# Emotionality ratings and electrodermal responses to university-related expressions in a native and a non-native language

**Author 1 (corresponding)**

Name: Norbert Vanek

Affiliation:  University of Auckland

E-mail: norbert.vanek@auckland.ac.nz

Address: School of Cultures, Languages and Linguistics, Faculty of Arts, University of Auckland, CLL Building, 18 Symonds Street, Auckland, 1010, New Zealand

ORCID: https://orcid.org/0000-0002-7805-184X


**Author 2**

Name: Artem Tovalovich

Affiliation: University of York

Email: tomtoff@mail.ru

# Abstract

To what extent does emotional reactivity differ when bilinguals process input in their native (L1) or non-native language (L2)? Does the L1 elicit a significantly stronger emotional arousal or can salient second language experience generate comparably strong associations between emotions and the L2? These questions were addressed through two measures of emotional arousal, (online) skin conductance responses (SCR) and (offline) emotionality ratings. Russian-English late bilinguals, UK university students, were presented different types of university-related expressions in English and Russian. The vocabulary types were university-related emotionally-laden expressions ("Плагиат"/"Plagiarism") and neutral words ("Круг"/"Circle"). Two main results emerged. First, in L1, SCRs showed a significantly increased electro-dermal activity when participants reacted to university-related words. Emotionality ratings showed contrasts based on stimulus type in both languages. These results indicate that university-related words qualify as a category of emotionally charged expressions. Second, between-language tests showed that electrodermal reactivity was not more reduced in L2 than in L1, which was also mirrored in emotionality ratings. These findings are located within the existing empirical context, and alternative interpretations are provided to further our understanding of how an emotionally salient L2 context contributes to shifts from mother tongue dominance to an increased emotional power of the second language.


Keywords: Bilingualism, Emotional reactivity, Skin conductance responses, Emotionality ratings, University-related expressions.

# Introduction

Speakers of multiple languages may find it intuitively appealing to assign dominance in emotional strength to their native language. While a lot of research points in this direction (e.g., Pavlenko, 2008; Grosjean, 2010), a noteworthy exception are early bilinguals who acquired their second language in naturalistic and emotionally salient contexts (Grabovac, & Pléh, 2014; Pavlenko, 2012; Caldwell-Harris, 2014). The question of whether a non-native language can match or exceed the intensity of the mother tongue in populations other than early bilinguals is resonant because of discrepancies in findings often attributable to different data elicitation methods. Studies based on self-reports often suggest greater psychological intimacy and power of the native language (e.g., Dewaele, 2004; Dewaele, 2008). Reports from studies using emotionality ratings and Stroop-like distractors have brought more mixed results (Eilola & Havelka 2011; Winskel, 2013). Experimental studies based on psychophysiological measures of emotional reactivity to date seem to align more closely with self-reports. For instance, in Caldwell-Harris, Tong, Lung and Poo (2011), and Harris (2004), skin conductance responses showed a greater autonomic

arousal towards the native language, unless the second language had become dominant, and participants underwent secondary affective socialization. The results of these approaches amplify the limitation of moving the debate of language-modulated emotional intensity forward by means of a single method. In response, this study combines two established tools into a set comprising an online (automatic) skin response measure with offline (decision-based) emotionality ratings.

Expressions in few semantic domains were observed to elicit a higher or comparable emotional arousal in the non-native language than in the mother tongue. An informative case comes from assessing the emotional weight of the phrase "I love you" (Dewaele, 2008) and other endearments (Caldwell-Harris et al, 2011), for which the non-native language was found to outweigh the emotional intensity of the mother tongue for two subgroups of learners (representing roughly 25% from the total of 1459 participants), namely for those with an early onset of L2 acquisition and also for late learners with emotionally rich experience linked to the non-native language. These results raise the question of what other types of expressions besides endearments could evoke a comparably strong – or perhaps even stronger – emotional reactivity in the L2 than in the L1. Considering that an emotionally rich context in a second language can dethrone the L1 from its privileged position, this study set out to test whether university-related expressions acquired in an L2 context represent another domain with a similar (or possibly higher) sensitivity to L2 expressions than to their L1 analogues. University-related expressions are an under-researched yet pertinent topic in emotionality research because exams and assignments are a potential source of anxiety for providing an emotionally salient context for secondary affective socialization (Pavlenko, 2008). This study was built on the rationale that L2 expressions from a university setting are a suitable candidate to contribute to our understanding of L2-modulated emotionality in bilinguals.

To define the key concepts and the scope, a bilingual is viewed here as a highly proficient participant exposed to two languages (in this case Russian and English) through a bilingual immersion context of a family, society or educational environment (Ferré, Anglada-Tort, & Guasch, 2018). More specifically, the type of bilinguals dealt with here are late bilinguals (with an age of onset of learning L2 English at the age of 7 or later). Regarding emotionally charged words, we adopt Pavlenko's (2008) functional approach defining them as expressions that elicit emotional arousal in interlocutors (e.g., "shame on you"), as opposed to emotion expressions that merely label affective states (e.g., "I'm anxious"). Several subcategories of emotionally charged words have been examined in studies on bilingual emotions to date, including taboo words (e.g., Dewaele, 2004; Harris, Ayçíçeǧí, & Gleason, 2003), childhood reprimands, insults and endearments (e.g., Dewaele, 2008; Caldwell-Harris *et al.*, 2011) or aversive words and interjections (Pavlenko, 2008). Pavlenko (2008) aptly notes that the borders between categories are not discrete (e.g., insults and swearwords can overlap). Another layer of flexibility important to consider is that words commonly considered as emotionally neutral may gain emotional connotations in the course of a new experience, such as during L2 acquisition. For example, while words like "*plagiarism*" or "*deadline*" may have little emotional charge for a pre-university student in an L1-dominant less rigorous educational system, their status

may shift from neutral to aversive when moving to an L2-dominant more rigorous university context.

