

## Flour or flower? Resolution of lexical ambiguity by emotional prosody in a non-native language

Adriana Hanulíková, Julia Haustein

University of Freiburg, Department of German studies, Germany

Adriana.Hanulikova@germanistik.uni-freiburg.de, Julia.Haustein@gmx.de

### Abstract

It is well known that a speaker's communicative intention and his/her emotional state affect the prosodic characteristics of an utterance. Emotional prosody can function as one type of contextual cue that listeners adopt to disambiguate word meaning or to derive word meaning from novel words in their native language [1, 3, 4]. In this study we asked whether non-native speakers of English integrate emotional prosody during resolution of lexical ambiguity. Based on a vocabulary test with 32 native speakers of German, we selected a subset of the original English homophone stimuli from [3]. In a two-alternative forced-choice task, 71 native speakers of German were required to choose the meaning of an English homophone (with a happy, sad, and neutral meaning) spoken in three different affective tones (happy, sad, and neutral) that were congruent, incongruent, or neutral with respect to the affective meaning. We found a significant emotion congruency effect for sad but not for happy homophones. Despite this asymmetry, the result suggests that non-native listeners use emotional prosody during non-native lexical selection.

**Index Terms:** emotional prosody, tone of voice, lexical ambiguity, non-native listeners, cross-linguistic recognition of emotional prosody

### 1. Introduction

Prosody fulfills different functions during communication. At the linguistic level, prosody can convey information about the syntactic structure of an utterance, or signals various types of utterances such as questions, exclamations, and statements. At the paralinguistic level, prosody can convey a speaker's emotional state. Although emotional prosody is often linked to affective and social aspects of communication, several studies have shown that it conveys meaning [1, 2, 3], and thus may facilitate various aspects of speech processing such as lexical processing and naming [5, 10, 11], lexical and referential disambiguation [3, 4, 9], or the assignment of meaning to novel words [1, 6, 7, 8].

Previous studies have shown that speakers frequently modulate their emotional prosody to convey additional information about objects they are referring to, such as, among others, size [1] and speed [9], and listeners use this information to resolve referential ambiguity. For example, novel adjectives such as "daxen" were typically produced louder, slower, and with a lower pitch when they were supposed to mean *big* as compared to *small* [1]. Speakers also increased their speaking rate when they referred to fast moving objects compared to slow moving objects [9]. In a two-

alternative forced-choice task (2AFC), listeners reliably identified slow or fast-moving objects based on the speech rate.

Emotional prosody can also function as one type of several contextual cues that listeners use to disambiguate words [3, 4]. Studies on lexical ambiguity resolution have often focused on the time-course of sentential and semantic effects on the selection of an appropriate meaning to disambiguate words [3]. However, studies that include nonlinguistic aspects in spoken-word processing suggest that emotional prosody is equally relevant for the lexical-selection process [3, 4, 5, 11]. To evaluate effects of emotional prosody on the lexical selection, [3, 4] used a transcription task and presented listeners ambiguous homophones with an emotional (sad or happy) meaning and a neutral meaning (e.g., *banned/band*). The only context provided that would allow listeners to resolve the lexical ambiguity was emotional prosody (e.g., *banned/band* spoken in a sad, happy, or a neutral tone of voice). Upon hearing an ambiguous homophone, listeners were more likely to transcribe the appropriate meaning when it was spoken in a congruent affective tone (e.g., *banned* in a sad tone of voice), suggesting that they integrated emotional prosody during lexical processing and constrained the selection of word meaning. In addition, [4] found that sad participants were more likely to transcribe sad meanings than were happy participants, suggesting that a subject's emotional state modulates processing of ambiguous lexical representations.

Emotional prosody can effectively serve as a contextual cue to help resolve lexical ambiguities. We asked here whether non-native (L2) listeners, too, are sensitive to emotional prosody as a cue to the intended meaning in homophones. So far, little is known about how communicative aspects of prosody are integrated during L2 speech processing. The classroom-based acquisition of foreign languages usually focuses on vocabulary, grammar, and pronunciation. Communicative aspects of prosody are not usually covered, possibly because it is often assumed that cross-linguistic variation in the suprasegmental modulation (such as pitch, speech rate and amplitude) that conveys emotions such as sadness and happiness may be relatively limited, although this issue is still under debate [12, 13, 14, 15]. It has been suggested that both native and non-native speakers show relatively high categorization scores of emotional prosody such as happiness and sadness, and that the processing of emotional prosody (e.g., sadness and happiness) does not seem to depend on L2 proficiency [15]. According to this view, one would assume that affective prosodic markers are universal and it would be less surprising that non-native speakers can use these cues given that native speakers have been able to do so. However, while L2 listeners may be able to recognize and categorize emotional prosody well, it remains an open

question whether they are able to integrate this information during L2 lexical processing.

