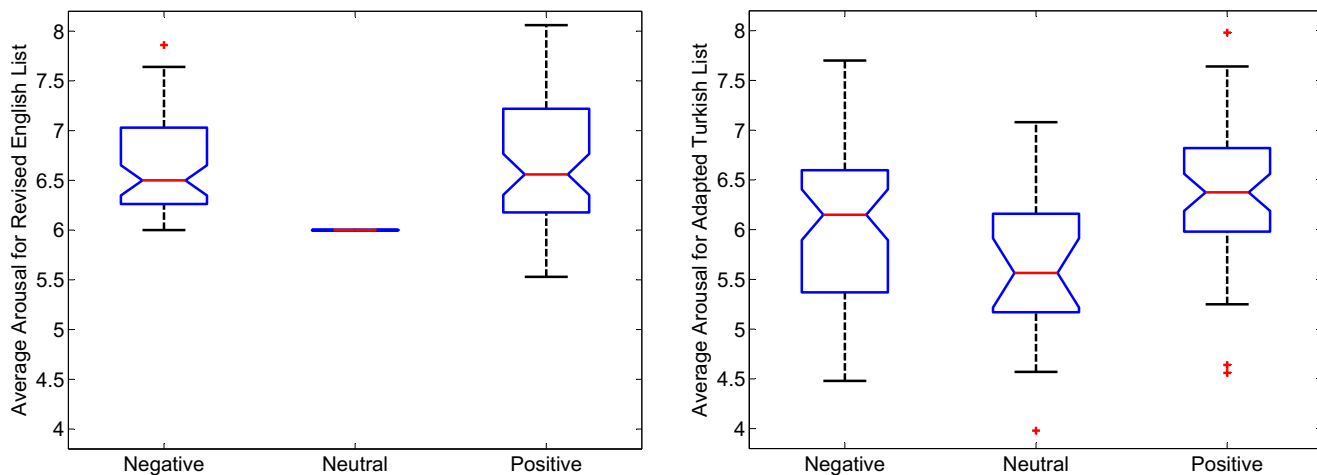


Fig. 3 The box plots for sample mean distributions of arousal ratings for the 127 words of (left) REL and (right) ATAL from three valence categories

Relationships Across Measures

Having obtained the valence-arousal scatter plots for the English and Turkish datasets as depicted in Fig. 4, a quadratic curve was fitted to the distribution of the REL stimuli described by $y = 0.057x^2 - 0.573x + 7.634$ with the adjusted $R^2_{EN,poly2} = 0.055$. The linear correlation for these English norms was weaker as $r = 0.034$ and $p > 0.7$. Although our initial word selection criteria limited the majority of English stimuli to be highly arousing and extremely pleasant or unpleasant, the respective valence-arousal distribution still resembled the familiar U-shaped curve as reported by Bradley and Lang (1999) and the majority of their adaptations. However, a gap was visible in the interval between scales 3 to 7. The linear correlation between the Turkish average norms was equal to $r =$

0.171 , $p = 0.054$, with the adjusted $R^2_{TR,poly1} = 0.022$. A stronger quadratic fit was nevertheless obtained for the ATAL ratings described by $y = 0.100x^2 - 0.919x + 7.493$ which explained 15.50% of the total model's variations. It can be inferred that words at either valence extreme sounded more arousing or agitating to the Turkish participants, while the perceived charge of neutral words was closer to the medium level. Furthermore, the positive ATAL stimuli were found more arousing than the negative words, although not statistically significant, and negative words had higher average arousal levels than the neutral category, $p < 0.05$.

Intercultural Differences in Affect Assessment

Correlations of affective ratings between the original REL dataset and the adapted Turkish affective list was examined

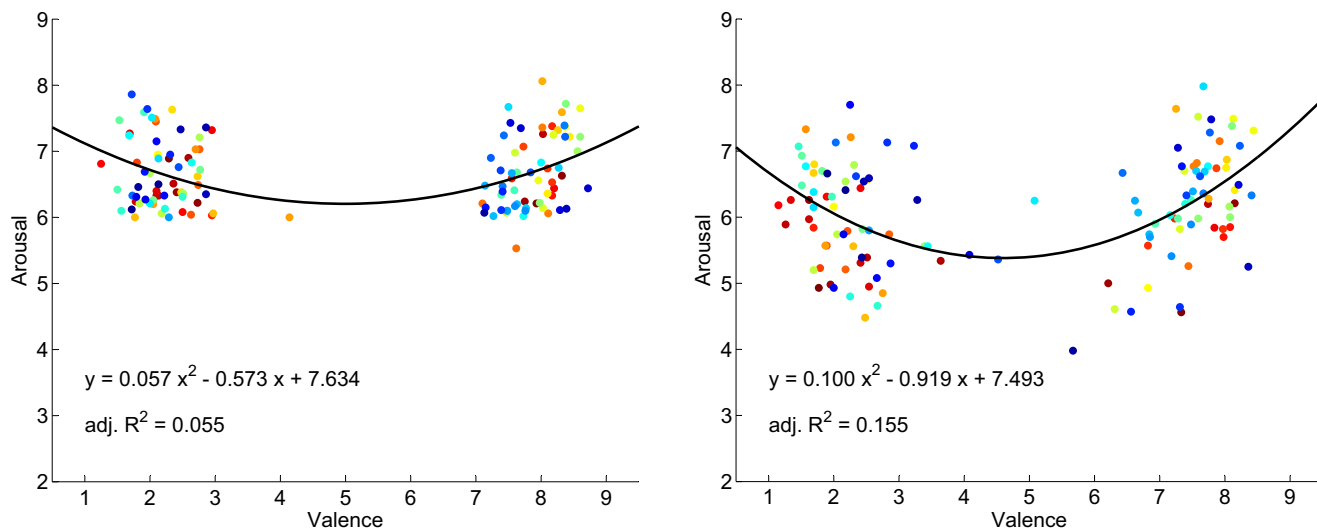


Fig. 4 The Valence × Arousal space for the Revised English List (left) and Adaptive Turkish Affective List (right)

to provide a measure of usability for the adaptation attempt and a clear picture of affect assessment for the Turkish emotional words. The means of affective valence ratings from the two datasets have a very strong and positive linear correlation with $r = 0.968$ and $p < 0.001$, indicating a high degree of consensus among Turkish and English speakers toward perceiving subjective pleasantness. However, analyzing the arousal norms reveals a medium-level correlation between the average arousal ratings of the two lists, $r = 0.482$ and $p < 0.001$.

The assessed valence and arousal levels for our adapted Turkish words were, in average, 0.20 ± 0.72 and 0.56 ± 0.72 points lower than the corresponding dimensions for their equivalent English words (Cohen's $d = 0.071$ and 0.825 , respectively). Thus, differences in arousal norms of English and Turkish words had a larger effect size. To investigate these variations in more depth, words with more than one unit difference in their valence and arousal ratings between the two languages were listed in Tables S4 and S5 of Supplementary Material C.

Table S5 contains the list of 40 words with more than one point difference in their average arousal levels between the Turkish and English ratings. However, more care should be taken into account for analyzing these variations. Perceiving the idea of emotional excitation or charge – and dominance – is more challenging than that of valence in general since the former variables depend on mood or degree of anxiety of the individuals and their recent life events (Montefinese et al. 2014). A qualitative look at words with different arousal ratings shows that the young Turkish students were highly excited by the word *aşk* and rated it, in average, 1.04 points more arousing than “love”. The Turkish versions of “couple”, “passion”, “romantic”, and “seduction” were, however, perceived at least one level less arousing for them. Even Turkish equivalents of “actor” (*aktör*) and “cash” (*nakit*) were considered less exciting, resulting in at least two points lower arousal scores compared to what was perceived by the native English speakers. Getting a “divorce”, being “enthusiastic”, thinking of someone as “genius”, telling a “joke”, “laughing”, or showing signs of “intelligence” did not trigger their agitation or excitement as high as the ANEW scores either.

Implications of Study 1

The current 127-word adaptation of ANEW into Turkish focuses on highly arousing words. Although only one neutral word was included in the final list of English stimuli, we observed that 20 Turkish words were in average rated as having neutral valence levels. This mapping by itself is informative; however, one shortcoming of this study is the small number of words which makes it hard to match and balance the emotionally-loaded stimuli with neutral words

in terms of age of acquisition, familiarity, and frequency of use in the destination language. In addition, respondents tend to assess the emotional loads of words with respect to a baseline, and it was suggested that the absence of neutral words could distort their ratings given to the originally positive and negative words which, subsequently, could change the distributions in the valence \times arousal and $M \times SD$ spaces. For these reasons, we decided to expand the number of stimuli in the dataset by adding more common words from ANEW 2010 and including respondents from more diverse age and socio-economic backgrounds to obtain insights into existence of any bias in the valence and arousal assessments of the contemporary Turkish language.

Study 2: The 508-word Expanded Turkish Affective List

Material and Methods

Stimuli Selection

In order to substantially increase the source word list, we first included all the ANEW 2010 words with neutral valence (between 4 and 6) and middle arousal levels. Next, words from all the three valence categories whose mean arousal level was higher than 6 were added to our highly arousing REL stimuli. Similar to the procedure described for the Original English List of “Study 1”, for words such as “annoy” and “drown” which appear in ANEW with two different lexical subclasses, the entry with the more extreme mean valence score was selected. However, this time we did not have a definite exclusion criteria regarding the taboo words. This selection process resulted in a total of 176 neutral, 110 positive, and 110 negative words in addition to the one neutral, 63 positive, and 63 negative words from the REL. The average arousal levels of all the selected English words were higher than 5. These 523 words were then submitted to four native speakers of Turkish, professionally fluent in English, for translating in isolation. When no consensus was reached among the translators for a given word, all the submitted suggestions were recorded. An example was the word “breathless” that, depending on the context, could be translated to *nefesi kesilmiş*, *nefessiz*, or *nefes nefese*.