## On the emotional intensity of L1 and L2 in bilinguals

Approaches to testing emotional reactivity towards one's native (L1) and non-native (L2) language boast a great degree of variation, and so do the reported results. Methods to date largely vary between those that measure less controllable reactions during stimulus presentation and those that rely on decision-based or memory-based responses. Automatic measures of emotional reactivity include skin conductance responses (e.g., Harris, Ayçíçeğí, & Gleason, 2003), electroencephalography (e.g., Jończyk, Boutonnet, Musiał, Hoemann, & Thierry, 2016), facial electromyography (e.g., Baumeister, Foroni, Conrad, Rumiati, & Winkielman, 2017), pupillary responses (e.g Iacozza, Costa, & Duñabeitia, 2017), functional magnetic resonance imaging and event-related potentials (e.g., Chen, Lin, Chen, Lu, & Guo, 2015). Less automatic measures include reaction times in an emotional Stroop task (e.g., Sutton, Altarriba, Gianico, & Basnight-Brown, 2007), emotionality ratings (e.g., Winskel, 2013), responses of patients during psychotherapy (e.g., Kokaliari, Catanzarite, & Berzoff, 2013) or by self-reported perceptions (e.g., Dewaele, 2008).

Results have been mixed across as well as within approaches. One example is self-report studies, in which participants rated words and phrases for emotional intensity. Dewaele (2004, 2008) conducted web-based questionnaires with nearly fifteen hundred bilinguals with different L2s, and reported that swear words, taboos and the phrase *I love you* tended to be stronger in the L1 for most participants (around 75% in Dewaele, 2008). These findings align with more recent research of Ożańska-Ponikwia (2019), who studied the perceived emotional weight of *I love you* and the Polish equivalent *Kocham Cię* in Polish-English bilinguals. However, it is noteworthy that L1 was not in a privileged position for a considerable proportion of participants, namely for 25% of the sizeable cohort in Dewaele (2008). Language proficiency and dominance, naturalistic context of acquisition, degree of L1 attrition, age of onset of learning and degree of socialization were argued as likely factors that may induce a shift towards increased emotional weight of the L2 (Dewaele, 2004, 2008). Statistical analyses in Ożańska-Ponikwia (2019) showed that the degree of socialization and frequency of L2 use accounted for more than half of the variance in participants who did not report L1 to be emotionally strongest.

Another example of inconclusive findings comes from studies with emotionality ratings, where participants needed to decide on the valences of emotionally salient and neutral words presented in L1 and L2 (e.g., Anooshian & Hertel, 1994; Caldwell-Harris & Ayçiçeği-Dinn, 2009; Winskel, 2013). For an illustration, in Winskel (2013), negative words like *death* and *pain* were rated 6 on average out of 7 (where 7 is 'most unpleasant') in both L1-Thai and L2-English, whereas neutral words had a rating of around 3.5 in both languages. But while in Winskel (2013), valences for negative words in the L1 were comparable to those in the L2, e.g., in Caldwell-Harris and Ayçiçeği-Dinn (2009) ratings for lies pointed to a higher emotional arousal in the

native language than in the L2. Discrepancies also arise when ratings are compared with other methods used with the same participants. For instance, in Winskel (2013) the patters found in ratings misaligned with those from an emotional Stroop task (i.e., measuring the degree of interference from emotionally charged words used as distractors), in which emotional arousal in L2 was reduced compared to the L1.

The Emotional Stroop task is an informative methodological approach aimed at examining automatic activation and processing of emotionally charged words in bilinguals. During the task, participants see words on the screen, some of which are neutral (e.g., *rabbit*, *chair*), while others are emotionally charged (e.g., *bastard*, *idiot*). The task is to name the colour of the word shown while ignoring the meaning of the text. The relative delay in colour naming is interpreted as interference of emotionality in information processing, allowing comparisons of the emotional power of native and non-native words (Eilola & Havelka, 2011). Numerous variations of this task exist, e.g., a face-word Stroop task (Fan *et al.*, 2018), where colours are replaced with facial expressions. The emotional Stroop paradigm, just like ratings, also generated mixed results, highlighting the need to consider nuances in participant characteristics. Grabovac and Pléh (2014) found that for early balanced Hungarian-Serbian bilinguals there were negligible differences between languages in the interference of emotionality, implying very similar emotional weight of taboo and negative words in L1 and L2. This is consistent with self-reports emphasising the impact of language proficiency and early onset of acquisition on emotional sensitivity to the L2. Findings tend to differ for late unbalanced bilinguals. For instance, Eilola and Havelka (2011) and Eilola, Havelka and Sharma (2007) found no effect of emotionality in the L2 condition with Greek-English and Finnish-English late bilinguals.

One potential source of criticism regarding the Emotional Stroop task lies in its questionable sensitivity. A study pointing in this direction is Eilola and Havelka (2011) who did not find differences in reactivity to negative and taboo words using an Emotional Stroop task but differences were observed when using a Skin Conductance Response test. Such a discrepancy could be attributed to the relatively lower sensitivity of the Emotional Stroop task compared to SCR, particularly when proficient (though still unbalanced) bilingual participants are involved. SCR is viewed as a direct method to access the emotional intensity of autonomic arousal based on a paradigm of increased electrodermal activity while processing emotionally charged stimuli (Harris, Gleason, & Aycicegi, 2006). The mechanism relies on a very low, imperceptible voltage administered to monitor electrical conductance of the skin on the index and middle fingers. As a physiological response to threat recognition, the autonomic nervous system increases the perspiration of the body, which, in turn, improves the skin's electrical conductance. Being used as a part of *Lie Detection* technique, SCR is widely recognised for its sensitivity, though result interpretation depends, as with most methods, on the researcher (Harris *et al.*, 2006).

