



Behavioral evidence for differences in emotional resonance during processing first and second language

Anna Weimer, Ina Koniakowsky, Tatjana Nazir, Anke Huckauf

► To cite this version:

Anna Weimer, Ina Koniakowsky, Tatjana Nazir, Anke Huckauf. Behavioral evidence for differences in emotional resonance during processing first and second language. *International Journal of Bilingualism*, 2022, 26 (6), pp.784-798. 10.1177/13670069211073578 . hal-03890723

HAL Id: hal-03890723

<https://hal.science/hal-03890723>

Submitted on 10 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Behavioral evidence for differences in emotional resonance during processing first and second language

International Journal of Bilingualism
2022, Vol. 26(6) 784–798
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13670069211073578
journals.sagepub.com/home/ijb



Anna Weimer* and **Ina Koniakowsky*** 

Ulm University, Germany

Tatjana A. Nazir

Univ. Lille, CNRS, UMR 9193—SCALab—Sciences Cognitives and Sciences Affectives, France

Anke Huckauf

Ulm University, Germany

Abstract

Aims and objectives/purpose/research questions: Currently available data show mixed results as to whether emotional resonance is stronger for words expressed in the mother tongue (L1) compared to a second language, acquired later in life (L2). One reason for these discrepancies could be differential effects of individuals' L2 learning history. We introduce an experimental paradigm that is sufficiently robust for testing outside the laboratory to reach a more diverse population. We illustrate this paradigm using 24 well-characterized Russian (L1)–German (L2) bilingual migrants.

Design/methodology/approach: The paradigm consists of displaying an array of random letters that may contain a word, which participants must identify. Stimuli are displayed until response and the proportion of correct identification is used as dependent measure. Performance for neutral words is contrasted to swear or taboo words.

Data and analysis: The interplay between language and word type is assessed with a 2×2 within subjects ANOVA.

Findings/conclusions: At the group level, a swear or taboo word superiority in L1 and its absence in L2 is observed. At the individual level, however, the data show a clear divide depending on the age of arrival at the L2 country. Participants who arrived after mid-adolescence show a clear language effect. By contrast, individuals who arrived earlier, present a swear or taboo word superiority in either L1, L2, or in both languages. The age of arrival should therefore be regarded as a critical variable and averaging over bilinguals with different ages of arrival can distort the results depending on the relative size of the respective groups.

*Anna Weimer and Ina Koniakowsky shared first authorship—authors contributed equally to this article.

Corresponding author:

Ina Koniakowsky, General Psychology, Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 47, 89081 Ulm, Germany.

Email: ina.koniakowsky@uni-ulm.de

Originality: The representativeness of test subjects is constrained by the availability of participants at the testing site. Testing outside the laboratory, at home or online, allows reaching larger and/or target populations.

Significance/implications: By removing constraints on the availability of bilingual participants, our paradigm enables refined insights into how emotion shapes language processing.

Keywords

Emotion, second language, swear/taboo words, word detection, behavioral measures

Introduction

Language conveys both meaning and emotion, which is particularly evident in the case of swear words. Children, for instance, seem to use such emotionally charged expressions even before knowing their exact literal meaning. But when learning a foreign language, the correct use of swear words is among the most perilous challenges. Interestingly, theories of lexical processing in bilinguals generally stipulate that information concerning both languages are activated in parallel during processing either of the two languages (for a review, see for example, Kroll & Ma, 2018). One of the currently prevalent theories, the Bilingual Interactive Activation plus (BIA+) model by Dijkstra and van Heuven (2002) assumes, for instance, that bottom-up cross-language processes are engaged from stimulus onset, and that these processes activate (common) semantic representations alongside representations that indicate membership to a particular language. Moreover, following work on the affective value of pictures and words, a word's affective value has been conceptualized as a tag (e.g., *positive*, *negative*) that is associated with conceptual information in the semantic system (e.g., De Houwer & Hermans, 1994; Glaser & Glaser, 1989; Sianipar et al., 2015). Thus, within such a theoretical framework, the affective representations of words in the first language (L1) and in a second language (L2)¹ are linked through a common semantic system and should therefore not differ. However, while some studies that used psychophysiological or behavioral measures support this notion, there are indications that *emotional resonance*² differs between L1 and L2. As we will detail below, in particular, self-reports suggest that processing words in L1 is more strongly modulated by emotional factors than processing the same words in L2. Such discrepancies in findings may result from the different experimental tasks used in these studies, but also from the small size of the tested populations, which may differ in their individual L2 learning history (operationalized, for example, by the age of L2 acquisition [AoA_{L2}]). The aim of this study is therefore to introduce a behavioral paradigm that allows studying the affective value of words in L1 and L2 outside the laboratory (e.g., at home or via the internet) to reach a larger and more representative sample of bilinguals. In what follows, we will provide a brief non-exhaustive summary of studies that aimed at capturing potential differences in emotional resonance between L1 or L2. Then, we will introduce our behavioral paradigm and report results from a well-characterized population of Russian–German bilingual migrants that were tested with this paradigm at their respective homes. Finally, we will shortly discuss potential implications of our findings for models of lexical access in bilinguals.