Subsequently, a list with 550 Turkish words was prepared for back-translation. Native speakers from the academia and general public were invited through posts on social media and the school bulletin board for back-translation. Each volunteer received a spreadsheet with 110 words. A total of 23 spreadsheets were returned, with each Turkish word having received between 3 and 6 English suggestions. The 550 English words, their original Turkish translations,

and the suggested back-translations were compared and contrasted by the authors, and the final stimuli were selected based on the following criteria: a) The high frequency of use in the contemporary Turkish language and the chance of correct perception by young Turks; b) English words with intact Turkish translations whose true meanings were not at all perceived or guessed by the back-translators or were rare in the contemporary Turkish culture were eliminated. Examples: “bicker”, “bumble”, “enforcement”, and “nudist”; c) English words whose Turkish translations consisted of three parts and required a thorough explanation for their concept to be fully comprehended were also removed from the list. Example: “Hiker”; d) If two or more English words were best described by one unique Turkish word, the least frequently used English option was removed. Examples: “Ambush” versus “lurk”, “applause” versus “cheer”, and “pistol” versus “revolver”. Finally, from the Revised English List introduced in “Study 1”, translations of two previously adapted words were updated accordingly: The original translation of *keyif* for “joy” was replaced with *neşe* while the word *zalim*, previously used for “violent”, was updated to *şiddet içeren*.

These criteria resulted in 508 adapted Turkish words, hereafter known as the Expanded Turkish Affective List (ETAL), used for assessment of emotional perception from native Turkish speakers. The finalized 508-word English list will be subsequently referred to as the *Expanded English List*. These words were divided to three valence categories: positive (mean ANEW valence above 6; 169 words), neutral (mean ANEW valence between 4 and 6; 167 words), and negative (mean ANEW valence below 4; 172 words). The middle column of Table 2 shows the range of valence and arousal norms as well as the mean, standard deviation,

and standard error of these ratings for these three valence categories.

Prepared Questionnaires

ETAL words were similarly divided to three valence categories, and each category was randomly permuted and divided into three sub-lists with 55 to 57 words. Word sub-lists A, B, and C were then constructed from the permuted concatenation of three different sub-lists, and were checked so that no more than two consecutive words belonged to the same valence category. This was done to ensure the emotional load of earlier words would not drive the perceived valence and numerical scale of the participants' answers to their subsequent stimuli.

Data Acquisition

To reach a wider audience from different geographical locations, a web-based survey was prepared using the Qualtrics Research Suite that could be completed online and at one's own pace. The test was not supervised for any assemble of undergraduate students to ensure all participants experienced similar test-tasking environments. Upon opening the survey links, participants would see a welcoming message and a brief set of instructions regarding the nature of the experiment and the necessity to choose the affective scales according to their own feelings regardless of what polarity the society would assign to each dimension. After they provided an informed consent by agreeing to share their anonymous information for research purposes, a page describing the standard valence instructions and an example depicting the selection method with SAM icons

Table 2 Mean, standard deviation (SD), and standard error (SE) of affective ratings based on the valence categories of the 508-word Expanded English and Turkish Affective lists

	Expanded English list	Expanded Turkish affective list
Range of Valence	1.25 – 8.82	1.222 – 8.422
Average Valence	4.999 ± 2.130 (SE = 0.095)	4.550 ± 2.025 (SE = 0.090)
Arousal Range	5.00 – 8.17	2.438 – 8.444
Average Arousal	6.075 ± 0.676 (SE = 0.030)	5.369 ± 0.893 (SE = 0.040)
Positive-Valence Category	N = 169 (33.27%)	N = 161 (31.69%)
Valence	7.494 ± 0.619 (SE = 0.048)	6.966 ± 0.610 (SE = 0.004)
Arousal	6.243 ± 0.713 (SE = 0.055)	5.199 ± 0.894 (SE = 0.006)
Neutral-Valence Category	N = 167 (32.87%)	N = 123 (24.21%)
Valence	5.047 ± 0.602 (SE = 0.047)	5.053 ± 0.592 (SE = 0.005)
Arousal	5.793 ± 0.533 (SE = 0.041)	4.868 ± 0.787 (SE = 0.006)
Negative-Valence Category	N = 172 (33.86%)	N = 224 (44.10%)
Valence	2.501 ± 0.542 (SE = 0.041)	2.538 ± 0.667 (SE = 0.003)
Arousal	6.183 ± 0.680 (SE = 0.052)	5.766 ± 0.763 (SE = 0.003)

would appear. Next, the online interface would uniformly and randomly assign one of the sub-lists A, B, or C to each participant. For easier matching of the 10-point scales with the SAM icons, each list was divided into three pages. Once the respondents had finished rating the valence dimension for all the words of their sub-list, a similarly short set of instructions were displayed for rating the arousal dimension. Other test design strategies were similar to those described in “Data Acquisition”. To follow the common practice in ANEW adaptations and emotional word list ratings, a number of demographic questions were implemented at the end of the survey to collect information about gender, age, handedness, first and second languages, highest academic degree and occupation, nationality, and cities of birth and residence. The online survey was set to record partial answers while requiring the participants to answer all the affective ratings in each page before proceeding to the next page or dimension. The experiment was not timed, but the majority of participants completed the task in less than 30 minutes.

Participants

The survey links were shared with the faculty, staff, and students of Sabanci University via the institution bulletin board, and with professionals and academicians in the fields of psychology and linguistics via social media and nationwide mailing lists. Participation was voluntary and was not compensated with monetary gifts or course credits. The partially saved data of approximately one third of respondents were discarded as they had only provided their informed consent to carry out the online survey collection, but had failed to complete the first page of emotional ratings. Thus, a total of 136 responses were recorded and used for analysis. Out of the 49 participants who had submitted their demographic information, 38 people (77.55%) were female and 11 were male (22.45%), with the age range of 18 to 64 years old ($M = 34.14$, $SD = 10.36$). Furthermore, 45 respondents (91.84%) identified themselves as right-handed. All participants but one were nationals of Turkey, and all but four (91.84%) were residents of Turkey at the time of survey completion. Respondents had indicated Turkish as their first and English as one of their second languages. Furthermore, eight people (16.33%) were familiar with German and five (10.20%) with French, while seven other European and middle-eastern languages were mentioned as a second or foreign language. In terms of the highest obtained academic degrees, six people (12.25%) had a high school diploma, one (2.04%) an associate degree, 14 (28.57%) a bachelor's degree, 11 (22.45%) had master's, 15 (30.61%) had a PhD or MD degree, and two people (4.08%) were medical specialists. Four people had identified themselves as instructors, eight as

academicians/professors, five as researchers or research assistants, 10 as undergraduate or master's students, seven as psychologists, and the rest as engineers, doctors, lawyers, musicians, etc.

Results

Descriptive Statistics

The three sub-lists of Expanded Turkish Affective List received a total of 136 partial and complete responses from our online survey. Out of the 46 responses recorded for sub-list A, 76.09% provided valence ratings for all the 170 words and only 34.78% proceeded all the way to the end of arousal norms. These ratios were equal to 77.55% and 42.86% for sub-list B with a total of 49 recorded responses, and 63.41% and 39.02% from 41 valid responses saved for sub-list C. In summary, each word had received 26 to 49 ratings for the valence dimension and 16 to 25 ratings for the arousal dimension. Table S6 in Supplementary Material D contains the weighted arithmetic mean, standard deviation, and standard error for the valence and arousal norms of the obtained affective ratings as well as the number of letters for the original English words and their 508 Turkish adaptations. Figure 5 shows histograms of the average valence and arousal ratings for the expanded Turkish list while the right column of Table 2 displays the number of positive words (with mean valence above 6), neutral words (with mean valence between 4 and 6), and negative words (with mean valence below 4) based on the affective norms obtained from our Turkish participants. As can be seen, the average valence of 44.10% of all the Turkish words was below 4, resulting in a shift towards perceiving words as unpleasant despite inclusion of neutral words in Study 2.

Reliability

The split-half procedure was performed to assess the reliability of calculated measures for the three sub-lists and the entire 508 Turkish words. This analysis, known for measuring the stability of collected affective ratings, was also used by Warriner et al. (2013) and Riegel et al. (2015), and Imbir (2016) although it is known for underestimating the actual reliability due to reducing the number of samples. The Pearson's linear coefficients revealed strong correlations of 0.975, 0.966, and 0.964 between the average valence norms of the two splits of our sub-lists, all statistically significant as $p < 0.001$. Furthermore, two-sample Student's t -tests revealed no statistically significant differences among the valence values of the aforementioned splits as $p > 0.4$ for all comparisons. The correlation between the average valence of the entire 508 words was also strong and significant, $r = 0.965$ and $p < 0.001$.