Studies built on the SCR method typically report on emotional intensity in terms of the privileged position of the mother tongue. Stimulus modality seems to make little difference, the L1 was found to be stronger than the L2 when presented as spoken or written text (Jankowiak & Korpal, 2018). As for stimulus type, Harris et al. (2003)

found sensitivity towards taboo and sex words and childhood reprimands to be greater in the native than in the non-native language. The explanations varied between greater emotional distance of the L2 due to lower proficiency, late immersion in the L2-speaking-country and late onset of acquisition. The latter argument found support in subsequent studies (Harris, 2004; Caldwell-Harris et al., 2011) documenting an increased sensitivity of early learners towards some L2 stimuli. Namely, early bilinguals exhibited no pronounced differences between L1 and L2 taboo and sex words, with the exception of reprimands, which were still stronger in the L1.

For late bilinguals, numerous findings converge on supporting native language dominance, and this seems to hold across a very rich spectrum of approaches. For an illustration of scope, resistance to shift to the L1 was reported by bilingual psychotherapists when L2 was used by patients as a defence mechanism against painful experience (Kokaliari *et al.*, 2013), or when late bilinguals subconsciously codeswitched to the L2 in discussions about traumatic events (Frie, 2011). Caldwell-Harris (2014) and Sutton *et al.* (2007) in their syntheses of findings on bilingual emotionality both emphasise the role of the context of acquisition. Studies reporting L1 dominance typically recruited participants with L1 acquired in a more naturalistic emotionally rich context and L2 usually learnt in a classroom without being deeply coded through application in various ways and experienced in numerous contexts. However, there is evidence that the power of the L2 does not necessarily fall behind (Dewaele, 2008; Caldwell-Harris et al., 2011; Ożańska-Ponikwia, 2019). For example, Caldwell-Harris et al. (2011) reported that endearments were more powerful in the L2. Similarly, studies of Dewaele (2008) and Ożańska-Ponikwia (2019) found that participants with stronger L2 tied emotional weight of "*I love you*" to their life experience and linguistic context in which they encountered this phrase.

Table 1. An overview of studies (participants, measures, findings) to portray the empirical context of the present study. (Notes: The more dominant of the bilinguals' languages is listed first; AoA = age of onset of acquiring the second language, SCR = skin conductance response)

| Study | Participants | Measure(s) | Main finding(s) |
|---|---|---|---|
| Anooshian & Hertel (1994) | English-Spanish bilinguals (mean L2 AoA >16 years) | Free recall & ratings of emotional and neutral words in L1 and L2 | More emotional words than neutral recalled in L1 but not in L2 |
| Baumeister, Foroni, Conrad, Rumiati, & Winkielman (2017) | Spanish-English bilinguals (mean L2 AoA >15 years) | Surprise memory task in L1 and L2 with facial motor resonance and SCR measures | For emotional content: Enhanced memory in L1, decreased motor resonance in L2 |
| Caldwell-Harris, Tong, Lung, & Poo (2011) | Mandarin-English bilinguals (mean L2 AoA >6 years) | SCRs and ratings of L1 & L2 endearments, insults, reprimands, taboo, neutral words | Reprimands rated as stronger in L1, SCRs for L2 endearments higher than L1, all else similar |
| Caldwell-Harris & Ayçiçeği-Dinn (2009) | Turkish-English bilinguals (mean L2 AoA >12 years) | SCRs and ratings of endearments, insults, reprimands & neutral w. | Overall greater SCRs in the L1, L1>L2 difference greatest for reprimands |
| Chen, Lin, Chen, Lu, & Guo (2015) | Chinese-English bilinguals (Mean L2 AoA >11 years) | Reaction times on pos/neg/neutral L1 & L2 word correctness, ERPs and fMRI | Robust L1 emotional word processing advantage through rapid, automatic attention |
| Dewaele (2004) | 1039 multilinguals with varying Lx AoAs | Perceived emotional force/ratings of swearwords and taboo words | Emotional force highest in L1, gradually lower in languages learned subsequently |