Behavioral measures

Behavioral measures such as reaction time and response accuracy to isolated words can provide information about the degree of emotional resonance during L1 and L2 processing. For example,

interference effects such as the Stroop effect (Williams et al., 1996) have been used to investigate selective attention to emotional information. In the emotional Stroop task, participants are asked to name the ink color of words with emotional or neutral valence. Typically, emotional words trigger an interference effect, resulting in an increase in reaction times (Ben-Haim et al., 2016; Williams et al., 1996). In bilinguals, this paradigm has provided mixed findings: Sutton et al. (2007) reported that Spanish (L1)–English (L2) bilinguals ($AoA_{L2} > 5$) with L2 dominance responded faster to neutral words than to emotional words. Yet, while the effect size for interference effects from emotional words was larger in the more proficient L2 than in L1, the interaction between word type and language did not reach significance. Similarly, Eilola et al. (2007) also found an emotional Stroop effect in both, L1 and L2 (AoA_{L2} between 7 and 13). The missing interaction between word type and language was attributed to the high proficiency of their participants in both languages. In a follow-up study, in addition to the Stroop task, electrodermal activity (EDA) was monitored (Eilola & Havelka, 2011). The results from the emotional Stroop task replicated the earlier findings in that no language specific effects were found. However, there was a selective enhancement of EDA to L1 swear words that was not seen in L2 ($AoA_{L2} > 6$). The emotional Stroop task thus fails to systematically capture effects that can be seized through EDA analysis, suggesting that the Stroop task is not suitable for investigating emotional resonance during language processing in bilinguals.

Another approach to capture emotional resonance to taboo or swear words was proposed by Colbeck and Bowers (2012). The researchers asked monolingual English speakers and bilingual (Chinese (L1)–English (L2); no information about AoA_{L2}) speakers to indicate the presence of a pre-specified target word within a list of distractors (i.e., pseudo words, swear words, neutral words). Compared to neutral words, the presence of a swear word in the list of distractors impaired the detection of the target word. This interference was significantly stronger in mono- than bilinguals, testifying language specific effects.

Incera et al. (2020) used a computer mouse-tracking paradigm in an auditory lexical decision task to examine taboo word effects in participants for whom English was L1³ and participants for whom English was L2 (mean $AoA_{L2} = 7.6$, no information about range of AoA_{L2}), using English taboo and neutral words. The researchers found an interaction between language and word type in the number of errors that participants made in the lexical decision task, in the profile of the mouse trajectories, but not in the reaction times. The lack of experimental effects, as noted in some of the previously discussed studies, might therefore be related to the sensitivity of the dependent measure that was used.

Psychophysiological approaches (EDA, ERP, and fMRI)

EDA, providing information about changes in the level of sympathetic arousal (e.g., Braithwaite et al., 2013), has also been used in this domain. Swear or taboo words elicit stronger EDA than neutral words (for a review, see Harris et al., 2006). Bowers and Pleydell-Pearce (2011) further specified that the sound structure of the verbal stimuli (cf. phonological information) could play a critical role in this amplification. By comparing the processing of swear words (e.g., “fuck”) with their semantically equivalent euphemism (“f-word”), the researchers revealed stronger EDA for the phonologically more familiar swear words. Harris et al. (2003) compared EDA in a bilingual Turkish (L1)–English (L2; $AoA_{L2} > 16$) population using written and spoken taboo words and childhood reprimands. Larger EDA responses in L1 than in L2 were observed in the auditory modality for childhood reprimands, but not for taboo words. The stronger emotional resonance triggered by spoken over written language in L1 was attributed to the fact that spoken language and respective emotional associations are acquired before written language. In a group of Spanish (L1)–English (L2) bilinguals, Harris (2004) showed that early bilinguals, for whom English was

the dominant language, displayed a similar pattern of EDA responses in both languages. By contrast, bilinguals arriving in the L2 country after the age of 11 years showed the hypothesized stronger EDA for childhood reprimands in L1 and its absence in L2. This suggests that the decline in emotional resonance during language processing occurs as age of L2 acquisition increases and language proficiency decreases (Harris et al., 2006).

While EDA reflects a rather unspecific cognitive or emotional arousal, electro-encephalography (EEG) can provide more specific information (for a review, see Kissler et al., 2006). For instance, compared to neutral words, words of high emotional valence generally trigger a stronger modulation of a prototypical emotion-related ERP component, that is, the EPN (Early-Posterior-Negativity, for example, Conrad et al., 2011; Kissler et al., 2007; Ortigue et al., 2004). The EPN peaks around 200–300 ms after stimulus onset over temporo-occipital regions and is elicited by various kinds of emotional stimuli (for a review, see Schindler & Bublatzky, 2020). In German–Spanish bilinguals ($AoA_{L2} > 12$ years), Conrad et al. (2011) found a general delay in the peak of the EPN in L2. However, they did not find evidence for the predicted interaction between language and the words' emotional valence. Opitz and Degner (2012), who examined the EPN to emotional words in German–French bilinguals (AoA_{L2} between 7 and 16 years), showed that emotional words triggered a larger EPN than neutral words in both languages. However, the peak amplitude in L2 occurred delayed compared to L1, which was interpreted as indicating that lexical access to emotional words in L2 could be subject to interferences from the L1 lexicon. Finally, Chen et al. (2015), who analyzed the EPN to emotional words in Chinese–English bilinguals ($AoA_{L2} > 7$ years), showed the predicted effect of emotional valence in L1 and its absence in L2. While participants' AoA_{L2} and/or L2 proficiency in these three studies cannot be directly compared, the data nonetheless suggest that one or the other of these two factors might affect the degree of emotional resonance during language processing.

Some studies also used fMRI (functional Magnetic Resonance Imaging) to capture potential differences in emotional resonance in L1 and L2. Hsu et al. (2015), for instance, monitored hemodynamic responses while German–English bilinguals (medium to high L2 proficiency; AoA_{L2} between 6 and 14 years) read text passages with happy or neutral content. Stronger responses to happy passages were seen for L1 but not for L2 in bilateral amygdala and the left precentral cortex. Such involvement of the limbic system (of which the amygdala is part of, for example, Kober et al., 2008) was also reported by Sulpizio et al. (2019). They compared the processing of taboo and non-taboo words in proficient Italian–English bilinguals (mean $AoA_{L2} = 6.36$ years; no information about range of AoA_{L2}) and revealed language specific modulations of brain activity in the anterior cingulate gyrus. Like the amygdala, the anterior cingulate serves affective regulation. However, unlike Hsu et al. (2015), Sulpizio et al. (2019) reported a stronger involvement of the limbic system during processing L2 instead of L1. Finally, in the previously mentioned study by Chen et al. (2015) with Chinese–English bilinguals, an increased activation for emotional L2 words was observed in the left cerebellum, which also regulates emotion through its connection to the limbic system (see Baumann & Mattingley, 2012). In all three studies, brain regions serving affective regulation are thus involved differently in the processing of emotional words in L1 and L2.