In this study, we employed a similar design and the original stimuli and recordings from [3], but changed two important aspects; we used only a subset of the original stimuli and selected words that German learners of English were likely to know. This was necessary, because listeners would not be able to disambiguate between the two meanings of a homophone if the words were unknown. Furthermore, instead of a transcription task, we opted for a two-alternative forced-choice task, a task that is frequently applied in the research of emotions [15]. The main reasons for choosing a 2AFC task instead of a transcription task were that a) L2 listeners' possible insecurities in homophone spelling could lead to biased responses, and b) the relative frequencies for both meanings of a homophone may differ substantially between L1 and L2 listeners, increasing biased responses for familiar or frequent words in an L2. We reasoned that a visual presentation of both meanings of a lexically ambiguous homophone following an auditory presentation would allow us to directly assess the effect of emotional prosody on ambiguity resolution and to minimize word frequency and familiarity effects in an L2. Although [3] have matched both meanings of a homophone for the relative L1 frequency use, it is difficult to assess whether the frequencies would be comparable from the perspective of an L2 listener.

We anticipated that participants' affective states could influence lexical responses. Previous research has shown that comprehenders' affective states can substantially modulate aspects of language processing; for example, semantic processing during discourse comprehension [16, 17], referential processing [18, 19], or mood incongruent word processing [4, 20, 21]. Since we wanted to keep the design as comparable as possible to [3], we did not use a mood induction procedure; instead, we asked L2 participants to rate their mood before starting the experiment.

## 2. Experiment

The main 2AFC experiment consisted of two parts: 1) listening to homophones one at a time and assigning each homophone to one of two possible meanings presented visually on a screen, and 2) a vocabulary questionnaire to determine L2 listeners' familiarity with the critical homophones. Before conducting the main experiment, a vocabulary translation test with an additional group of participants was run to determine L2 speakers' familiarity with both meanings of a homophone, and to select homophones for which each of the two meanings were correctly translated more than 50% of the time.

We predicted that if L2 speakers are able to integrate emotional prosody during lexical processing, we should obtain similar results as those for native listeners: more happy-meaning responses for the happy/neutral homophones when the words were produced in a happy tone of voice; and similarly, more sad-meaning responses for the sad/neutral homophones when the words were produced in a sad tone of voice. Based on previous research on recognition of emotions [see 15, for a review] and the results in [3, 4], we also expected the results to be slightly more pronounced for the sad/neutral homophones.

## 2.1. Methods

### 2.1.1. Participants

Participants were 103 students from the University of Freiburg, all native speakers of German. Thirty-two students (mean age 22.8, range 21-29, 7 men, average years learning English 13.2, range 9-20 years) participated in the vocabulary translation test. These students were enrolled in German language and literature studies at the department of German studies. Seventy-one students (mean age 23.8, range 20-31, 16 men, average years learning English 9.7, range 4-14 years) participated in the 2AFC experiment. All 71 participants studied English and were enrolled at the department of English studies. All participants volunteered to participate. None of the students reported hearing difficulties.

### 2.1.2. Materials

For the vocabulary translation test, both meanings of all 35 homophones (e.g., *flower/flour*) used in the original study [3] along with 17 filler (non-homophone) words were presented visually (with spellings) to native speakers of German in a randomized order. The two possible meanings of each homophone were never presented in pairs.

Based on the results of the vocabulary test, we selected 24 homophone pairs for the 2AFC experiment. We used the original recordings from [3], and only the female amateur actor from these. This speaker was a native speaker of American English (and the recordings were done in the USA for the purposes described in [3]). In the original study, all homophones were matched for frequency, and the affective meaning of each homophone pair was determined in a separate rating experiment. A detailed description of the affective homophone meaning ratings, the homophone selection, and the stimuli recordings can be found in [3].