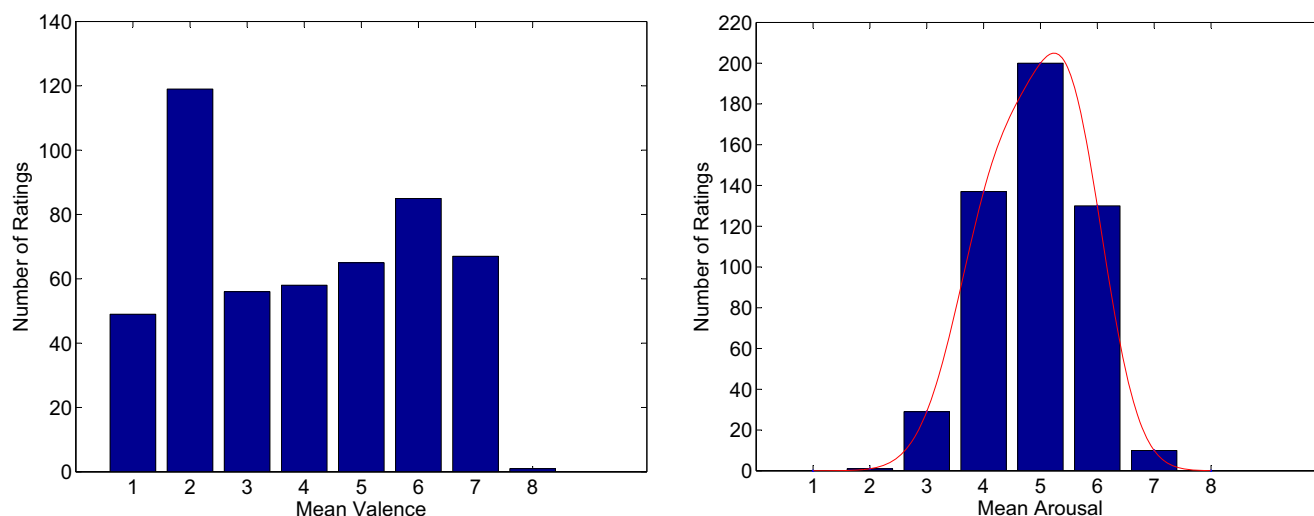


Fig. 5 Frequency histograms of valence (left) and arousal (right) normative ratings of 508 Turkish words

For the arousal norms, linear relationships were significant for sub-lists A and B ($r = 0.316$ and 0.587 , $p < 0.001$), and weaker for sub-list C as its arousal pages were rated by 16 to 19 participants. Two-sample Student's t -tests showed statistically significant differences only between the average arousal norms of sub-list B, $p < 0.001$, Cohen's $d > 0.15$. A significant linear correlation of 0.276 , $p < 0.001$, was obtained between the average arousal scores of the entire words, and it was decided to carry on the analyses focusing on all the words instead of individual sub-lists.

Homogeneity Plots

Figure 6 represents the $M \times SD$ planes for valence dimension of the Expanded English List and ETAL. The new selection criteria of “Study 2” result in spanning the whole 9-point range for both English and Turkish valence norms. The linear Pearson correlation for M - SD of Turkish valence norm is equal to 0.268 , $p < 0.001$, with the linear fit explaining only 7.00% of the total variations in the model. The reversed U-shape quadratic curve fitted to the same norms has an adjusted $R^2_{TRValence}$ of 0.401. The quadratic fit thus reveals more congruence among respondents in rating the valence dimension of extremely pleasant and unpleasant stimuli than the neutral words.

The small but statistically significant correlation of $+0.268$ between the mean and standard deviation of Turkish valence norm is in line with the heavier density towards the lower tail of the right plot in Fig. 6. To better analyze the congruence for different valence categories, Pearson correlation coefficient was calculated for the positive, neutral, and negative intervals of this plane. A strong agreement was found in rating the valence of most extremely loaded words as $r = +0.641$ for negative words, $p < 0.01$, and $r = -0.545$ for positive words, $p < 0.01$.

The linear correlation for the neutral interval of the $M \times SD$ plane was small and insignificant, $r = -0.030$.

Similarly, Fig. 7 shows the homogeneity plots for the arousal dimension of the Expanded English List and ETAL. The selected English words all have arousal levels higher than five, and a negative correlation of $r = -0.128$ ($p < 0.01$) exists between the SD and mean of their arousal ratings. For the Turkish words, however, the distribution of arousal norms in the $M \times SD$ plane is closer to an elliptic cluster around the medium ranges. The small and insignificant correlation of $+0.033$ between the mean and SD of Turkish arousal ratings fails to describe the distribution of this cluster. A downward quadratic curve that implies more congruence in rating the arousal of more exciting or stimulating words describes only 6.15% of the total variance as well. A higher, although small, level of agreement is observed in rating the charge levels of extremely pleasant words, $r = -0.198$, $p < 0.05$.

Relationships Between Affective Variables

Analyzing relationships among affective variables and comparing the results with similar studies has two important implications. First, it enables researchers to compare their adaptation results with earlier studies – especially for datasets composed of smaller number of stimuli (Moors et al. 2013). Second, it provides a general idea on how well their raters were able to distinguish between emotional states elicited by the utilized stimuli. The average arousal of positive, neutral, and negative words from the Expanded English List were equal to 6.24 ± 0.71 , 5.79 ± 0.53 , and 6.18 ± 0.68 , respectively, with one-way ANOVA revealing a significant effect of valence category on the arousal level, $F(2, 504) = 23.93$, $p < 0.001$. The two-sample Student's t -test, however, showed no statistically

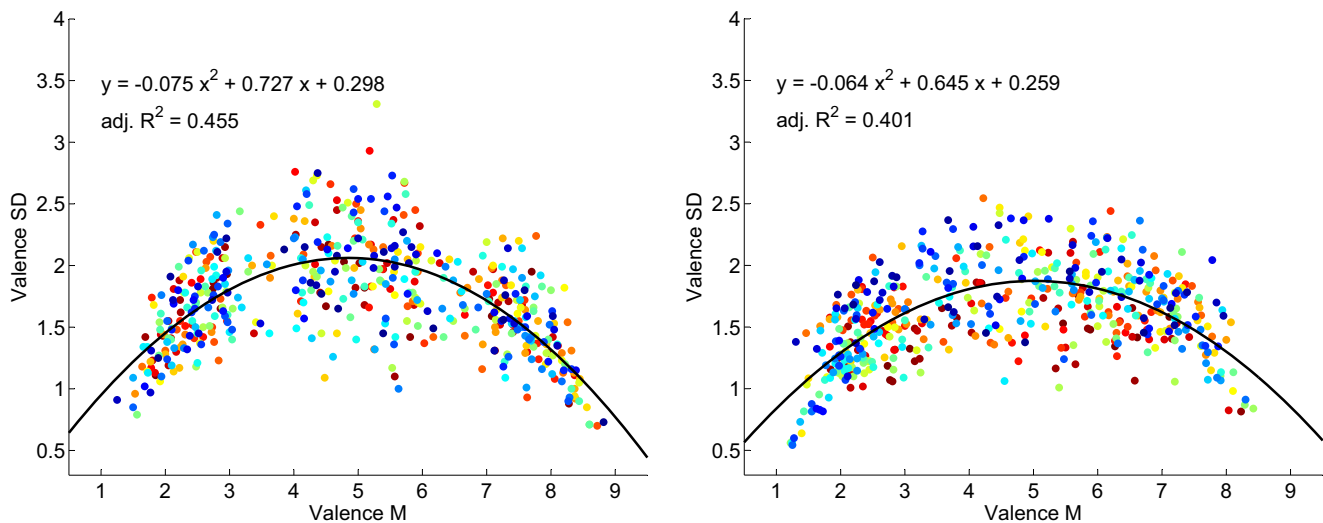


Fig. 6 Homogeneity plots for valence ratings of the Expanded English List (left) and Expanded Turkish Affective List (right)

significant differences between the arousal norms of the extreme valence categories in the selected English words, $p > 0.2$.

For words in the Expanded Turkish Affective List, the average arousal norms had reduced for all valence categories as can be seen in the box plots of Figure 8. The effect of valence category on the mean arousal levels of these Turkish stimuli was more significant than before as $F(2, 504) = 53.78, p < 0.001$. Furthermore, one-tailed Student's t -test showed that Turkish words that were perceived more unpleasant were also rated more stimulating compared to the positive ones, Cohen's $d_{TR}(neg - pos) = 0.683, p < 0.001$, and more arousing than neutral words, Cohen's $d_{TR}(neg - neut) = 1.160, p < 0.001$. This finding

was different from the results reported for "Study 1" where positive words were found to be more exciting.

Figure 9 demonstrates the Valence \times Arousal planes for the extended English and Turkish lists. Both distributions have the familiar U-shaped curve, summarizing the observation that words of either valence extreme are found to be more arousing and exciting by native speakers of each language while the neutral words are generally thought to have medium arousal levels. Similar curves were reported by Bradley and Lang (1999) and the majority of their adaptations. The goodness-of-fit for obtained ratings in both expanded lists explain their respective models' variances more than the linear fits for the valence-arousal plane of English (adjusted $R^2_{EN,poly1} = 0.000$) and Turkish (adjusted $R^2_{TR,poly1} = 0.121$) words.