| | | | |
|---|---|---|---|
| Dewaele (2008) | 1459 multilinguals with varying Lx AoAs | Perceived emotional weight of the phrase *I love you* | Overall, *I love you* was strongest in the L1, but for 25% in the Lx, related to emotionally rich experience in Lx |
| Eilola & Havelka (2011) | English monolinguals and Greek-English bilinguals (Mean L2 AoA >8 years) | Stroop task and SCR with emotional (pos/neg) taboo and neutral words | L1 and L2 groups responded in very similarly on the Stroop task, but SCR L1>L2 for neg, and taboos |
| Eilola, Havelka, & Sharma (2007) | Finnish–English bilinguals (Mean L2 AoA >9 years) | Stroop task with emotional (pos/neg) taboo and neutral words in L1 and L2 | No differences in the size of the interference were present between languages |
| Fan, Xu, Wang, Xu, Yang, & Lu (2018) | Chinese-English bilinguals (Mean L2 AoA >8 years) | Face-word Stroop task with emotion words in L1 and L2 | Emotional face-word Stroop effect greater in the L1 than in the L2 |
| Ferré, Anglada-Tort, & Guasch (2018) | Spanish-Catalan early bilinguals (Mean AoA <3 years) and Spanish-English late bilinguals (Mean AoA >8 years) | Free recall of emotional (pos, neg) and neutral words in L1 and L2 | Recall for emotional words was very similar in L1 and L2, regardless of L2 AoA or language similarity |
| Grabovac & Pléh (2014) | Hungarian-Serbian early bilinguals (Mean AoA <3 years) | Stroop task with negative, positive and neutral written words in L1 and L2 | Equal interference in the two languages, similar levels of emotionality |
| Harris (2004) | Spanish-English early bilinguals (Mean AoA <4 years) and Spanish-English late bilinguals (Mean AoA >7 years) | SCR measured while listening to L1 and L2 taboos, sexual terms, childhood reprimands and neutral words | AoA effects on SCRs were found, L1>L2 for late learners' SCRs, but L1=L2 for early learners' SCRs |
| Harris, Ayçíçeğí, & Gleason (2003) | Turkish-English bilinguals (Mean AoA >12 years) | Ratings & SCR measured while hearing or reading L1 and L2 neutral, positive, aversive, taboo words and reprimands | Greater autonomic arousal to taboos and reprimands in L1 than in L2, especially in the auditory modality |
| Iacozza, Costa, & Duñabeitia (2017) | Spanish-English bilinguals (Mean AoA >5 years) | Pupil size changes while reading aloud emotional sentences in L1 and L2, plus emotionality ratings | Pupillary responses, but not ratings, showed a larger effect of emotion in L1 than in L2 |
| Jankowiak & Korpal (2018) | Polish–English bilinguals (Mean AoA >9 years) | SCRs and self-reports using emotionally-laden narratives presented in L1 and L2, in the visual and auditory modality | Decreased reactivity (SCRs) to L2 compared to L1 stimuli, stronger SCRs to visual than auditory stimuli |
| Jończyk, Boutonnet, Musiał, Hoemann, & Thierry (2016) | English monolinguals, Polish–English bilinguals (Mean AoA >11 years) | Valence ratings and ERPs in response to sentential pos/neg/neut (in)congruent adjectives | Increased N400 for odd sentences in L1 Polish but not for negative L2 English sentences |
| Ożańska-Ponikwia (2019) | Polish–English short-stay bilinguals (1-3 months in an English-speaking country) and long-stay bilinguals (2-27 years) | Perceived emotional strength of *I love you* in English and Polish | 'Long-stay' bilinguals ranked emotionality of *I love you* in L2 significantly higher than 'short stay' bilinguals |
| Sutton, Altarriba, Gianico, & Basnight-Brown (2007) | Spanish–English bilinguals (Mean L2 AoA >4 years) | Stroop task with L1 & L2 emotional (neg), and neutral words | Equal interference effects in both the L1 and the L2 |
| Winskel (2013) | English native speakers and Thai–English bilinguals (Mean L2 AoA >7 years) | Stroop task with negative emotion(al) and neutral words in L1 and L2 | An emotional Stroop effect found in the dominant L1 but not in the weaker L2, ratings similar in L1 and L2 |

In sum, a great variety of methods and stimulus types used for triangulation purposes collectively signal that the mother tongue is typically more emotional for late unbalanced bilinguals (Table 1). Language proficiency, age of onset of acquisition as well as the context of acquisition were found to play a role especially when an emotionally salient experience is associated with L2 usage (Pavlenko, 2008). These background variables are thus important to consider when testing emotional reactivity of bilinguals towards L2-linked university-related expressions. Such expressions represent an under-researched stimulus type in the area. They are of particular interest in this study because of the stress-and-anxiety-generating potential of exams and assignments for students. A further rationale comes from the literature comparing demands and academic standards in the Russian and UK universities (West & Frumina, 2012), reporting differences and increased demands in favour of the latter. One might therefore expect that the L2 context of acquisition and use of English vocabulary related to university study in the UK might have led to secondary affective socialization exhibited in Russian students as an emotional reactivity match between reactions to L2 English expressions compared with reactions to their L1 Russian equivalents. This prediction is derived from the *emotional context of learning* hypothesis (Caldwell-Harris, 2014), positing that verbal expressions develop an emotional feel through learning and habitual use in an emotionally resonant context. This is the main idea the present study set out to test.

RQ: To what extent does emotional reactivity to university-related expressions in L2 differ from that to the same expressions in the L1 when measured online via skin conductance responses and offline via emotionality ratings?

H: If subsequent recall is autobiographic and based on context-dependent learning (Marian & Kaushanskaya, 2008), the emotional punch of L2 university-related expressions made salient in the L2 environment should not be significantly reduced in comparison with their L1 equivalents.

## Methods

Participants

Twenty-eight native speakers of Russian were recruited for this study at a UK university. All participants reported Russian to be their dominant language at the time of testing, and preferred Russian as a language of instruction during the experimental session. G*Power (Erdfelder, Faul, & Buchner, 1996) was used to determine that the sample size of N = 28 would be sufficient to reach a power of 0.8 with a medium effect size of >0.4 at the standard 0.5 alpha error probability. Data from eight participants were affected by partial recording malfunction for some of the trials so their data underwent list-wise deletion and were not included in the analyses. The twenty participants included in the analyses were aged 19-39 years

(M = 24.20, SD = 4.42), and the female-to-male ratio was 12:8. All participants were full-time university students in the UK at the time of the test, with a comparable spread across the undergraduate (N = 8) and the postgraduate levels (N = 12). All participants met the inclusion criterion of having spent at least one term at a UK university and having received feedback on their study results.

Most participants began to learn English either at a primary or a secondary school in a formal context in a Russian-speaking environment. The age of onset of acquisition varied from 7 to 18 years, with a mean age of 10.4 years (SD = 3.56). On average, the length of learning English was 13.8 years (SD = 5.33). Almost no exposure to English in a naturalistic setting was reported before participants had enrolled in their studies in the UK. Only one participant had lived in an English-speaking country for 5 years during their secondary school. Following the university entry requirements at the time of testing, all students had to exceed the minimum threshold for English language knowledge set at 6.5 in the IELTS test, signalling that all participants had an upper-intermediate or higher level of proficiency in English. However, the diagnostic potential of a past IELTS test needs to be viewed with caution because of the likelihood that the participants' language skills during their studies and residence in the UK improved. Considering these background variables, the working label for the participant base here is unbalanced bilinguals with L1 Russian and L2 English.