In short, most studies that used physiological measures indicate differences in emotional resonance in L1 and L2 processing. However, the reported results cannot be easily summarized because of the variability in the type of stimuli that have been used (e.g., happy vs. neutral words; taboo vs. non-taboo words) and differences in participants' L2 acquisition history (i.e., AoA_{L2} and language proficiency).

Questionnaires and self-report approaches

Evidence in favor of reduced emotional resonance in L2 also comes from the data collected with the “Bilingualism and Emotions Questionnaire” (Dewaele & Pavlenko, 2001–2003). In this largest online survey to date on bilingualism and emotions, Dewaele and Pavlenko collected data from over 1,500 multilinguals including participants’ answers to eight questions concerning the relationship between language-use and emotion. One interesting finding from this study is that participants were more likely to report that they use L1 to express positive and negative affect, such as in emotional self-talk (Dewaele, 2015) or anger expression (Dewaele & Qaddourah, 2015). This database also allowed revealing that participants attributed a greater emotional force to swear or taboo words presented in their L1 compared to their L2 (Dewaele, 2004). Moreover, AoA_{L2}, self-rated proficiency, and frequency of use were identified as predictors for the perceived emotional force of swear words in L2 (Dewaele, 2004). Caldwell-Harris et al. (2012), who looked at predictors of perceived emotionality in the mother tongue of Russians immigrated to the United States (English) found that self-reported emotional resonance for L1 was higher for individuals that arrived late compared to those that arrived early in the United States. Since age of arrival constrains language-use and proficiency, the authors suggested that in judging the emotional quality of a language, these factors are more important than the age at which the language was acquired (see also Caldwell-Harris, 2014).

In summary, introspective methods consistently provide evidence for differences in emotional resonance during language processing in L1 and L2, and point to factors related to AoA_{L2}, acquisition context, language proficiency, and language-use as potential critical variables.

Research goal

Currently available data from the literature show mixed results with respect to the question of whether emotional resonance during word processing differs between L1 and L2. We believe that such a discrepancy in findings results from the heterogeneity of participants’ L2-acquisition histories. Hence, the smaller the sample, the greater the likelihood that idiosyncratic patterns of L2 learning will affect the results. The goal of this study is therefore to introduce an experimental paradigm that is sufficiently robust for testing outside the laboratory (at home or via the Internet) to reach more diverse samples of bilinguals who use both languages in daily communication. The paradigm involves displaying an array of random letters that may contain a word, which participants are asked to identify. When the array contains a word, the word can have a specific emotional value (i.e., a taboo/swear word or a neutral word). Stimulus display remains on the screen until participants respond *yes* or *no*. This self-paced design helps to minimize confounding variables related to inattention and to perceptual issues, which are related to short visual displays. In addition, since the arrays of letters are sufficiently complex, the number of correct answers—instead of reaction times—can be used as dependent measure. With this paradigm, we seek to reveal differences in processing taboo or swear words in L1 and L2, through the testing of a well-characterized sample of bilinguals.

Methods

Participants

A sample of 24 Russian immigrants (11 female) living in Germany, with Russian as L1 and German as L2, aged between 19 and 61 years ($M=39.2$, $SD=13.5$) participated in the study. Participants

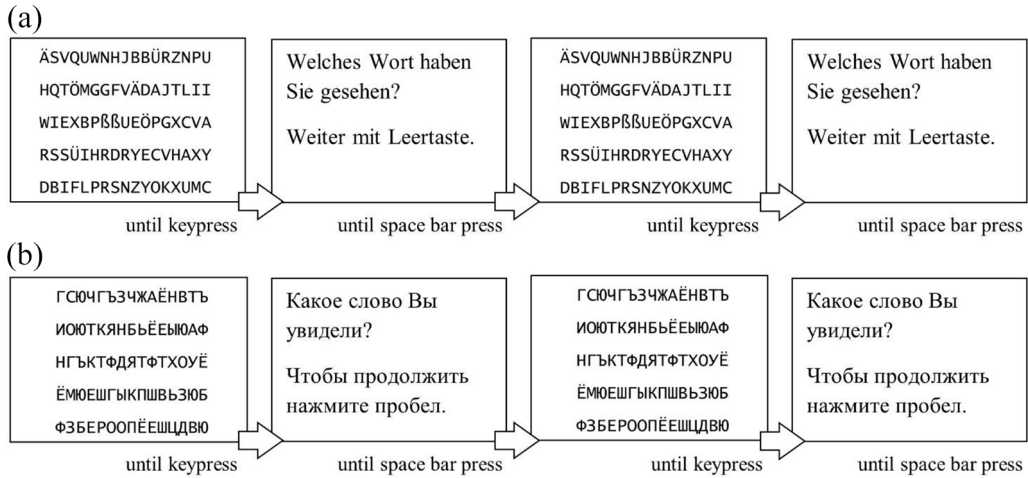


Figure 1. Examples of two trials in the word detection task with German (a) and Russian (b) instructions. Half of the participants were instructed in German, the other half in Russian. The instruction text between the trials says, “Which word did you see? Press space bar to continue.”

reported to use both languages in writing and speaking. They were recruited from the experimenter’s social circle, in the university and on social media. Seven participants indicated their L1 as dominant language, eight L2, and nine described L1 and L2 as equally dominant. Age of arrival⁴ in L2 (AoA_{L2}) ranged between 2 and 42 years ($M=20.0$, $SD=13.0$). All immigrated from a Russian-speaking country. The average time they lived in Germany was 19.3 years ($SD=5.9$). Except for one participant, L2 learning in a natural context was accompanied by some instructed learning (school or language courses). Detailed descriptive data of participants and self-ratings of language proficiency are provided in the Supplemental material. For their participation in the study, participants were rewarded either with 10€, with sweets, or with a one university course credit. Two participants were subsequently excluded from the analyses due to insufficient Cyrillic L1 reading skills, defined by extremely slow reaction times in L1 trials ($>2 SD$ from the mean). Results are thus based on a sample of $n=22$ participants.