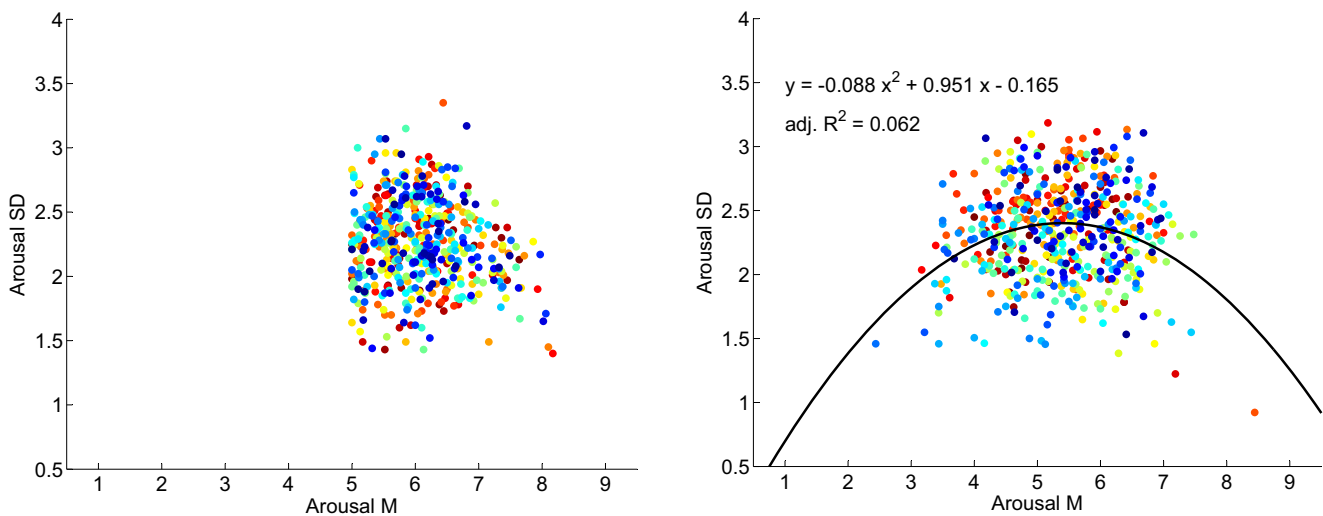


Fig. 7 Homogeneity plots for arousal ratings of the Expanded English List (left) and Expanded Turkish Affective List (right)

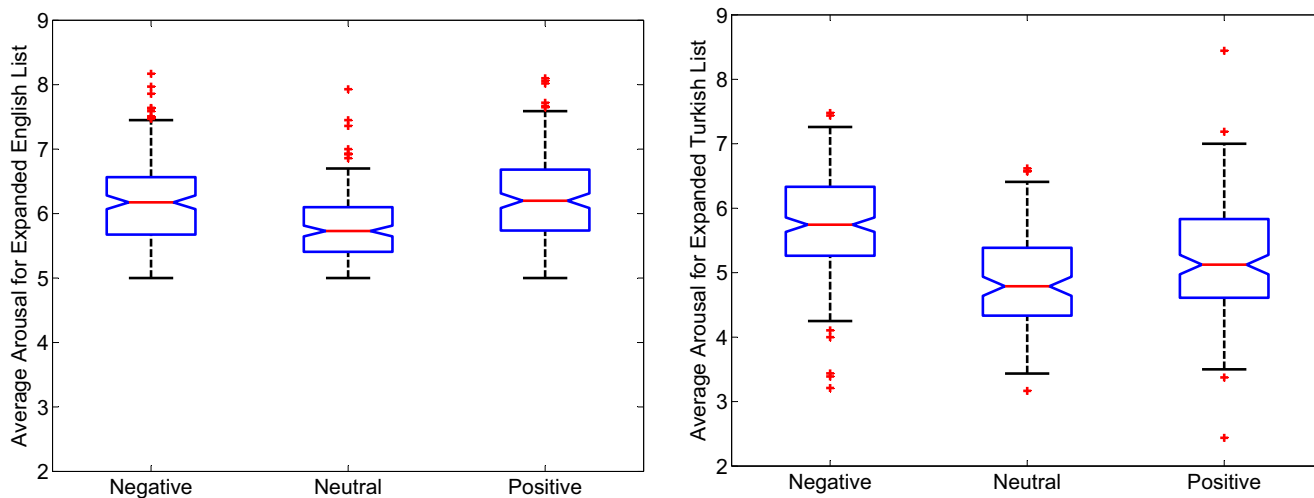


Fig. 8 Median, 25th, and 75th percentiles for the average arousal of three valence categories of the Expanded English List (left) and Expanded Turkish Affective List (right)

Analyzing the specific regions of the valence-arousal plane tends to be highly informative from the linguistics perspective as well. We observed a significant negative correlation between the valence-arousal of the negatively rated words in ATAL, while the other two valence categories did not reveal any significant correlations. Focusing on categorization criteria mentioned in Table 2, arousal and valence of negatively rated words in ETAL also carry a negative correlation ($N = 224$, $R = -0.575$, $p < 0.001$), while weaker linear correlations are observed between the arousal and valence of positively rated words ($N = 161$, $r = +0.202$, $p < 0.01$) and neutrally rated words ($N = 123$, $r = +0.045$, $p > 0.6$).

Intercultural Differences in Affect Assessment

In the 127-word Turkish adaptation list, the Pearson's coefficient of $r = +0.171$ ($p = 0.055$) between the average valence and arousal norms depicted positive words in ATAL were slightly more stimulating for our Turkish participants. A similar analysis on the ratings of expanded Turkish words, on the other hand, revealed a significantly negative correlation as Pearson's $r = -0.351$, $p < 0.001$.

Plots in Fig. 10 depict the distribution of valence and arousal norms in the TR \times EN space. The collected Turkish valence norms and those of the original English words are found to be strongly correlated as $r = +0.878$, $p < 0.001$.

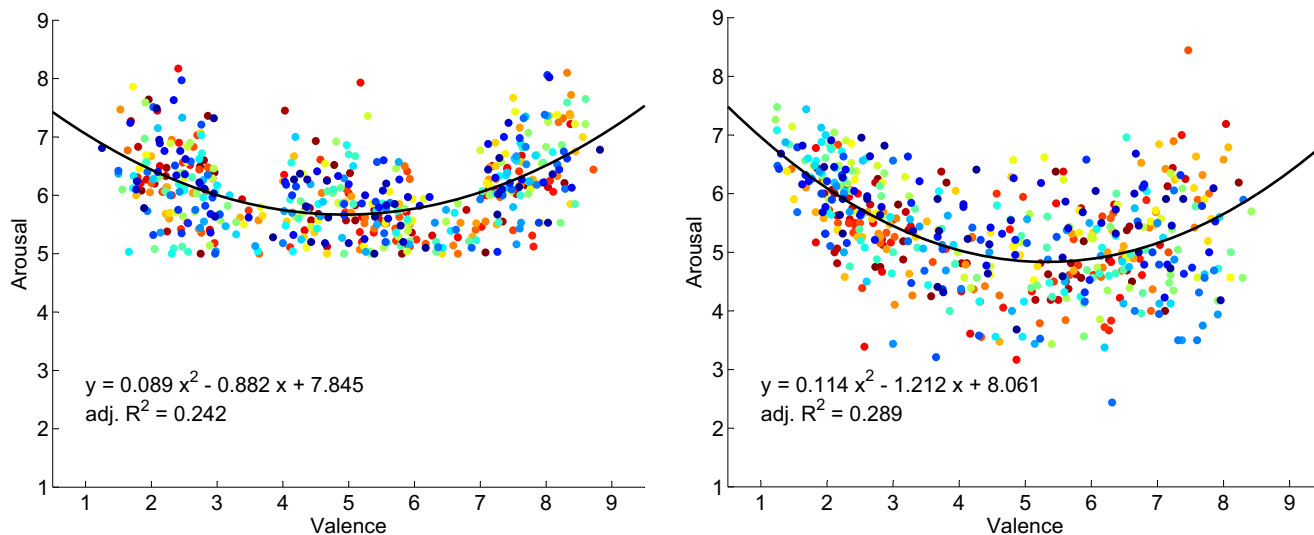


Fig. 9 The Valence \times Arousal space for the Expanded English List (left) and Expanded Turkish Affective List (right)

Looking at the right plot of Fig. 10, the linear correlation between the arousal norms of our expanded English and Turkish datasets is smaller than that of valence ratings, $r = +0.450$, but still statistically significant at $p < 0.001$.

To better compare the obtained ETAL ratings with the original ANEW affective norms of the Expanded English List, Table 2 demonstrates the number of words from each valence category of the expanded English and Turkish lists. The difference between the average valence levels of the entire Expanded English List with those obtained from ETAL was equal to 0.445 ± 1.032 ($SE = 0.046$) with a small effect size of Cohen's $d = 0.216$. Interestingly, although the three valence categories were almost uniformly distributed in the original Expanded English and its sub-lists, 44.10% of their equivalent Turkish words had mean valence levels below 4 and only 31.69% were perceived as being pleasant. The large effect size of $d_{pos.valence} = 0.861$ between the valence norms of positive words confirms this large distinction. This shift towards the lower perceived pleasantness can be observed in the left scatter plot of Fig. 10 which is more dense in the left side of the linear fit.

The arousal dimension of Turkish words saw a pattern of lowered perceived levels as well and was shifted towards the middle range. The original words in the Expanded English List were chosen to have medium to high charge levels; however, similar to the results of Study 1, their equivalent Turkish words were in general perceived to be less stimulating. Considering the whole 508 words, the difference between English and Turkish arousal norms was equal to 0.706 ± 0.843 ($SE = 0.037$) with a large effect size of 0.891. Computing the inter-dataset effect sizes results in obtaining Cohen's $d_{pos.arousal} = 1.291$, $d_{neg.arousal} = 0.576$, and $d_{neut.arousal} = 1.376$, thus observing the largest difference between the arousal levels of neutral words of the two languages.