We created three experimental lists so that each homophone occurred in only one of the three affective tones (happy, sad, neutral) per list. The order of stimuli was randomized so that tone of voice and homophone type were spread across the list. Each list contained 12 homophones with happy/neutral affective meanings (*flower/flour, ate/eight, knows/nose, rose/raws, won/one, tide/tied, medal/metal, dear/deer, peace/piece, presents/presence, heal/heel, sweet/suite*) and six with sad/neutral meanings (*bored/board, missed/mist, blue/blew, banned/band, thrown/throne, lone/loan*). We also included six filler homophones with neutral/neutral meanings (e.g., *pause/paws, choose/chews, cord/chord, heard/herd*).

### 2.1.3. Procedure

All participants were tested in small groups in a quiet seminar-room setting. Participants in the vocabulary translation test were asked to provide a German translation for each English word presented visually on a list. They were told to guess the meaning of a word in cases where they were uncertain.

For the 2AFC experiment, participants were seated in front of a large screen in a seminar room. Before starting the experiment, they had to indicate their mood on a seven-point Likert-type scale (with 1 indicating a very good mood). They were told that they would hear (over loudspeakers) words that can have two meanings, and that they would see the two possible meanings on the screen. They were required to indicate on an answer sheet which of the two meanings the

speaker intended to say. Participants had four seconds to make a decision. The experiment started with a practice trial including two example homophone items. After the experiment, they were invited to fill in a language-history questionnaire and to provide feedback on the experiment. To evaluate participants' strategic responses, we asked them whether they were aware of the presence of homophones and their emotional meanings.

## 2.2. Results

### 2.2.1. Vocabulary translation study

All translation responses were coded manually; each response was categorized either as an incorrect translation, no response, a correct translation, a related meaning, or a translation of the other homophone (e.g., *Blume* 'flower' when presented with *flour*). We excluded all homophone pairs that obtained more than 50% incorrect responses (no response and a translation of the other homophone counted as incorrect). Pairs for which one of the two meanings was incorrectly translated more than 50% of the time were also excluded (e.g., *tow/toe*, *hall/haul*, *bridle/bridal*, *petal/pedal*, *die/dye*, *groan/grown*, *pain/pane*, *poor/pore/pour*, *hair/hare*). We also excluded cognates or similarly sounding words in German (e.g., *caller/collar*, *medal/metal*). Based on these exclusion criteria, a set of 12 happy/neutral and six sad/neutral homophones remained for the 2AFC experiment.

### 2.2.2. Lexical disambiguation study

As can be seen in Table 1 (overall proportions) and Figure 1 (estimated effects), non-native speakers of English were more likely to choose the sad meaning of a sad/neutral homophone pair when it was presented in a sad tone of voice. This replicates the effect observed for native English speakers in [3]. However, unlike in [3], a less clear effect of tone of voice on lexical disambiguation is visible for the happy homophones.

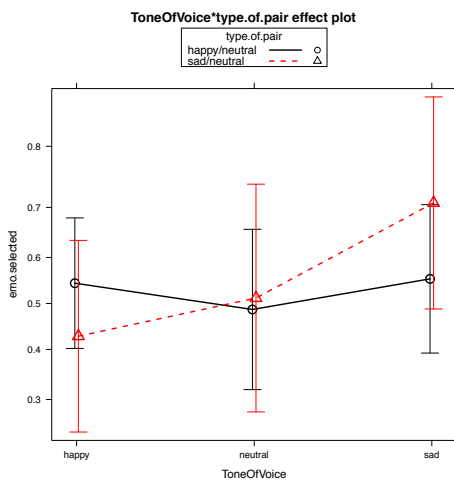


Figure 1: Estimates of emotional homophone choices (*emo.selected*) as a function of Tone of Voice and Homophone Type (*type.of.pair*). The error bars represent 95% confidence intervals.