Word detection task

The word detection task was modified from Hamada (2017) and was implemented in PsychoPy version 3.2.4 (Peirce et al., 2019). This behavioral paradigm allows displaying stimuli for an unlimited viewing duration and to measure non-time-critical hit rates. Measuring response accuracy makes the paradigm suitable for participants of all age groups, regardless of the devices used, and allows testing in less controlled environments. Therefore, it can be easily adopted to testing at home or online.

A stimulus display consisted of five lines of random sequences of upper-case letters plotted in the *Lucida Console* proportional font (see Figure 1). The sequences of upper-case letters were created with a random letter generator. Words created by chance were eliminated manually by swapping the letters. When a target word was present in the array, it appeared in one of the three center lines, either at the beginning (starting at the third letter position in the line), in the middle or at the end of the line (ending at the third letter position before the end of the line). Since on average, L2

words were two letters longer than L1 words, a line consisted of 17 characters in L2 and of 15 characters in L1. One-third of the stimulus displays did not contain a target word (i.e., distractor trials). Participants' task was to decide as quickly and correctly as possible whether a word was present or not by pressing one of two predefined keys on the keyboard (right button *yes*, left button *no*). After each keystroke, participants had to either name the detected word or report that no word had been detected. To name taboo words, participants could choose to utter the entire word or to say only the first letter of the word. The experimenter was present throughout the whole experiment and recorded the verbal responses. After the verbal response, participants triggered the next trial by pressing the space bar (see Figure 1).

Word stimuli (L1 and L2)

Four L1 taboo words of comparable levels of tabooeness were selected from Golodnaya (2017). All words are considered as strongly vulgar, and all form the stem of many other Russian swear words which implies a high degree of familiarity. L1 taboo words were matched with L2 words of comparable levels of tabooeness and corpus entries (i.e., frequency of occurrence). For L1, the Russian National Corpus (2003–2021) was used as reference, for L2 the Leipzig Corpora Collection (2018).

Due to their rare use in literary works and official media, taboo words have generally a low frequency of entries in text corpora. Neutral control words were therefore matched for frequency, as well as for word length, number of syllables, and number of phonemes. A full list of stimuli and their corpus entrance frequency is provided in the Supplemental material.

Design

The word detection task was based on a 2×2 within subjects factorial design with language (L1 vs. L2) and word type (taboo vs. neutral) as factors. The experiment included a total of 216 trials (108 per language). Two-thirds of the trials contained a word (target trials; 72 per language) and one-third contained no words (distractor trials; 36 per language). Each word was presented once in each of the $3 \times 3 = 9$ possible positions in the array. To avoid sequence effects, three lists of pseudo-random sequences were created. The same number of participants saw each of the three lists. The proportion of correct target detection was used as dependent variable. In addition, an implicit measure of the level of taboo-shame was assessed, defined by whether participants chose to utter the taboo words aloud or to name just the first letter of the words. We also monitored reaction time, measured from the onset of the stimulus array until the response key was pressed.

Procedure

All details of the Helsinki declaration were applied to the study. Figure 2 summarizes the experimental procedure. Participants were first informed about the purpose of the experiment and that they may see offensive words. They were also informed that they could stop the experiment at any time without penalty. Once fully informed, participants gave their written consent for their participation. The experimenter tested participants individually in a quiet room at their respective homes, isolated from other people. A 15.6-inch HD notebook (HP 250 G7) was used for testing. Testing started with an online version of the German LexTALE (Lexical Test for Advanced Learners of English) to determine L2 proficiency. This was followed by a target word knowledge test, adopted from Drijvers and Özyürek (2020) and implemented in PsychoPy. The test contained 16 target words and 8 pseudo words that were created using a random word generator and/or by swapping single letters in existing words. Stimuli were presented one by one in a random order and

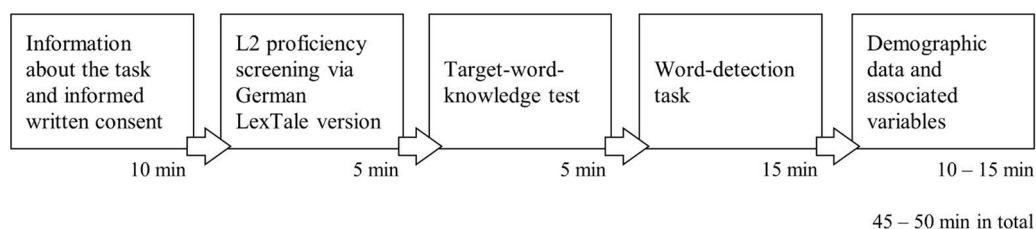


Figure 2. During the first 10 minutes, information about the task and written consent were given. Next, language proficiency was assessed through the German LexTale version. The target-word-knowledge test took about 5 minutes and was followed by the word-detection task. Finally, participants filled out a questionnaire on demographic data and associated variables.

participants had to decide whether the presented word was an existing word or not. The ensuing word detection task of the main experiment was preceded by 30 practice trials with other stimuli than those used in the main experiment. The experiment started when participants felt comfortable with the task. On average, the word detection task took 15 minutes. Participants could take a break at any time during the experiment.