Finally, we look at degrees of agreement in rating the emotional stimuli among the native speakers of each language: the downward quadratic fit and small but statistically significant correlation of $+0.268$ between the mean and standard deviation of Turkish valence norm is in line with the heavier density towards the lower tail of the right homogeneity plot in Figure 6, and shows that Turkish respondents had even more agreement in ratings extremely loaded words, especially the words they found unpleasant. This correlation becomes smaller and in the opposite direction for valence norms of English words, $r = -0.121$ and $p < 0.01$. To be more precise, correlations between the mean and SD of English valence norms denote that $r = +0.656$ for the negative words and $r = -0.608$ for the positive English words ($p < 0.01$). Thus, there was more congruence in rating the valence levels of words at either extreme in both languages. Regardless of the valence category, there is less dispersion in valence norms

of Turkish words than those of English words as confirmed by a one-tailed Student's t -test, $p < 0.001$. A one-tailed Student's t -test on the SD of arousal norms, however, shows that American raters had more agreement than Turkish respondents in rating the arousal levels of the 508 stimuli, $p < 0.01$.

Linguistics-affective Relationships

To investigate correlations between the linguistic and affective variables of the Turkish adaptation dataset, means, standard deviations, and standard errors for the valence and arousal norms of each word, number of letters in the English and Turkish words, and the frequency of usage of Turkish words – reported as the number of instances per million were considered. Table 3 includes a summary of these linear relationships and their respective significance levels. In addition to correlations between the affective variables, the homogeneity patterns, and the opposite direction of linear correlations in the valence-arousal space, a small and significant linear correlation of $+0.128$ was obtained between the frequency of Turkish words and their average valence levels. This correlation is larger than the Pearson's $r = 0.06$ obtained for Polish words in the Nencki database (Riegel et al. 2015) and smaller than linear correlation coefficients of 0.15 for Dutch words (Moors et al. 2013), 0.136 reported for the ANPW-R Polish list (Imbir 2016), and 0.182 for 13,915 English lemmas (Warriner et al. 2013). Furthermore, an increase in the number of letters of Turkish words – which could be due to suffixes such as *etmek* (“to make”) or *-mak/mek* for the infinite forms – is correlated with higher levels of arousal induced in Turkish participants, $r = +0.126$. A similar analysis showed a slightly weaker linear relationship between the number of letters of the expanded English stimuli and their average arousal levels, $r = +0.102$.

Discussion

Recent years have seen an increasing interest in sentiment analysis and affective lexicon processing from Turkish speech and audiovisual stimuli (Oflozoğlu and Yıldırım 2013; Çakmak et al. 2012; Aydın Oktay et al. 2015). Although valuable collections of sentence-level polarity ratings from Turkish speakers exist, developing a comprehensive affective lexicon independently or through adaptation from existing English datasets have been considered a time-consuming task. Thus, sentiment analysis researchers have directly transferred normative ratings from English datasets for translated Turkish words (Aydın Oktay et al. 2015). Even for TUDADEN (Gökçay and Smith 2008), a detailed comparison with the corresponding English words does not

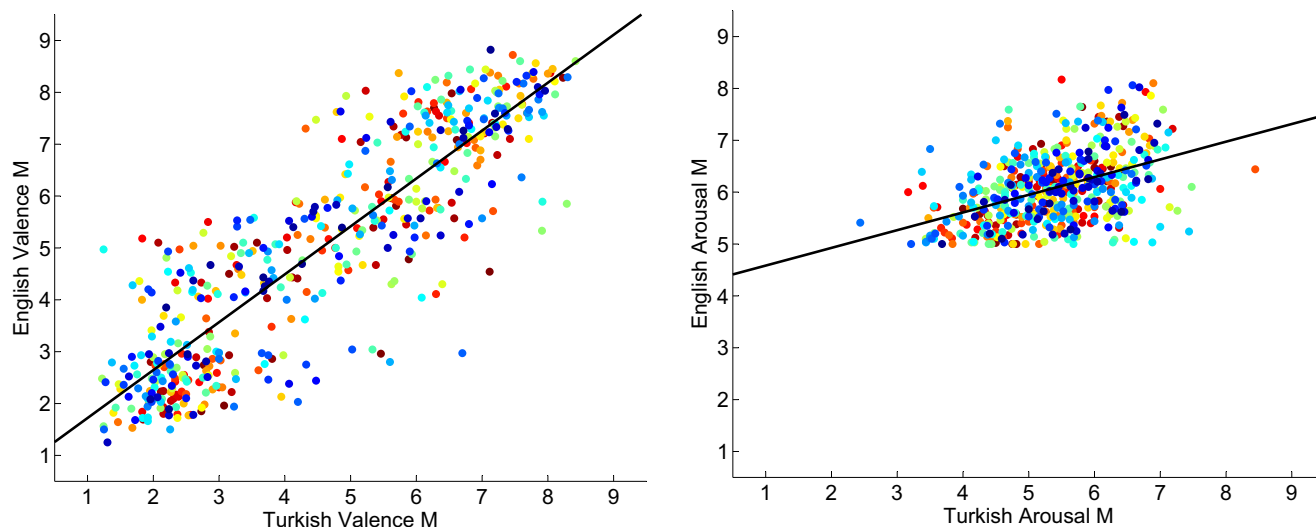


Fig. 10 (Left) Average valence and (Right) arousal of 508 words from the Expanded English List versus their equivalent measures from the 508 adapted Turkish words in ETAL. The Pearson's linear correlations are equal to $r = +0.878$ for valence and $r = +0.450$ for arousal norms, $p < 0.001$

exist. Noticing the inaccessibility of a database derived from direct emotional assessment of Turkish speakers to written verbal stimuli, we have presented the adaptation of 127 highly arousing, emotionally-loaded words and 508 positive, negative, and neutral words from ANEW 2010 to Turkish as the first partial adaptation efforts based on subjective assessment of Turkish speakers. The methods and analyses reported in this work have followed the standard procedures for developing such affective lexicon databases carried out in various European languages adapted from ANEW and non-English resources.

At the time of its submission, ATAL contained the first publicly available database that provided information for cross-linguistic comparisons. Since this adaptation was carried out with a special attention to the role of sociocultural differences in emotional perception (Soares et al. 2012), it avoided the direct use of existing English

norms for translated words that had been done for earlier Turkish studies on semantic analysis (Aydın Oktay et al. 2015) and affect assessment (Halfon et al. 2016), and still provided a ground for cross-linguistic discussions. After completing the data collection sessions for Study 2, we were notified of a study by another group of Turkish researchers who had translated – but not back-translated – 2,031 words from ANEW to Turkish, and obtained ratings for valence and arousal dimensions as well as five discrete emotion categories, including happiness, sadness, anger, fear, and disgust (Kapucu et al. 2018). This study gives a thorough picture of relations between these dimensions and emotions and the ratings of negative words for Turkish speakers, but it does not offer a comparison of obtained Turkish ratings with those of the English words, nor does it contain a discussion of homogeneity plots.

Table 3 Linear correlation patterns among affective and linguistic variables

Correlation Rho	EN ValSD	EN AroMN	EN AroSD	TR ValMN	TR ValSD	TR AroMN	TR AroSD	TR Freq.	EN Letters	TR Letters
EN ValMN	-0.121	0.044	-0.188	0.878	0.276	-0.294	0.008	0.131	0.012	-0.087
EN ValSD	1.000	-0.181	-0.013	-0.151	0.224	-0.062	-0.114	-0.038	0.044	-0.056
EN AroMN		1.000	-0.128	0.003	-0.027	0.450	0.113	-0.063	0.102	0.026
EN AroSD			1.000	-0.150	-0.150	0.095	0.101	0.015	0.044	-0.092
TR ValMN				1.000	0.268	-0.351	0.049	0.128	-0.019	-0.062
TR ValSD					1.000	-0.195	-0.071	0.105	0.047	-0.049
TR AroMN						1.000	0.033	-0.043	0.026	0.126
TR AroSD							1.000	-0.028	0.041	0.00
TR Freq.								1.000	-0.074	-0.183
EN Letters									1.000	0.222

Light gray and dark gray shades denote statistical significance at $p < 0.05$ and $p < 0.001$, respectively. EN: English; TR: Turkish; ValMN: Mean of valence ratings; ValSD: Standard deviation of valence ratings; AroMN: Mean of arousal ratings; AroSD: Standard deviation of arousal ratings; TR Freq.: Number of word instances per million. EN and TR Letters: Number of letters in the English and Turkish words, respectively

Congruence in Affective Ratings

As observed in the homogeneity plots of Fig. 6, a reverse U-shape curve was fitted to the *SD* versus *M* scatter plot of Turkish valence ratings that depicted lower variation in ratings for words with more extreme valence. Although it was suggested that elimination of neutral words from REL and ATAL could have caused a bias in their respective homogeneity plots, we observed that the emotionally-loaded database and the expanded one showed similar trends in terms of valence homogeneity plots. This is in agreement with similar fits obtained for the 13,915 English lemmas (Warriner et al. 2013) and 4,300 Dutch words (Moors et al. 2013), and in contrast with the corresponding plots of the Polish ANPW (Imbir 2015, 2016) and Italian (Montefinese et al. 2014) adaptations. A similarly reverse U-shape curve fitted to the homogeneity plot of the arousal dimension in Fig. 7 was observed for the majority of other affective lexicons. It can be inferred that the more arousing and exciting Turkish stimuli resulted in more consistent arousal ratings, but this is in contrary to the observation for 13,915 English lemmas and Polish adaptation of Imbir (2016).