To control for subject and item level variability in the data analysis, a generalized mixed-effects logistic regression model (implemented in R package *lme4*, [22]) for binary responses

was fitted. The proportion of selected emotional meaning per homophone pair was taken as the dependent variable. Treatment coding was used on the independent variables Tone of Voice (happy, neutral, and sad) and Homophone Type (happy/neutral and sad/neutral). The model includes random intercepts for listeners and items, and random slopes for Tone of Voice (by items). A model including random slopes for mood (by items) as well as the interaction of Tone of Voice and Homophone Type (by subjects) did not converge. The summary results of the final model are shown in Table 2. The results show that the interaction between Tone of Voice and Homophone Type is driven by the responses to the sad/neutral homophones in the sad tone of voice. We then applied Anova (type III Wald chisquare tests) to the model and found a significant main effect of Tone of Voice ( $X^2(2) = 6.81, p = 0.03$ ) and a significant interaction between Tone of Voice and Homophone Type ( $X^2(2) = 6.11, p < 0.05$ ).

Table 1. Percentage of emotional homophone choices for all participants and separately for participants in a happy, sad, and neutral mood (*ToV* = Tone of Voice).

	Happy/Neutral Homophones			Sad/Neutral Homophones		
	ToV: happy	neutral	sad	happy	neutral	sad
Good mood	55.2	43.9	52.5	42.7	49.4	68.3
Sad mood	58.3	56.3	66.0	41.7	37.5	56.5
Neutral mood	45.8	52.8	48.6	44.4	58.3	65.7
<b>All (n = 71)</b>	<b>53.0</b>	<b>48.2</b>	<b>54.1</b>	<b>44.4</b>	<b>49.6</b>	<b>65.7</b>

Table 2. Mixed-effects model result summary with coefficient estimates  $\beta$ , standard errors *SE*, z-scores and p values (*ToV* = Tone of Voice, *HType* = Homophone Type, *neut* = neutral).

Predictor	$\beta$	SE	z	p
Intercept	0.162	0.288	0.56	0.57
ToV neut	-0.243	0.261	-0.93	0.35
ToV sad	0.005	0.265	0.02	0.99
HType sad/neut	-0.509	0.484	-1.05	0.29
Mood neut	-0.008	0.161	-0.05	0.96
Mood sad	0.174	0.189	0.92	0.36
ToV neut*HType sad/neut	0.608	0.447	1.36	0.17
ToV sad*HType sad/neut	1.139	0.464	2.46	0.01

Note: Formula: *emo.selected* ~ *ToneOfVoice* \* *type.of.pair* + *mood* + (1 | *VP*) + (*ToneOfVoice* | *Item*)

Separate models restricted to pairwise comparisons for the sad/neutral and happy/neutral homophone pairs showed that the percentage of sad homophones chosen was significantly larger in the sad tone of voice relative to the happy tone of voice ( $\beta = -1.176, z = -2.51, p = 0.012$ ) or to the neutral tone of voice ( $\beta = -0.898, z = -2.41, p = 0.016$ ), but there was no significant difference between the happy tone of voice and the neutral tone of voice. The choices within the happy/neutral homophone pairs did not significantly differ from each other. This result indicates that non-native listeners detect the relevance of emotional prosody for lexical disambiguation, but they do so reliably only in the sad tone of voice. Similar

asymmetric patterns have been previously reported [e.g., 3, 4], and we will come back to this asymmetry in the discussion.

To assess whether participants' mood modulated the integration of affective prosody during lexical meaning disambiguation, we asked participants to indicate their mood on a seven-point Likert-type scale ahead of the experiment. Based on this information, three groups emerged (see Table 1 for percentage of homophone disambiguations for each mood group): 40 participants in a happy mood (1-3 on the scale), 18 participants in a neutral mood (4 on the scale), and 12 participants in a sad mood (5-7 on the scale). However, since no main effect of mood emerged and the analysis is based on subjective ratings provided by the participants, further research would be necessary to evaluate the nature of mood effects on L2 lexical disambiguation.

We further analyzed the vocabulary questionnaire and listeners' feedback on the experiment. In the vocabulary questionnaire, 27 participants indicated that they did not know some of the homophones. Across all items for this group of participants, 51 words were marked as unknown, the majority of which ( $n = 36$ ) were fillers (neutral/neutral homophones). The remaining 15 critical unknown items consisted of the following words: *lone* ( $n = 4$ ), *loan* and *mist* (each three times), *rows* (two times), and *thrown*, *heel*, *won* (each once). This suggests that L2 participants' knowledge of the homophone lexical meanings was very high and cannot explain the lack of a congruency effect for happy/neutral homophones.