A questionnaire was used to collect demographic data and associated variables. This included a self-assessment of language ability in L1 and L2 (self-rating of L1/L2) and an estimation of the proportion of L2 use in relation to L1 use. Demographic data on the proportion of life spent in Germany in relation to total life span and the duration of education in Germany in years was also collected. In addition, participants were asked to rate their use of taboo words on a scale of 0–4 ranging from “(almost) never” to “every hour.” The overall study lasted about 45–50 minutes per participant.

Data analysis

Trials including target words were analyzed. Prior to the analysis, outliers were identified in the following way. Words that were not recognized as real words were identified by the word knowledge test (see Supplemental material). As a matter of fact, a number of participants had initially failed to identify some target words although they acknowledged afterwards that they knew the words. For five of these words (one Russian neutral word, three German neutral words, one German taboo word), performance in the word knowledge test correlated with performance in the main experiment. These items were therefore excluded from the data of the concerned participants (i.e., 144 trials). Following Ratcliff (1993), trials with response latencies exceeding 7 seconds were also excluded from analysis (22 trials). After elimination, a total of 3,002 trials remained for the analysis. Data were analyzed using Microsoft Excel 17.0 and IBM SPSS Statistics 26.0. Following a normal distribution test, using Shapiro–Wilk test, a 2×2 within subjects ANOVA was performed, using Bonferroni corrected alpha levels. Data for all reported effects and further supplementary documents can be retrieved through the open science framework (https://osf.io/x8gkt/?view_only=d364b3261788464197223eb2985172c5).

Results

Correct responses in the word detection task

Averaged reaction time in all trials that contained words was 2.75 seconds ($SD=0.76$ seconds). The proportion of correct word detection is depicted in Figure 3. For neutral words, it was 0.66 seconds

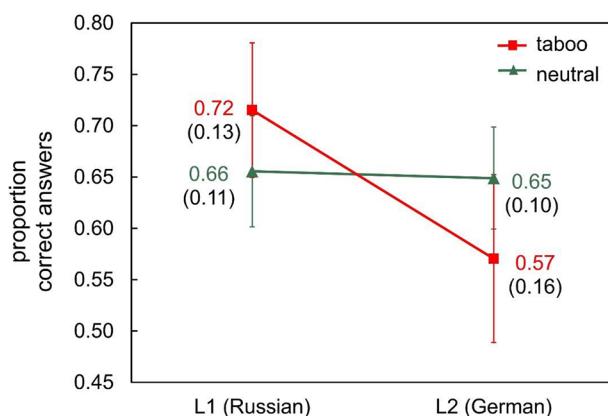


Figure 3. Mean proportion of correct responses in L1 and L2, separately for neutral and taboo words. Error bars indicate 95% confidence intervals, corrected for between-participants variance.

($SD=0.11$ seconds) in L1 and 0.65 seconds ($SD=0.10$ seconds) in L2; for taboo words, it was 0.72 seconds ($SD=0.13$ seconds) in L1 and 0.57 seconds ($SD=0.16$ seconds) in L2.

The ANOVA showed no significant main effect for language, $F(1, 21)=3.32$, $p=.083$, nor for word type, $F(1, 21)=.20$, $p=.66$. However, a *taboo word superiority*, that is, better performance for taboo words compared to neutral words, was evident in L1, whereas it was not in L2. This interaction reached statistical significance, $F(1, 21)=6.98$, $p=.015$, $\eta_p^2=.25$, confirming the different effect of word type in the two languages. Boxplots for reaction times and proportion of correct answers are provided in the Supplemental material.

AoA_{L2} and taboo word superiority

For analyzing the impact of AoA_{L2} on the emergence of the taboo word superiority in L2, we first determined whether AoA_{L2} or the time (in years) participants spent in the L2 country was a better predictor of language skills in L2. The latter was estimated through performance in the LexTALE. A linear regression analyses showed that AoA_{L2} predicted 52% of the variance in the LexTALE (adjusted $R^2=.522$, $p<.001$), while the absolute number of years spent in the L2 country did not predict performance (adjusted $R^2=.013$, n.s.). AoA_{L2} was therefore used as predictor.

For L2, AoA_{L2} predicted performance for neutral words with an adjusted R^2 of .32 ($p=.003$) and for taboo words with an adjusted R^2 of .48 ($p<.001$). Participants who arrived in the L2 country earlier performed better in L2 than those who arrived later. For L1, AoA_{L2} did not predict performance. Figure 4(a) (L1) and 4(b) (L2) plot the sizes of the taboo word superiority, that is, the difference in performance ([taboo words]–[neutral words]) for each of the 22 participants as a function of AoA_{L2}.

As evident from Figure 4(a), in L1, the large majority of participants showed a taboo word superiority except for participants arriving in the L2 country when they were 15 years or younger (cf. data on the left side of the dotted vertical line). As shown in Figure 4(b), in L2, a nearly mirror reversed pattern is observed, indicating better performance for neutral words. An ANOVA with AoA_{L2} (≤ 15 vs. > 15 years; this partition splits our participants into two equal groups of 11 individuals) as between group factor and language (L1 vs. L2) as within group factor, revealed a significant interaction between AoA_{L2} and language, $F(1, 20)=7.171$, $p=.014$, $\eta_p^2=.264$ (Figure 4(c)).