Emotionality Across Languages

Our computations denote an interesting finding in affective characteristics of different languages. Although REL contains equal numbers of positive and negative stimuli, 44.88% of the rated Turkish words in ATAL were considered negative, 15.75% neutral, and only 39.37% positive. Likewise, 44.10% of the ETAL words were found to be rated negative, 24.21% neutral, and 31.69% positive although the original Expanded English List was almost equally composed of the three valence categories. In a recent study for developing a Persian affective lexicon, the balance between equally selected positive and negative words was also distorted: 52.67% of stimuli were rated as positive, 21.00% as neutral, and 26.33% as negative (Mokhlesin et al. 2015). This variation was attributed to differences in generations and political, social, and cultural environments that the middle-aged researchers and college-aged raters had experienced. If we extend the definition of positivity to any word with valence level above 5, 49.61% of our ATAL words are considered positive – in line with our word selection criteria– while 58.93% of words from ANEW 2010 and 57.28% of the English lemma database have valence levels above 5.

This positivity bias in the English language (Kloumann et al. 2012; Warriner et al. 2013) is thought to show the tendency of individuals to have pleasant and pro-social communications and is in line with the two new Polish affective ratings. Nevertheless, all the ANEW adaptations as well as the results of our study show very high

cross-linguistic correlations for the valence dimension which represents the ease of transferring and perceiving valence levels across different cultures (Warriner et al. 2013). Arousal dimension, on the other hand, is similar to familiarity and originality in exhibiting more variations across different languages.

Valence-arousal Relationships

In the original ANEW 2010 dataset, a negligible linear relationship existed between the average valence and arousal norms as $r = -0.003$. This property was maintained for the 127 words of the REL and 508 words of the Expanded English List as $r = +0.034$ and $+0.044$, respectively. These correlations increased to $r = +0.171$ ($p = 0.055$) for ATAL in Study 1. For the 508 words of ETAL, this correlation became significantly negative as Pearson's $r = -0.351$, $p < 0.001$. The 2,031 words of Kapucu et al., on the other hand, bear a correlation of 0.004 in their valence \times arousal space.

The British English and Finnish adaptations of ANEW 1999 with 210 words, including 34 taboos, also demonstrate insignificant linear valence-arousal correlations with $r = -0.057$ and $+0.066$. However, our computations show that other affective norms databases with larger number of stimuli mostly carry negative and significant linear correlations, such as in Spanish and European Portuguese with 1,034 ANEW words ($r = -0.149$ and -0.393 , respectively, $p < 0.001$), French FEEL with 835 adjectives ($r = -0.798$, $p < 0.001$), English lemmas with 13,915 words ($r = -0.185$, $p < 0.001$), Polish ANPW and ANPW-R with 1,586 and 4,905 words ($r = -0.170$ and -0.464 , $p < 0.001$), respectively, and Polish Nencki BAWL-R with 2,902 words ($r = -0.099$, $p < 0.001$). The quadratic correlations for most of the above databases are stronger than their linear fits and all follow the boomerang shape shown in Fig. 9.

Intercultural Similarities in Valence and Differences in Arousal Perception

Correlation analysis between ANEW and the adapted Turkish lists provides a measure of intercultural similarities and a clear picture of affect assessment between Turkish emotional words and their source English words. When considering the 127 highly arousing English words, the means of affective valence ratings from the Revised English and ATAL had a very strong and positive linear correlation of $r = 0.968$, $p < 0.001$. When the dataset was expanded to include neutral and medium-level arousing words, this correlation reduced to $r = +0.878$ ($p < 0.001$) which still shows the high degree of consensus among Turkish and English speakers toward perceiving subjective word

pleasantness. This level of adaptability from the expanded dataset is comparable to valence correlations for most other ANEW adaptations such as Spanish (0.916), European Portuguese (0.92), British English (0.97), Finnish (0.96), Italian (0.917), and the German-Polish study (0.85).

Analyzing the arousal norms reveals a medium-level correlation between the average arousal ratings of the 127-word lists, $r = +0.482$, which reduces to $r = +0.450$ for the expanded lists, $p < 0.001$. These linear relationships between the arousal levels of utilized American English stimuli and the adapted Turkish words are weaker than those of the Spanish (0.746), British English (0.67), Finnish (0.62), European Portuguese (0.65), Italian (0.625), and German-Polish (0.55) adaptations, and closer to the arousal correlation for taboo words in the ANEW - Finnish study (0.51).

The scatter plot demonstrates interesting differences in the absolute values and ranges of pleasantness between the two languages. An interesting outlier in the left plot of Fig. 9 is the word *patlama* (“explosion”) with a neutral valence of 5.18 ± 2.93 in English while being perceived extremely unpleasant in Turkish with an average valence of 1.833 ± 1.341 ($SE = 0.245$). Two more outliers are worth mentioning here: *aşk* (“love”) has an average arousal of 6.44 ± 3.35 in ANEW but is unanimously perceived as highly arousing in Turkish ($M = 8.444$, $SD = 0.922$, $SE = 0.217$), and *cevap* (“answer”) which has an average arousal of 5.41 ± 2.43 in English and a much lower arousal norm of 2.438 ± 1.459 ($SE = 0.365$) in Turkish.

Inspired by the discussion of results on average arousal ratings of the three valence categories performed for ATAL in Table 1 and Fig. 3, the 9-scale point datasets of ANEW 2010 (Bradley and Lang 2010), Spanish (Redondo et al. 2007), European Portuguese (Soares et al. 2012), English lemmas (Warriner et al. 2013), Polish ANPW (Imbir 2015), and Polish ANPW-R (Imbir 2016) were analyzed by considering the valence levels between 3 to 7 forming the neutral valence category. This range was reduced to 3.5 to 6.5 for British English and Finnish datasets of Eilola and Havelka (2010), 2.5 to 5.5 for the Dutch list of Moors et al. (2013), and -1.5 to $+1.5$ for the Polish BAWL-R list of Riegel et al. (2015).

In the ANEW 2010 and Spanish lists, average arousal values of positive and negative categories are found to be higher than that of the neutral category ($p < 0.001$), but no significant difference exists between the arousal ratings of the two extreme valence categories. The arousal ratings for the British English, Finnish, and Dutch adaptations in the descending order follow the positive, negative, and neutral categories as well with $p < 0.001$ for all two-way comparisons. In the French FEEL, English lemmas, and three Polish affective lists, arousal ratings in descending

order belong to negative, positive, and neutral categories ($p < 0.001$ for all one-tailed comparisons). Finally, in the European Portuguese adaptation, the negative words have higher arousal than positive ones ($p < 0.001$), but the arousal levels of neutral and positive words are not significantly different from each other.

In the Revised English List (REL), average arousal levels in the descending order belong to positive, negative, and neutral categories; but the differences are not statistically significant ($p > 0.4$). For the ATAL stimuli, positive words are more arousing than negative words, although not statistically significant, and negative words have higher average arousal levels than the neutral category with $p < 0.05$. This polarity changes in the 508 words of ETAL as negative words become significantly more arousing than the positive and neutral ones, $p < 0.001$. Having stronger reactions to extremely unpleasant words is visible from the more dense distribution towards the left half of the Turkish valence-arousal plot in Figure 9. The low number of highly charged Turkish words from the perspective of our native speakers in the same plot demands an analysis of its own. Out of only 11 Turkish words with average arousal levels equal to or greater than 7, six were rated as more stimulating than English words by a difference of at least 1 unit: “accost” (*sarkıntılık etmek*), “bullet” (*mermi*), “torture” (*işkence*), “harass” (*taciz etmek*), “execution” (*infaz*), and “love” (*aşk*). Higher arousal levels for the emotionally loaded words in the positive and negative affect groups of ATAL and ETAL are in line with all other similar studies as well as the recent Turkish affective norms of Kapucu et al.. This shift in polarity could be attributed to the wider age range, socioeconomic backgrounds, and education levels of our participants in Study 2 compared to Study 1, the differences in arousal perception and representation between Eastern and Western cultures, and the inherent sociocultural features of Turkey and Turkish respondents.

An idea presented in the “Introduction” section was the cross-cultural differences reported for arousal norms. It has been argued that Westerners value high arousal emotions more than Easterners (Tsai et al. 2007): while the Japanese and Chinese participants were reported to conceive happiness as a low arousing, positive emotion, the conception of happiness for American participants was a high arousing and more desirable feeling compared to that of their Asian counterparts (Uchida and Kitayama 2009; Lu and Gilmour 2004). Findings from studies focusing on arousal ratings corroborate these ideas. For example, American participants rated themselves to be in high arousal – stimulated, excited – emotional states whereas Asians were in the low arousal – calm, sluggish – states (Kacen and Lee 2002). These patterns are attributed to the cultural self-controls and socialization of members of these cultures

(Tsai et al. 2007; Lim 2016). Such cultural differences could explain the lower cross-linguistic correlation in the arousal ratings of the present study. Turkey has been categorized as a collectivistic country (Hofstede 1984) and, based on the cultural self construals literature (Markus and Kitayama 1991), Turkish participants would possess an interdependent self-construal that acts as a mental filter in responding to external stimuli. An interdependent self would promote relatedness goals as well as social harmony (Markus and Kitayama 1991; Cross et al. 2011), and in order to preserve group harmony and connectedness, intense –i.e., highly arousing– emotions may need to be suppressed while low arousal emotions such as calmness are promoted.