In the feedback requested from participants, only six participants indicated that they did not realize that homophones had been used, and 15 reported that they did not realize that the homophones had emotional meanings. In describing how they made their choices, they chose between deciding intuitively ( $n = 29$ ), based on the prosody ( $n = 6$ ), based on the word frequency ( $n = 5$ ), and based on a match between the pronunciation and the meaning ( $n = 15$ ); the remaining participants did not provide an answer. When asked to guess the aim of the experiment, participants indicated that the study examined effects of intonation ( $n = 16$ ), emotions ( $n = 16$ ), and mood ( $n = 8$ ) on word selection. Other responses included the study of gender, frequency, word recognition, comprehension, minimal pairs, associations, learning processes, and language competence. None of the participants reported being aware of the presence of one emotional and one neutral meaning of a given homophone, or the link between a homophone meaning and emotional prosody. We therefore assume that participants did not develop a specific strategy that would explicitly match the experiment's demands.

### 3. Discussion

In a 2AFC experiment, we used homophones with either a happy/neutral emotional meaning (e.g., *flower/flour*) or a sad/neutral emotional meaning (e.g., *banned/band*) to examine whether German listeners make use of emotional tone of voice to resolve lexical ambiguity in English single words. For sad/neutral homophones, results were, as predicted, that a sad tone of voice led to more sad homophone choices than a happy or neutral tone of voice. For happy/neutral homophones, however, no clear effect of tone of voice on lexical disambiguation emerged.

There are several possible explanations for why we observed an asymmetry in the modulation of lexical

disambiguation by a sad and a happy tone of voice. First, studies conducted on various languages suggest that sad emotional prosody is frequently more accurately recognized than happy emotional prosody [see 15, for a review]; and our result extends this tendency to lexical processing. A similar asymmetric pattern of results emerged for native speakers of English in [3], with disambiguation effects being clearer for sad/neutral homophones.

Second, it is possible that L2 participants did not activate the same emotional meaning for some of the English homophones which would have made them comparable to the emotional meaning ratings provided by L1 speakers. It has been previously suggested that emotional meaning associations may even vary among native speakers [4]; it is therefore reasonable to assume that cross-cultural differences would play a role. For example, the word *hymn*, associated with happy emotions by native speakers in [3], may be associated with a neutral meaning and would then not differ from the neutral homophone *him*. Future cross-linguistic studies should collect associations with English homophones to make sure that the appropriate emotional meanings are activated.

Last but not least, it is possible that participants' moods contribute to asymmetric patterns in the use of emotional prosody. However, we only collected subjective mood ratings and did not expose participants to a happy or a sad mood induction; different results might be obtained with a systematic mood manipulation [see 4].

Taken together, these findings indicate that non-native speakers are able to make use of emotional prosody to disambiguate homophones in their L2, in particular for sad/neutral homophones. Put differently, L2 listeners can access meanings of ambiguous words that are congruent with affective properties of the speakers' utterances. However, unlike L1 listeners, they may be slightly less effective in constraining their selection of word meaning. While L2 listeners detect the meaning associated with the emotional prosody, the activation spread to emotion-related lexical representations may require increased computational effort (it usually takes longer to access an L2 than an L1 lexical representation) as well as increased exposure to prosodic information of a given speaker and a given non-native language. Further research is needed that establishes a) cross-linguistic affective ratings for each pair of homophones, and b) effects of mood on L2 word processing by manipulating participants' emotional states. Furthermore, future studies could address how presentation of single words with different prosody will scale up in a sentential context, and consider the use of online tasks.

### 4. Conclusions

We showed that non-native listeners are able to use emotional prosody for lexical disambiguation, but they do so mainly for sad/neutral homophones. The present result partly replicates previous research on the role of emotional prosody in the resolution of lexical ambiguity [3, 4] and extends this research to a non-native population. This result forms the basis for further research into the role of emotional prosody on lexical processing and ambiguity resolution in L2 listeners. A successful integration of contextual cues such as emotional prosody could be a beneficial tool for vocabulary learning, as it would facilitate the mapping process between labels and referents.

## 5. Acknowledgements

We would like to thank Lynne Nygaard for providing her original stimuli and recordings, Lars Konieczny for statistical advice, and the reviewers for very helpful and constructive feedback.