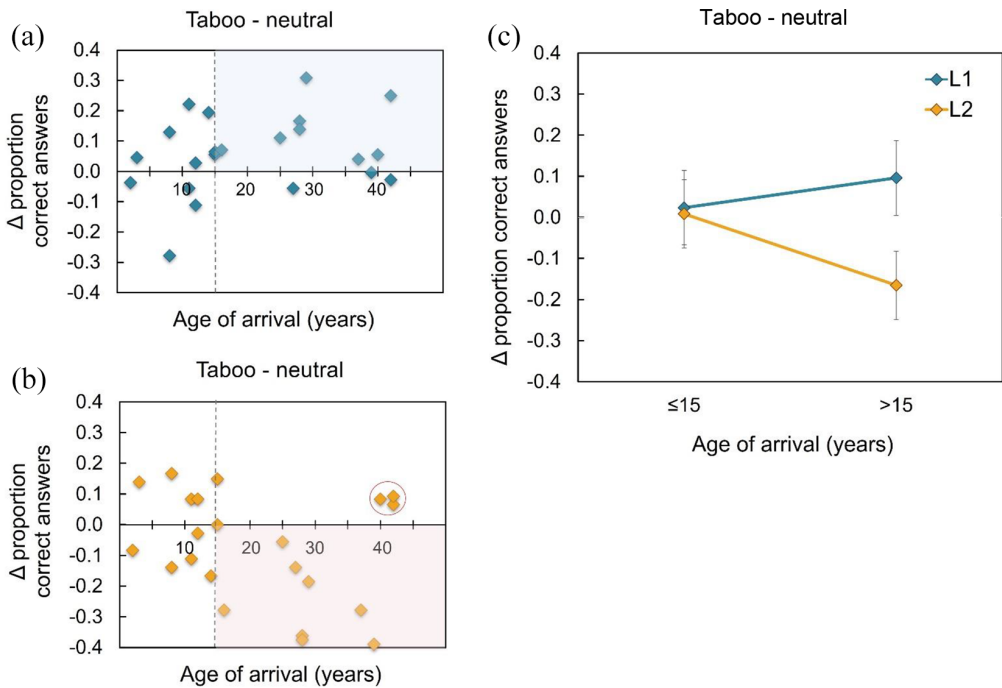


Figure 4. Left: taboo word superiority in L1 (a) and L2 (b) as a function of age of arrival in the L2 country. The vertical dotted line marks an AoA_{L2} of 15 years. Right: split-half taboo word superiority for AoA (plotted with 95%CI) (c). Positive values indicate better performance for taboo words.

Discussion

Using a behavioral paradigm with a well-characterized sample of bilinguals, we demonstrated a taboo word superiority effect in L1 and the absence of such an effect in L2 (see Figure 3). This result is consistent with the hypothesis that emotional resonance differs when processing words in a mother tongue compared to words in a language acquired later in life. Critically, when taking *individual* AoA_{L2} into account, our data showed a fairly clear divide in the performance pattern of our participants. When participants arrived at the L2 country before the age of around 15, there was virtually no taboo word superiority on the group level, neither in L1 nor in L2. The absence of such an effect stems from the fact that some of these participants showed a taboo word superiority, while others showed better performance for neutral words (see Figure 4(a) and (b)). This variability was observed in both languages with no discernable systematics at the level of the participants. By contrast, when participants arrived after the age of around 15, a clear effect of language became evident, with a taboo word superiority in L1 and no such effect in L2 (see Figure 4(c)).

Note that these results are in line with several findings summarized in the "Introduction" section. Eilola et al. (2007), for instance, who tested bilinguals with an $AoA_{L2} \leq 13$ using the emotional Stroop paradigm, and Opitz and Degner (2012), who examined the EPN to emotional words in bilinguals with an $AoA_{L2} \leq 16$ years did not find language specific effects. Harris (2004), by contrast, showed that while early bilinguals displayed a similar pattern of EDA responses in L1 and L2, bilinguals with an $AoA_{L2} > 11$ years showed stronger EDA for childhood reprimands in L1 only. Our results thus allow demonstrating the impact of the heterogeneity of a small bilingual

population on the outcome of the experiment: The magnitude of the taboo word superiority in L1 and L2 depends on the number of participants with an $AoA_{L2} > 15$. With a slightly higher number of participants with an $AoA_{L2} \leq 15$, we may have joined the studies that failed to show differential pattern of emotional resonance in L1 and L2 on the group level (i.e., Eilola et al. 2007; Opitz & Degner, 2012; Sutton et al., 2007). It is important to emphasize that our sample size is too small to precisely define the critical AoA_{L2} , above which emotional resonance during language processing declines. The here suggested age of 15 years is therefore only a rough estimate that should be refined with a larger sample.

While our data in Figure 4(b) suggests that emotional resonance during processing of L2 taboo or swear words strongly diminishes with an $AoA_{L2} > 15$, 3 of our 11 participants with an $AoA_{L2} > 15$ were an exception to this rule (i.e., red circled dots). These three were among the five participants who were reluctant to utter the Russian (L1) taboo words aloud, and all three left the L1 country late ($AoA_{L2} > 40$; see Supplemental material). Depending on the period and the cultural context of L1 acquisition, the use of taboo words (particularly by children) may be more or less severely punished. In a recent study with bilinguals, Grégoire and Greening (2020) demonstrated that fear conditioning, for example, reinforcing a neutral word of one language with a mild electrical shock, automatically generalizes to the same word in the other language (cf. semantic generalization). Besides AoA_{L2} , semantic generalization due to fear conditioning could therefore be another critical factor affecting the taboo word superiority in L2. The latter finding thus signals that several independent learning mechanisms might shape the affective value of a word. To understand the role played by these different mechanisms, it is important to reach and test specific target samples of bilinguals. Such an endeavor would benefit from test facilities like ours, which make it possible to collect data beyond the classic laboratory environment.

In this study, beyond the assessment of language proficiency using standardized language tests and self-ratings, participants' knowledge about the target words was also tested. Self-ratings of language proficiency using a scale of 0 (very bad) to 3 (very good) resulted in an average proficiency estimation of 2.5 for L1 and 2.3 for L2. Hence, self-rated language proficiency could be considered as high for both languages. However, for both languages, the word knowledge test revealed, that a number of participants failed to recognize certain of the target items as being real words. Also, while certain participants knew that a given L2 taboo word was a word, when discussing the meaning of these words, the experimenter realized that sometimes the assigned meaning was incorrect and often less offensive. Requesting participants to give a brief definition of target words (e.g., via multiple choice) may therefore be helpful.