Higher arousal levels for negatively-rated words were also observed in a study on Persian affective norms, and were interpreted as negative stimuli causing more arousal and agitation due to being more threatening (Mokhlesin et al. 2015). Finally, the very nature of the rating methodology could play a role in the cross-linguistic differences of the excitability dimension. The self-report questionnaires and ratings of the kind are of reflective nature rather than immediate reactions to the stimuli (Sheldon et al. 2017; Aydin 2018), and related to the interpretation or the appraisal of the emotional reactions. Recently, Fontaine et al. (2007) identified emotional dimensions other than valence and arousal. Among them are appraisals of events and emotion regulation strategies, both of which are reported to vary across cultures; see Ford and Mauss (2015) for a review. Considering these cross-cultural factors provides a helpful and cautionary approach in interpreting the reported cross-linguistic findings as well as the future adaptation and norming studies.

Summary and Future Directions

The discussions presented in this work have focused on features of the ANEW 2010 release, the small English and Turkish lists with 127 highly arousing positive and neutral words, the expanded lists with 508 words of medium to high arousal levels spanning the whole 9-point valence level, and 11 other affective normative databases. Seven of these databases were partially or completely adapted from the 1,034 emotionally loaded words in the 1999 release of ANEW while the rest were either independently developed or obtained from non-English affective databases. Nevertheless, they were all affected by the role of common geographic locations and highly intertwined cultural elements. We believe that any future Turkish adaptation of ANEW or any independent work on developing a Turkish affective lexicon should consider and discuss the role of language emotionality and its effect on

excitability of native speakers from a variety of age and socioeconomic backgrounds.

To the best of our knowledge, this paper contains the first publicly available databases that provide information for cross-linguistic comparisons. Since we carried out this adaptation with a special attention to the role of sociocultural differences in variability caused by emotional perception (Soares et al. 2012), we refrained from using the existing English norms for translated words. A future application for the ATAL and ETAL ratings includes using them as the ground truth in the scope of a study on characterizing neural and electrophysiological correlates of emotional and verbal processing from affective Turkish texts, an application that will be immensely interesting for researchers and practitioners in the fields of cognition and emotion development, neuromarketing, pattern recognition, and affective computing.

Finally, the current adaptation of ANEW into Turkish focuses on medium to highly arousing words; however, while the majority of reported databases show a tendency towards words with lower emotional charges, over 60% of words in the Spanish adaptation of ANEW1999 have higher than average arousal levels. It would thus be interesting to expand these Turkish adaptations to cover all the ANEW2010 words and include other literary resources –as has been done by the Polish and Dutch linguists– to obtain insights into existence of any bias in the valence and arousal dimensions of the contemporary Turkish language. Rating dominance as an extra affective dimension and psycholinguistic variables such as familiarity, originality, imageability, and concreteness of adapted Turkish stimuli are other directions that would be regarded highly useful for studies on autism, hyperactivity disorders, schizophrenia, semantic analysis, affective computing, and electrophysiological correlates of emotional processing in the Turkish language.

Database Description

Tables S1 and S2 demonstrate the ratios of lexical classes in the Adapted Turkish Affective List (ATAL) and Revised English List (REL). Table S3 contains the list of 127 Turkish words from ATAL with valence and arousal ratings obtained from 61 native Turkish speakers. The word numbers of the equivalent words in the REL, retrieved from the ANEW 2010 dataset, the number of letters for both the English and Turkish words, case-insensitive word frequency reported as number of instances per million, and information on lexical classes are also included. Columns L and M, respectively, contain the ratio of valence ratings between 4 and 6 and arousal ratings equal to 5 for each word. Table S4 has a list

of 15 words with more than one unit difference between their average Turkish and English valence ratings while Table S5 contains the data for 40 words with over one point difference in their average arousal norms. Finally, Table S6 includes all the 508 Turkish words from the Expanded Turkish Affective List (ETAL), the ANEW words numbers for equivalents in the Expanded English List, number of letters, case-insensitive frequency for Turkish words, and the mean, standard deviation, and standard errors for the valence and arousal norms from our Turkish native speakers.

Acknowledgements The authors would like to thank Dr. Achille Pasqualotto from the Psychology Program at Sabanci University for his helpful suggestions regarding the web-based survey design and for arranging his students' participation in the experiments. The authors also extend their gratitude to all the participants and researchers in Turkey and abroad who enthusiastically devoted their time to attend the data collection sessions, and to Dr. Huseyin Ozkan for sharing his comments on the initial version of this manuscript. The first author also thanks Mr. Mostafa Mehdipour Ghazi for his inspiring discussions on the role of native and acquired languages in emotional perception.

Compliance with Ethical Standards

Conflict of interests The authors declare that there is no conflict of interest regarding the publication of this article.

Informed Consent Informed consent was obtained from all the participants included in the study.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References



- Ayçiçeği, A., & Harris, C. (2004). BRIEF REPORT: Bilinguals' recall and recognition of emotion words. *Cognition and Emotion, 18*(7), 977–987.
- Aydın Oktay, E., Balcı, K., Salah, A.A. (2015). Automatic assessment of dimensional affective content in Turkish multi-party chat messages. In *Proceedings of the international workshop on emotion representations and modelling for companion technologies* (pp. 19–24).
- Aydın, C. (2018). The differential contributions of visual imagery constructs on autobiographical thinking. *Memory, 26*(2), 189–200. <https://doi.org/10.1080/09658211.2017.1340483>.
- Bailey, K., & Chapman, P. (2012). When can we choose to forget? An ERP study into item-method directed forgetting of emotional words. *Brain and Cognition, 78*(2), 133–147.
- Beck, T.W. (2013). The importance of a priori sample size estimation in strength and conditioning research. *The Journal of Strength & Conditioning Research, 27*(8), 2323–2337.
- Bloom, L. (1998). Language development and emotional expression. *Pediatrics, 102*(Supplement E1), 1272–1277.
- Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.
- Bradley, M.M., & Lang, P.J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Tech. Rep.). Technical Report C-1, the Center for Research in Psychophysiology, University of Florida.
- Bradley, M.M., & Lang, P.J. (2010). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Tech. Rep.). Technical Report C-2, the Center for Research in Psychophysiology, University of Florida.
- Çakmak, O., Kazemzadeh, A., Yıldırım, S., Narayanan, S. (2012). Using interval type-2 fuzzy logic to analyze Turkish emotion words. In *Signal & information processing association annual summit and conference (apsipa asc), 2012 asia-pacific* (pp. 1–4).
- Chanel, G., Kronegg, J., Grandjean, D., Pun, T. (2006). Emotion assessment: Arousal evaluation using EEG and peripheral physiological signals. *Multimedia Content Representation, Classification and Security, 530–537*. https://doi.org/10.1007/11848035_70.
- Citron, F.M.M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language, 122*(3), 211–226.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. NJ: Lawrence Erlbaum Associates, 2. ISBN 978-0805802832.
- Cross, S.E., Hardin, E.E., Gercek-Swing, B. (2011). The what, how, why, and where of self-construal. *Personality and Social Psychology Review, 15*(2), 142–179. <https://doi.org/10.1177/1088868310373752>.
- Degner, J., Doycheva, C., Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition, 15*(1), 181–189.
- Eilola, T.M., & Havelka, J. (2010). Affective norms for 210 British English and Finnish nouns. *Behavior Research Methods, 42*(1), 134–140.
- Fontaine, J.R., Scherer, K.R., Roesch, E.B., Ellsworth, P.C. (2007). The World of Emotions Is Not. *Psychological Science, 18*(12), 1050–1057. <https://doi.org/10.1111/j.1467-9280.2007.02024.x>.
- Ford, B.Q., & Mauss, I.B. (2015). *Culture and emotion regulation* (Vol. 3). <https://doi.org/10.1016/j.copsy.2014.12.004>.
- Gilet, A.L., Grün, D., Studer, J., Labouvie-Vief, G. (2012). Valence, arousal, and imagery ratings for 835 French attributes by young, middle-aged, and older adults: The French Emotional Evaluation List (FEEL). *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 62*(3), 173–181.
- Gökçay, D. (2011). Emotional axes: Psychology, psychophysiology and neuroanatomical correlates. In *Affective computing and interaction: Psychological, cognitive and neuroscientific perspectives* (pp. 56–73). IGI Global.
- Gökçay, D., & Smith, M.A. (2008). TÜDADEN: Türkçede Duygusal ve Anlamsal Değerlendirmeli Norm Veri Tabanı. In *Brain-computer workshop* (p. 4).
- Gomes, C.F.A., Brainerd, C.J., Stein, L.M. (2013). Effects of emotional valence and arousal on recollective and nonrecollective recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 663.
- Grün, D., & Smith, J. (2008). Characteristics for 200 words rated by young and older adults: Age-dependent evaluations of German adjectives (AGE). *Behavior Research Methods, 40*(4), 1088–1097.
- Göz, I., Tekcan, A.İ., Erciyes, A.A. (2017). Subjective age-of-acquisition norms for 600 Turkish words from four age groups.