## Stimuli

In line with related research targeting bilingual participants (e.g., Harris et al., 2003), three different types of expressions were used in both English and Russian. The first type were the critical expressions, i.e., university-related emotionally charged words or phrases (e.g., *deadline*, *exam*). The second type were control words, i.e., neutral words (e.g., *chair*), presented just before the critical items to provide a comparative baseline for recognising emotional reactivity. The third type were buffer words, i.e., also neutral words sequenced to follow critical expressions with the function to resume the state of neutral emotionality. Following Eilola and Havelka (2011), all 20 English neutral words – 10 control and 10 buffer – were taken from the Affective Norms for English Words database (Bradley & Lang, 1999). Their mean valence rating was 5.24 (SD = 0.68) and their mean arousal rating was 4.19 (SD = 0.58).

To ensure that the study used emotionally charged expressions that relate to university context, 20 international students in the UK who did not participate in the main study were asked to think of a list of 5 English words or phrases related to their university experience which made them feel most anxious or frightened. Responses were compiled into a frequency-based list. The ten most frequent English expressions were chosen for the experiment, constituting an authentically generated list of critical items with high emotional intensity. A frequency-based list of phrases solicited from university students serves here as an exploratory step mapping which university-related expressions can count as potential sources of anxiety. To take a step further and establish the actual level of anxiety in the tested participants, it will be advantageous for future work in this domain to also include a task-independent
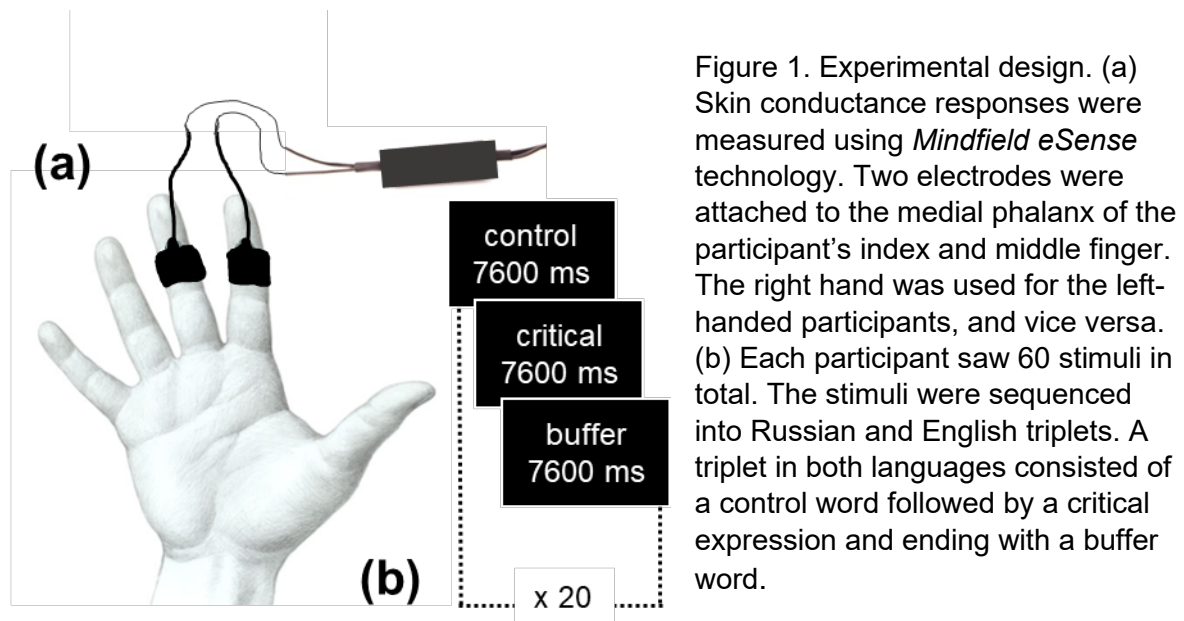
measure of anxiety, such as the Beck Anxiety inventory (Beck, Epstein, Brown, & Steer, 1988). Once the critical items were identified, the 30 English expressions (including 10 critical, 10 control and 10 buffer) were translated into 30 Russian. When direct translations sounded unnatural, they were substituted with their closest Russian analogues. Two Russian native speakers performed the translations to double-check semantic equivalence.

Each of the 30 English and 30 Russian expressions were organised into triplets. Within each triplet, the stimulus ordering conformed to control –> critical –> buffer items. Triplet formation resulted in 10 Russian triplets corresponding to 10 English triplets. To neutralise possible order effects, the triplets for English were randomly shuffled into three lists. Then, each list in English was exactly mirrored in terms of triplet sequencing with a list in Russian so that triplet order across languages was always matched. An additional layer of counterbalancing was implemented for language order. Half of the bilinguals saw the stimuli first in English then in Russian, and the other half vice versa.

Emotionality ratings were collected using a form adapted from Harris et al. (2003). After hearing and reading each stimulus, participants were asked to give an anxiety rating on a 4-point scale, ranging from "Not anxious" to "Very anxious". The rating sheet was two-sided, for 30 items for English stimuli on the one side, and 30 items for Russian stimuli on the other. Participants were asked not to compare their ratings for different languages.

## Procedure

The stimuli were presented using Microsoft PowerPoint on a laptop with a 14-inch screen. Participants saw 60 slides, each slide presenting one stimulus. Exposure to each stimulus was kept constant at 7.6 sec, after which the next slide appeared automatically. Stimulus presentation was in two modalities, audio-recorded speech and writing presented simultaneously to maximize emotional reactivity. Stimuli were read out by a proficient Russian-English bilingual, and their audio recordings were co-presented with the corresponding written stimuli. Using a voice of a bilingual speaker was purposeful and rationalised by the fact that a significant proportion of the participants' lecturers were not native English speakers, therefore university experience was not exclusively linked to native English pronunciation. Two measures were taken during stimulus presentation, skin conductance responses and emotionality ratings.

Figure 1. Experimental design. (a) Skin conductance responses were measured using *Mindfield eSense* technology. Two electrodes were attached to the medial phalanx of the participant's index and middle finger. The right hand was used for the left-handed participants, and vice versa. (b) Each participant saw 60 stimuli in total. The stimuli were sequenced into Russian and English triplets. A triplet in both languages consisted of a control word followed by a critical expression and ending with a buffer word.