Moreover, in this experiment, participants indicated their responses through keystroke and had to name the detected word or report that no word was detected in parallel. These verbal responses matched in 94% of the cases with their key press. An analysis of the key press revealed that different error types (i.e., incorrect key press with correct verbal response and vice versa) were equally distributed over all factor levels, indicating that there was no systematic influence on the outcome. Therefore, in an online study, only key press could be used as a dependent measure. Still, the experiment was conducted in a rather controlled setting at the participants' respective homes with an experimenter on site. Hence, future research is necessary, validating the paradigm online or without supervision.

Furthermore, given their strong emotional valence, taboo words have been frequently used in studies on emotional resonance during language processing. However, controlling for equivalence of taboohood between words of different languages is often difficult. Moreover, taboo words are mainly used in spoken language and can have a much higher frequency of occurrence than estimated by written language corpora, as we did in this study. It may therefore be preferable to work with other types of verbal items of varying affective values.

Theories of lexical processing in bilinguals (see Kroll & Ma, 2018) stipulate that cross-language processes are engaged from stimulus onset and activate common semantic representations. While the words' affective value is not central to such theories, work with pictures and words have been taken to suggest that the affective value of words is associated with conceptual information in the (common) semantic system (e.g., De Houwer & Hermans, 1994; Sianipar et al., 2015). Within such a theoretical framework, emotional resonance should therefore not differ between L1 and L2. However, our results add to a series of findings that indicate that emotional resonance varies between the two languages (e.g., Chen et al., 2015; Colbeck & Bowers, 2012; Harris, 2004). While associating affective values of words with the conceptual level could account for fear conditioning (cf. Grégoire & Greening, 2020), earlier processing stages should be considered as potential candidates to account for differences that are observed in the processing of L1 and L2 taboo or swear words. The level at which the sound structure of verbal stimuli (i.e., the phonological features that differentiate the word “fuck” to the “f-word”) is elaborated, as suggested by Bowers and Pleydell-Pearce (2011), could be one possibility.

In short, differences in emotional resonance between a first and a second language can be demonstrated by testing a well-characterized sample of bilinguals with a sufficiently sensitive behavioral experimental paradigm. As suggested by these data, the age of second-language acquisition is probably one of the major factors in the formation of language-driven emotional resonance.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project is partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 425867974, to Anke Huckauf) and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

ORCID iD

Ina Koniakowsky  <https://orcid.org/0000-0002-4075-1234>

Supplemental material

Supplemental material for this article is available online.

Notes

1. In line with Pavlenko (2012), we use the terms first language (L1) and second language (L2) to denote the chronological order of language acquisition from birth.
2. We took the term *emotional resonance* from the work by Caldwell-Harris and colleagues (e.g., Harris et al., 2003) to refer to the feeling of emotional intensity during language processing. The assumption is that emotional activation is stronger, more automatic, or develops faster in L1 compared to L2. When using behavioral measures, differential recognition performance during the processing of the same words in the two languages is therefore expected (e.g., Pavlenko, 2012).
3. Incera et al. (2020) do not specify whether the L1 participants in their experiment are mono- or bilinguals.
4. For this population, the age of arrival is the age at which L2 learning began. Therefore, the age of arrival is also the age of acquisition of L2. Consequently, the abbreviation AoA_{L2} is used.

References

- Baumann, O., & Mattingley, J. B. (2012). Functional topography of primary emotion processing in the human cerebellum. *NeuroImage*, 61(4), 805–811. <https://doi.org/10.1016/j.neuroimage.2012.03.044>
- Ben-Haim, M. S., Williams, P., Howard, Z., Mama, Y., Eidels, A., & Algom, D. (2016). The emotional Stroop task: Assessing cognitive performance under exposure to emotional content. *Journal of Visualized Experiments*, 112, 53720. <https://doi.org/10.3791/53720>
- Bowers, J. S., & Pleydell-Pearce, C. W. (2011). Swearing, euphemisms, and linguistic relativity. *PLOS ONE*, 6(7), Article e22341. <https://doi.org/10.1371/journal.pone.0022341>
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, 49(1), 1017–1034.
- Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, 5, Article 1055. <https://doi.org/10.3389/fpsyg.2014.01055>
- Caldwell-Harris, C. L., Staroselsky, M., Smashnaya, S., & Vasilyeva, N. (2012). Emotional resonances of bilinguals' two languages vary with age of arrival: The Russian–English bilingual experience in the US. In P. Wilson (Ed.), *Dynamicity in emotion concepts* (pp. 373–395). Peter Lang.
- Chen, P., Lin, J., Chen, B., Lu, C., & Guo, T. (2015). Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese–English bilinguals. *Cortex*, 71, 34–48. <https://doi.org/10.1016/j.cortex.2015.06.002>
- Colbeck, K. L., & Bowers, J. S. (2012). Blinded by taboo words in L1 but not L2. *Emotion*, 12(2), 217–222. <https://doi.org/10.1037/a0026387>
- Conrad, M., Recio, G., & Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: A cross cultural ERP study with German–Spanish bilinguals. *Frontiers in Psychology*, 2, Article 351. <https://doi.org/10.3389/fpsyg.2011.00351>
- De Houwer, J., & Hermans, D. (1994). Differences in the affective processing of words and pictures. *Cognition and Emotion*, 8, 1–20. <https://doi.org/10.1080/02699939408408925>
- Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development*, 25(2–3), 204–222. <https://doi.org/10.1080/01434630408666529>
- Dewaele, J.-M. (2015). From obscure echo to language of the heart: Multilinguals' language choices for (emotional) inner speech. *Journal of Pragmatics*, 87, 1–17. <https://doi.org/10.1016/j.pragma.2015.06.014>
- Dewaele, J.-M., & Pavlenko, A. (2001–2003). *Web questionnaire “bilingualism and emotions”*. University of London.
- Dewaele, J.-M., & Qaddourah, I. (2015). Language choice in expressing anger among Arab–English Londoners. *Russian Journal of Linguistics*, 19(4), 82–100.
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197. <https://doi.org/10.1017/S1366728902003012>
- Drijvers, L., & Özyürek, A. (2020). Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Language and Speech*, 63(2), 209–220. <https://doi.org/10.1177/0023830919831311>
- Eilola, T. M., & Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, 15(3), 353–369. <https://doi.org/10.1177/1367006910379263>
- Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion*, 21(5), 1064–1076. <https://doi.org/10.1080/02699930601054109>
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118(1), 13–42. <https://doi.org/10.1037//0096-3445.118.1.13>
- Golodnaya, V. N. (2017). Obscene words in positive meanings (on the examples of English and Russian invectives). *Russian Linguistic Bulletin*, 1(9), 64–66. <https://doi.org/10.18454/RULB.9.11>