- Behavior Research Methods*, 49(5), 1736–1746. <https://doi.org/10.3758/s13428-016-0817-y>.
- Halfon, S., Aydin Oktay, E., Salah, A. (2016). A Assessing affective dimensions of play in psychodynamic child psychotherapy via text analysis. In *International workshop on human behavior understanding* (pp. 15–34).
- Harris, C.L., Ayçiçeği, A., Gleason, J.B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24(4), 561–579.
- Hinojosa, J.A., Méndez-Bértolo, C., Pozo, M.A. (2010). Looking at emotional words is not the same as reading emotional words: {Behavioral} and neural correlates. *Psychophysiology*, 47(4), 748–757.
- Hofstede, G. (1984). *Culture's consequences: International differences in work-related values*. SAGE.
- Imbir, K.K. (2015). Affective norms for 1,586 Polish words (ANPW): Duality-of-mind approach. *Behavior Research Methods*, 47(3), 860–870.
- Imbir, K.K. (2016). Affective norms for 4900 Polish words reload (ANPW_R): assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and, age of acquisition. *Frontiers in Psychology*, 7, 1081. <https://doi.org/10.3389/fpsyg.2016.01081>.
- Kacen, J.J., & Lee, J.A. (2002). The influence of culture on consumer impulsive buying behavior. *Journal of Consumer Psychology*, 12(2), 163–176. <https://doi.org/10.1207/153276602760078686>.
- Kapucu, A., Kılıç, A., Özkılıç-Kartal, Y., Sarıbaz, B. (2018). Turkish emotional word norms for arousal, valence, and discrete emotion categories. *Psychological Reports*, 0(0), 1–22. <https://doi.org/10.1177/0033294118814722>.
- Kaviani, H., Sagan, O., Pournaseh, M. (2015). Emotion-Related Words in Persian Dictionaries: Culture, Meaning and Emotion Theory. *International Journal of Linguistics, Literature and Culture*, 2(3), 1–11.
- Kloumann, I.M., Danforth, C.M., Harris, K.D., Bliss, C.A., Dodds, P.S. (2012). Positivity of the English language. *PLoS One*, 7(1), e29484.
- Kroupi, E., Yazdani, A., Ebrahimi, T. (2011). EEG correlates of different emotional states elicited during watching music videos. *Affective Computing and Intelligent Interaction*, 457–466. https://doi.org/10.1007/978-3-642-24571-8_58.
- Lang, P.J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. *Technology in Mental Health Care Delivery Systems*, 119–137.
- Lim, N. (2016). Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integrative Medicine Research*, 5(2), 105–109. <http://linkinghub.elsevier.com/retrieve/pii/S2213422016300191>. <https://doi.org/10.1016/j.imr.2016.03.004>.
- Lu, L., & Gilmour, R. (2004). Culture and conceptions of happiness: Individual oriented and social oriented SWB. *Journal of Happiness Studies*, 5(3), 269–291.
- Luu, S., & Chau, T. (2008). Decoding subjective preference from single-trial near-infrared spectroscopy signals. *Journal of Neural Engineering*, 6(1), 16003.
- Markus, H.R., & Kitayama, S. (1991). Culture and the Self - Implications for Cognition, Emotion, and Motivation. *Psychology and Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295x.98.2.224>.
- Mokhlesin, M., Ahadi, H., Bakhtiari, J., Ahmadzadeh, Z., Kasbi, F. (2015). Persian norms for affective dimensions and lexico semantic features of words. *Koomesh*, 17(1), 60–76.
- Montefinese, M., Ambrosini, E., Fairfield, B., Mammarella, N. (2014). The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior Research Methods*, 46(3), 887–903.
- Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Harmelen, V.an., Brysbaert, A.L.M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, 45(1), 169–177.
- Nicolaou, M.A., Gunes, H., Pantic, M. (2011). Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing*, 2(2), 92–105.
- Nigg, J.T., & Casey, B.J. (2005). An integrative theory of attention-deficit/hyperactivity disorder based on the cognitive and affective neurosciences. *Development and Psychopathology*, 17(3), 785–806.
- Oflazoğlu, Ç., & Yıldırım, S. (2013). Recognizing emotion from Turkish speech using acoustic features. *EURASIP Journal on Audio, Speech, and Music Processing*, 2013(1), 26.
- Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961–1967.
- Passarotti, A.M., Sweeney, J.A., Pavuluri, M.N. (2010). Differential engagement of cognitive and affective neural systems in pediatric bipolar disorder and attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society*, 16(1), 106–117.
- Patrick, R.E., Kiang, M., Christensen, B.K. (2015). Neurophysiological correlates of emotional directed-forgetting in persons with Schizophrenia: An event-related brain potential study. *International Journal of Psychophysiology*, 98(3), 612–623.
- Paulmann, S., Bleichner, M., Kotz, S.A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, 4, 345. <https://doi.org/10.3389/fpsyg.2013.00345>.
- Pietro, C., Silvia, S., Giuseppe, R. (2014). The pursuit of happiness measurement: A psychometric model based on psychophysiological correlates. *The Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/139128>.
- Redondo, J., Fraga, I., Padrón, I., Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior Research Methods*, 39(3), 600–605.
- Riegel, M., Wierzba, M., Wypych, M., Żurawski, Ł., Jednoróg, K., Grabowska, A., Marchewka, A. (2015). Nencki affective word list (NAWL): The cultural adaptation of the Berlin affective word list–reloaded (BAWL-R) for Polish. *Behavior Research Methods*, 47(4), 1222–1236.
- Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective, and Behavioral Neuroscience*, 9(1), 28–43.
- Sezer, T., & Sezer, B.S. (2013). TS Corpus: Herkes için Türkçe derlem. In *Proceedings of the 27th national linguistics conference. antalya. hacettepe university, linguistics department* (pp. 217–225).
- Sheldon, S., Amaral, R., Levine, B. (2017). Individual differences in visual imagery determine how event information is remembered. *Memory*, 25(3), 360–369. <https://doi.org/10.1080/09658211.2016.1178777>.
- Soares, A.P., Comesaña, M., Pinheiro, A.P., Simões, A., Frade, C.S. (2012). The adaptation of the Affective Norms for English words (ANEW) for European Portuguese. *Behavior Research Methods*, 44(1), 256–269.
- Sylvester, T., Braun, M., Schmidtke, D., Jacobs, A.M. (2016). The Berlin affective word list for children (kidBAWL): exploring processing of affective lexical semantics in the visual and auditory modalities. *Frontiers in Psychology*, 7, 969. <https://doi.org/10.3389/fpsyg.2016.00969>.
- Sze, W.P., Liow, S.J.R., Yap, M.J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, 46(1), 263–273.
- Toglia, M.P., & Battig, W.F. (1978). *Handbook of semantic word norms*. Lawrence Erlbaum.

- Trimmer, P.C., Paul, E.S., Mendl, M.T., McNamara, J.M., Houston, A.I. (2013). On the evolution and optimality of mood states. *Behavioral Sciences*, 3(3), 501–521.
- Tsai, J.L., Louie, J.Y., Chen, E.E., Uchida, Y. (2007). Learning what feelings to desire: Socialization of ideal affect through children's storybooks. *Personality and Social Psychology Bulletin*, 33(1), 17–30. <https://doi.org/10.1177/0146167206292749>.
- Uchida, Y., & Kitayama, S. (2009). Happiness and unhappiness in east and west: themes and variations. *Emotion*, 9(4), 441.
- Vecchiato, G., Cherubino, P., Maglione, A.G., Ezquierro, M.T.H., Marinozzi, F., Bini, F., Babiloni, F. (2014a). How to measure cerebral correlates of emotions in marketing relevant tasks. *Cognitive Computation*, 6(4), 856–871.
- Vecchiato, G., Toppi, J., Maglione, A.G., Olejarczyk, E., Astolfi, L., Mattia, D., Babiloni, F. (2014b). Neuroelectrical correlates of trustworthiness and dominance judgments related to the observation of political candidates. *Computational and Mathematical Methods in Medicine*.
- Vò, M.L.H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M.J., Jacobs, A.M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534–538.
- Wang, Y., Qu, C., Luo, Q., Qu, L., Li, X. (2014). Like or Dislike? Affective Preference Modulates Neural Responses to Others' Gains and Losses. *PLoS One*, 9(8), e105694.
- Warriner, A.B., Kuperman, V., Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207.
- Yılmaz, B., Korkmaz, S., Arslan, D.B., Güngör, E., Asyalı, M.H. (2014). Like/dislike analysis using EEG: Determination of most discriminative channels and frequencies. *Computer Methods and Programs in Biomedicine*, 113(2), 705–713.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Mastaneh Torkamani-Azar¹  · Sumeyra Demir Kanik¹ · Ayse Tuba Vardan^{1,2} · Cagla Aydin³ · Mujdat Cetin^{1,4} 

¹ Signal Processing and Information Systems Laboratory,
Faculty of Engineering and Natural Sciences,
Sabanci University, Tuzla, 34956 Istanbul, Turkey

² Sheldon B. Lubar School of Business,
University of Wisconsin-Milwaukee, Milwaukee, WI, USA

³ Psychology Program, Faculty of Arts and Social Sciences,
Sabanci University, Istanbul, Turkey

⁴ Department of Electrical and Computer Engineering,
University of Rochester, Rochester, NY, USA