Emotional reactivity was recorded by a Mindfield eSense mobile application preinstalled on an Android smartphone. Hook-and-loop fasteners were fastened to two fingers neither too tight, which would restrict blood flow, nor too loose, which could lead to a signal loss. Participants were asked to read the expressions on the screen while listening to them being read out loud. The items were presented in triplets (Figure 1), each of which started with a control (neutral) item to provide a comparative baseline, followed by a critical (emotional) item to test the degree of emotional intensity in the target domain, and ending with a buffer (neutral) item to neutralise possible carryover effects. Buffer items served the function of inter-stimulus intervals, used in response to the finding that, on average, it takes between 4–8 seconds for an SCR to return to the baseline value (Hugdahl, 2001). For each item, 5 values were recorded per second, which in a 7.6 sec interval for each item resulted in 38 values recorded per item. In total, 2280 skin conductance response values were collected per participant.

In parallel with measuring skin conductance responses, participants were asked to use their free hand, i.e., their dominant hand with no electrodes attached, to provide emotionality ratings. Their task was to first carefully read and listen to the given expression and then use the provided rating sheet to mark the option they felt best reflected their anxiety level. This step was repeated 60 times, collecting participants' emotional weight ratings for each stimulus. The experimenter was trained to check for each session that the start of the PowerPoint presentation was time-locked to the start of SCR recordings.

At the start of the experiment, each participant was asked to self-assess their language dominance and to choose whether they preferred Russian or English as the language of instruction during the experimental session. All participants reported Russian as their stronger language and opted for instructions in Russian. Language history and background details were collected after the participant had completed the experimental task. Besides questions about length of studying English and the

language(s) used in their pre-university education, participants were asked to indicate in which language, English or Russian, words and phrases related to education they thought were greater anxiety generators for them. Four multiple-choice options were provided (a) Stronger in Russian, (b) Stronger in English, (c) Equally strong regardless of the language, or (d) Equally neutral (not emotional). Responses (seventeen As and only three Cs) were largely concordant with the self-assessed language dominance at the start of the experiment.

Participants were tested individually in a silent room with a temperature kept constant at 22 degrees Celsius. Testing started after participants had washed their hands, got familiarised with the procedure, and signed a consent form. Participation was remunerated. Two pilot sessions were run to test instruction clarity and equipment functionality. The pilot test participants (not included in the analyses) reported the task to be easy to understand. All procedures were approved by the Ethics Committee of the Department of Education, University of York.

# Results

## Emotionality ratings

As a measure of perceived emotional intensity, we calculated averages of participants' ratings of both stimulus types (critical vs. control) separately per language (Russian vs. English). This approach allowed us to compare (a) whether emotionally charged university-related expressions qualify as aversive stimuli when processed offline, i.e., if they elicit higher anxiety ratings than neutral words, and (b) whether the ratings for the L2 stimuli (English) are comparable to those for the L1 stimuli (Russian). In terms of stimulus types, the results (Figure 2) showed that the average ratings (1-4) for critical stimuli (M = 2.26, SD = 0.95) substantially exceeded the average ratings for neutral stimuli (M = 1.17, SD = 0.41). In terms of language differences, the average ratings for the L2 English stimuli (M = 2.30, SD = 0.94 for critical items; M = 1.18, SD = 0.44 for control items) closely matched those for the L1 Russian stimuli (M = 2.22, SD = 0.96 for critical items; M = 1.16, SD = 0.39 for control items).
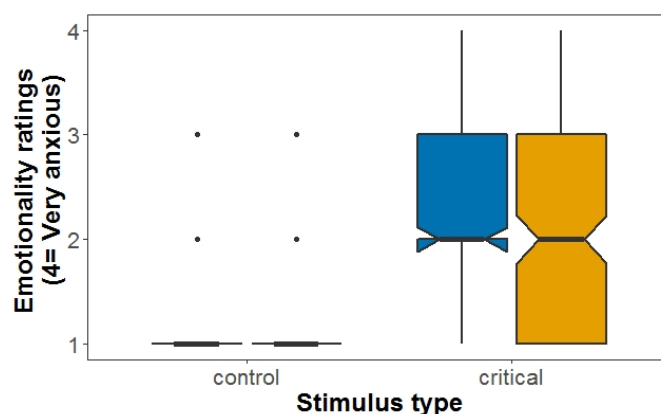


Figure 2. Notched plots of emotionality ratings (1 = not anxious, 4 = very anxious) elicited in response to control stimuli (left) and critical stimuli (left) presented to participants in the English L2 (blue) and the Russian L1 (orange).

A series of mixed-effects models were built using R (R Core, Team, 2016) and lme4 (Baayen, Davidson, & Bates, 2008) to examine the effect of stimulus type on emotionality ratings. We specified *Language* (Russian/English), *Type* (critical/control) and *Order* (English-Russian/Russian-English) as fixed effect factors and *Participant* and *Item* as random effect factors. We also tested the interaction between *Language* and *Type*. The lme4 formula used to fit the model was rating ~ language * type + order + (1 | participant) + (1 | item). The model returned a significant effect of *Type* ($\beta$ = 1.11, *SE* = 0.14, *t* = 8.19, *p* < .001), pointing to an overall higher level of anxiety marked for critical items. No other effects were significant. To establish whether the used university-related expression qualify as emotionally charged aversive stimuli in each language, in the next step we ran mixed effects regressions separately in the Russian and the English subset of the data (the same model structure as above but excluding *Language*). The models showed significantly higher perceived anxiety in the critical compared to the neutral stimuli both in the Russian ($\beta$ = 1.06, *SE* = 0.15, *t* = 7.45, *p* < .001) and in the English subset ($\beta$ = 1.11, *SE* = 0.13, *t* = 8.59, *p* < .001). These models provide a direct confirmatory test of the hypothesis that when it comes to emotionality measured via offline ratings, the perceived strength of L2 expressions matches that of their L1 analogues.