- Grégoire, L., & Greening, S. G. (2020). Fear of the known: Semantic generalisation of fear conditioning across languages in bilinguals. *Cognition and Emotion*, 34(2), 352–358. <https://doi.org/10.1080/0269931.2019.1604319>
- Hamada, M. (2017). L2 word recognition: Influence of L1 orthography on multi-syllabic word recognition. *Journal of Psycholinguistic Research*, 46(5), 1101–1118. <https://doi.org/10.1007/s10936-017-9483-1>
- Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development*, 25(2–3), 223–247. <https://doi.org/10.1080/01434630408666530>
- Harris, C. L., Aycicegi, 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. <https://doi.org/10.1017/S0142716403000286>
- Harris, C. L., Gleason, J. B., & Aycicegi, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. *Bilingual Education and Bilingualism*, 56, 257–283. <https://doi.org/10.21832/9781853598746-012>
- Hsu, C.-T., Jacobs, A. M., Citron, F. M. M., & Conrad, M. (2015). The emotion potential of words and passages in reading Harry Potter—An fMRI study. *Brain and Language*, 142, 96–114. <https://doi.org/10.1016/j.bandl.2015.01.011>
- Incera, S., Tuft, S. E., Fernandes, R. B., & McLennan, C. T. (2020). Using mouse tracking to investigate auditory taboo effects in first and second language speakers of American English. *Cognition and Emotion*, 34(6), 1291–1299. <https://doi.org/10.1080/02699931.2020.1741339>
- Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: Insights from ERP studies. *Progress in Brain Research*, 156, 147–183. [https://doi.org/10.1016/S0079-6123\(06\)56008-X](https://doi.org/10.1016/S0079-6123(06)56008-X)
- Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: Early cortical responses to emotional words during reading. *Psychological Science*, 18(6), 475–480. <https://doi.org/10.1111/j.1467-9280.2007.01924.x>
- Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, 42(2), 998–1031. <https://doi.org/10.1016/j.neuroimage.2008.03.059>
- Kroll, J. F., & Ma, F. (2018). The bilingual lexicon. In E. M. Fernández & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 294–319). Wiley Blackwell.
- Leipzig Corpora Collection. (2018). *German news corpus based on texts crawled 2018*. https://corpora.uni-leipzig.de/de?corpusId=deu_newscrawl-public_2018
- Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961–1967. <https://doi.org/10.1016/j.neuropsychologia.2012.04.021>
- Ortigue, S., Michel, C. M., Murray, M. M., Mohr, C., Carbonnel, S., & Landis, T. (2004). Electrical neuroimaging reveals early generator modulation to emotional words. *NeuroImage*, 21(4), 1242–1251. <https://doi.org/10.1016/j.neuroimage.2003.11.007>
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology / Journal International De Psychologie*, 47(6), 405–428. <https://doi.org/10.1080/00207594.2012.743665>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- Russian National Corpus. (2003–2021). National Corpus of the Russian Language. <https://ruscorpora.ru/new/search-main.html>
- Schindler, S., & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, 130, 362–386. <https://doi.org/10.1016/j.cortex.2020.06.010>

- Sianipar, A., Middelburg, R., & Dijkstra, T. (2015). When feelings arise with meanings: How emotion and meaning of a native language affect second language processing in adult learners. *PLOS ONE*, 10(12), Article e0144576. <https://doi.org/10.1371/journal.pone.0144576>
- Sulpizio, S., Toti, M., Del Maschio, N., Costa, A., Fedeli, D., Job, R., & Abutalebi, J. (2019). Are you really cursing? Neural processing of taboo words in native and foreign language. *Brain and Language*, 194, 84–92. <https://doi.org/10.1016/j.bandl.2019.05.003>
- Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish–English bilingual speakers. *Cognition and Emotion*, 21(5), 1077–1090. <https://doi.org/10.1080/02699930601054133>
- Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, 120(1), 3–24. <https://doi.org/10.1037/0033-2909.120.1.3>

Author biographies

Anna Weimer completed her M.Sc. studies in Psychology at Ulm University, Germany, now working in the neuropsychological department of a rehabilitation clinic. The current publication substantially derives from her master thesis.

Ina Koniakowsky studies psychology in her master's degree at Ulm University, Germany, and is engaged in projects on language and emotion and eye tracking in the context of her work as a research assistant.

Tatjana A. Nazir is a Cognitive Scientist and CNRS Research Director at the SCALab, University of Lille, France. Her research focuses on embodied cognition and language.

Anke Huckauf serves as Professor for General Psychology at Ulm University, Germany. Collaborating as PI in national and international projects, her research focuses on applied visual perception.