## 6. References

- [1] L. C. Nygaard, D. S. Herold, and L. L. Namy, "The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning", *Cognitive Science*, vol. 33, pp. 127-146, 2009.
- [2] M. D. Pell, A. Jaywant, L. Monetta, and S. A. Kotz, "Emotional speech processing: Disentangling the effects of prosody and semantic cues", *Cognition and Emotion*, vol. 25, no. 5, pp. 834-853, 2011.
- [3] L. C. Nygaard and E. R. Lunders, "Resolution of lexical ambiguity by emotional tone of voice," *Memory & Cognition*, vol. 30, no. 4, pp. 583-593, 2002.
- [4] J. B. Halberstadt, M. Niedenthal, J. Kushner, "Resolution of lexical ambiguity by emotional state", *Psychological Science*, 6, pp. 278-282, 1995.
- [5] L. C. Nygaard and Queen, "Communicating emotion: linking affective prosody and word meaning", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 34, no. 4, pp. 1017-1030, 2008.
- [6] D. S. Herold, L. C. Nygaard, K. Chicos, and L. L. Namy, "The developing role of prosody in novel word interpretation", *Journal of Experimental Child Psychology*, 108, pp. 229-241, 2011.
- [7] L. C. Nygaard, A. E. Cook, L. L. Namy, "Sound to meaning correspondences facilitate word learning", *Cognition*, 112, pp. 181-186, 2009.
- [8] E. Reinisch, A. Jesse, and L. C. Nygaard, "Tone of voice guides word learning in informative referential contexts", *Quarterly Journal of Experimental Psychology*, vol. 66, no 6, pp. 1227-1240, 2013.
- [9] H. Shintel, H. C. Nusbaum, and A. Okrent, "Analog acoustic expression in speech communication", *Journal of Memory and Language*, 55, pp. 167-177, 2006.
- [10] A. Schirmer, A. S. Kotz, and A. D. Friederici, "Sex differentiates the role of emotional prosody during word processing", *Cognitive Brain Research*, 14, pp. 228-233, 2003.
- [11] A. Schirmer, A. S. Kotz, and A. D. Friederici, "On the role of attention for the processing of emotions in speech: Sex differences revisited", *Brain Research*, vol. 24, no. 3, pp. 442-452, 2005.
- [12] M. D. Pell and V. Skorup, "Implicit processing of emotional prosody in a foreign versus native language", *Speech Communication*, vol. 50, no 6, pp. 519-530, 2008.
- [13] D. A. Sauter, F. Eisner, P. Ekman, and S. K. Scott, "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations", *PNAS*, vol. 107, no. 6, pp. 2408-2414, 2009.
- [14] K. Scherer and H. G. Wallbott, "Evidence for universality and cultural variation of differential emotional response patterning", *Journal of Personality and Social Psychology*, vol. 66, no. 2, pp. 310-328, 1994.
- [15] H. Bąk, *English Emotional Prosody in the native and non-native mind*, Doctoral Dissertation at the University Poznań, 2015.
- [16] G. Egidi and H. C. Nusbaum, "Emotional language processing: how mood affects integration processes during discourse comprehension", *Brain and Language*, vol. 122, no 3, pp. 199-210, 2012.
- [17] G. Egidi and H. C. Nusbaum, "How valence affects language processing: negativity bias and mood congruence in narrative comprehension", *Memory and Cognition*, 37, pp. 547-555, 2009.
- [18] J. J. A. van Berkum, D. de Goede, P. M. van Alphen, E. R. Mulder, and J. H. Kerstholt, "How robust is the language architecture? The case of mood", *Frontiers in Psychology*, vol. 4, art. 505, 2013.
- [19] C. T. Vissers, D. Virgillito, D. A. Fitzgerald, A. E. Speckens, I. Tendolkar, I. van Oostrom, et al., "The influence of mood on the processing of syntactic anomalies: evidence from P600", *Neuropsychologia*, 48, pp. 3521-3531, 2010.
- [20] M. Kiefer, S. Schuch, W. Schenck, and K. Fiedler, "Mood states modulate activity in semantic brain areas during emotional word encoding", *Cerebral Cortex*, 17, pp. 1516-1530, 2006.
- [21] N. L. Pratt and S. D. Kelly, "Emotional states influence the neural processing of affective language", *Social Neuroscience*, 3, pp. 434-442, 2008.
- [22] D. M. Bates and M. Mächeler, "lme4: Linear mixed-effects models using Eigen and Eigen++, R package version 0.999999-0, 2009.