## Skin conductance responses

As a measure of online sensitivity to emotional expressions, we identified the peak skin conductance response values for each item for every participant. The peak values were considered to indicate the highest level of emotional reactivity towards a stimulus, in line with related studies measuring electrodermal activity with fingertip electrodes (e.g., Harris et al, 2003; Caldwell-Harris et al, 2011).

The aims of the SCR analyses closely matched those for the emotionality ratings, namely, to compare (a) whether emotionally charged university-related expressions qualify as aversive stimuli when processed online, i.e., if they elicit higher skin conductance responses than neutral words, and (b) whether the SCRs for the L2 stimuli (English) are comparable to those for the L1 stimuli (Russian). The mean SCR values for critical stimuli (M = 1.72, SD = 0.78) exceeded the mean SCR peak values for neutral stimuli (M = 1.64, SD = 0.76). Considering between-language similarities, the average SCRs for the critical items (M = 1.71, SD = 0.94 for L2 English stimuli, M = 1.74, SD = 0.96 for the L1 Russian stimuli) were in both languages higher than the average SCRs for the control items (M = 1.64, SD = 0.44 for L2 English stimuli, M = 1.64, SD = 0.39 for the L1 Russian stimuli).
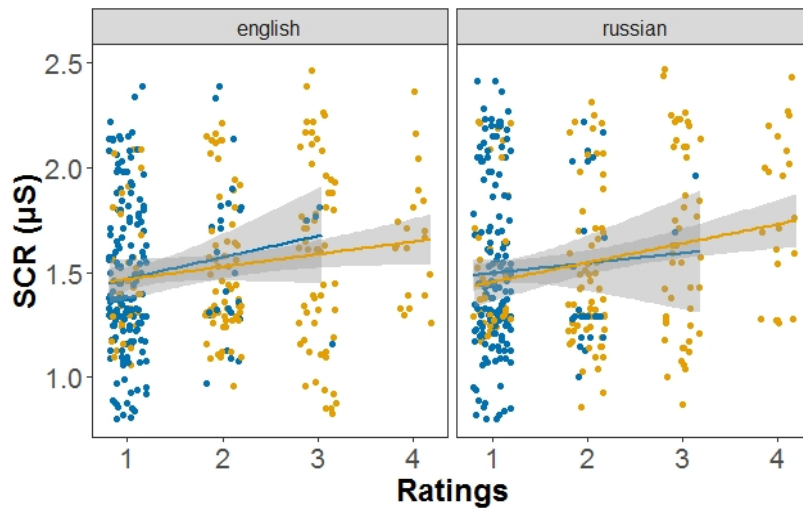
Figure 3. A by-participant correlation plot showing the relationship between participants' ratings and SCR responses (in Micro Siemens) to the control items (blue) and critical items (orange) presented in L2 English (left) and L1 Russian (right). No meaningful relationship was found between increases in ratings and increases in electrodermal responses.

First, our analyses examined a possible correlation between emotionality ratings and skin conductance ratings (Figure 3) to see the strength with which the two measures correlate within participants. Pearson tests showed no correlation between the two variables ($r$ = .010, $p$ = 0.789).

Next, to statistically assess whether university-related emotionally charged expressions qualify as aversive stimuli when processed online in the speakers' native language, a mixed effects model (with the same structure as for the ratings) was built using the Russian subset of the SCR data. The model returned a significant effect of *Type* ($\beta$ = 0.05, *SE* = 0.02, $t$ = 3.08, $p$ = 0.006), suggesting an overall stronger reactivity to critical items presented in the speakers' native language. This result aligns with the emotionality ratings of Russian stimuli and strengthens the idea that expressions from a university setting can be viewed as emotionally intense, i.e., aversive. The same pattern was not mirrored in the L2 data. The mixed effects model fitted to the English subset of the SCR data did not return a significant effect of *Type* ($\beta$ = 0.04, *SE* = 0.03, $t$ = 1.43, $p$ = 0.169). This quantitative result on its own does not align with the prediction that university-related expression would qualify as emotionally charged distinctly aversive stimuli when processed online in the L2. On average, the difference of reactions to critical vs. control L2 expressions measured via online skin conductance responses was not statistically significant, unlike for the corresponding L1 expressions. However, what stands out, on the level of analysis by item, that the strongest electrodermal reactions were elicited not by L1 but by two L2 expressions (Figure 4).
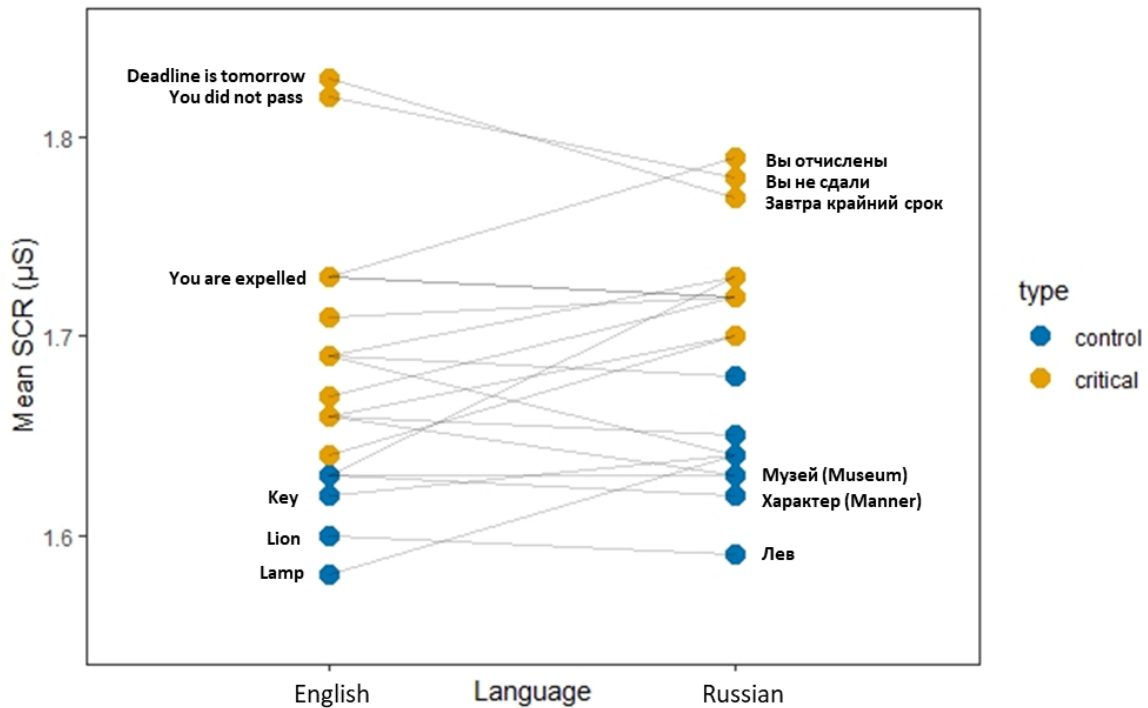
Figure 4. Mean skin conductance responses by item. Each line connects the participants' average SCR to an L2 English stimulus (left) with the average SCR to the corresponding L1 Russian stimulus (right). To aid interpretation, the three highest and three lowest scoring items in each language are labelled (with translations in brackets for those Russian items where the corresponding English score is not among the top or bottom three).

While there was variation across items (Figure 4), in three cases the SCRs to university-related expressions in L2 English exceeded those to L1 Russian, namely *Deadline is tomorrow* ($M_{ENG}$ = 1.83, SD = .83; $M_{RUS}$ = 1.77, SD = 0.96), *You did not pass* ($M_{ENG}$ = 1.82, SD = .78; $M_{RUS}$ = 1.78, SD = 0.96), and *Assignment* ($M_{ENG}$ = 1.73, SD = .78; $M_{RUS}$ = 1.72, SD = 0.79). However, for other word pairs, participants reacted more strongly in the Russian condition than in the English condition; for instance, *Fees payment* ($M_{ENG}$ = 1.63, SD = .75; $M_{RUS}$ = 1.73, SD = 0.88), *You are expelled* ($M_{ENG}$ = 1.73, SD = .79; $M_{RUS}$ = 1.79, SD = 0.75), or *You failed the exam* ($M_{ENG}$ = 1.69, SD = .80; $M_{RUS}$ = 1.73, SD = 0.77). Although some items included as neutral stimuli, such as *Plant* ($M_{ENG}$ = 1.73, SD = .79; $M_{RUS}$ = 1.72, SD = 0.40), scored higher than some university-related expressions, such as *Plagiarism* ($M_{ENG}$ = 1.64, SD = .76; $M_{RUS}$ = 1.70, SD = 0.75), there was a general observable trend of university-related expressions to elicit higher SCRs compared to neutral words. The full list of average ratings and skin conductance responses per item presented in English and Russian is provided in the Appendix.

# Discussion

University-related expressions qualify as emotionally charged stimuli

This study set out to investigate how emotional reactivity to L2 words compares to the corresponding L1 words. One novel element compared to previous work was the choice of lexical domain, namely university-related expressions. Two levels of verification were provided to document that university-related expressions qualify as emotionally charged stimuli, offline ratings and online skin conductance responses. Both English university-related vocabulary and their Russian equivalents received significantly higher emotionality ratings when compared to the neutral stimuli. However, these offline results were only partially mirrored by online skin conductance responses. Electrodermal activity increased significantly more in response to critical than to control items when L1 Russian was used as the input language. Even though the two highest mean emotional responses were in reaction to L2 input (*Deadline is tomorrow*, *You did not pass*), L2 English critical items did not elicit significantly higher SCRs than the L2 English control items. In sum, these results are interpreted as a confirmation of the hypothesis that university-related vocabulary can represent a new category of emotional expressions (Pavlenko, 2008). However, L2 results suggest that presenting this type of expressions in the L2 does not elicit the same emotional reactivity measured online when the stimulus processing unfolds (Jankowiak & Korpal, 2018). Debriefing discussions with participants and self-reported perceptions of emotional weight of university-related expressions align with the findings from offline ratings and L1-based SCRs pointing to an emotional nature of the critical stimuli used in the experiment. One might wonder whether the stimulus strength in this study is comparable to studies using emotional stimuli from other domains, for which standardised differences between two means (of the critical and control items) can be useful indicators. University-related expressions were found emotionally charged but they only evoked mild emotional responses in comparison to the other types of emotional words – taboo and sexual terms, childhood reprimands and insults – used in previous studies (e.g., Caldwell-Harris et al, 2011; Dewaele, 2004; Harris, Ayçíçeğí, & Gleason, 2003). The absence of a direct comparison between university-related expressions and other emotion-laden word types reported in the literature limits the informative strength of the present design. Future work will benefit from manipulating emotion-laden categories, in which university-related expressions will be tested alongside well-established emotional word types such as taboos or reprimands.

There are other open questions related to the emotional charge of university-related expressions. One question is the potential role of relevance and its variation across items and individual learners or learner groups. For instance, in 'Submit your dissertation!' the full phrase was relevant for (post)graduate students but arguably just its first part to undergraduate students. While in this study most participants

came from (post)graduate cohorts, potential relevance effects cannot be excluded. To address this question, future studies could ask participants to rate individual items for (recent) relevance and use these ratings to test whether the degree of phrase relevance is a significant predictor of emotional intensity observed in SCRs. Further open questions relate to the potential role of stimulus length, frequency and complexity. While potential word or phrase length and frequency effects in processing emotional stimuli cannot be ruled out (Larsen, Mercer, Balota, & Strube, 2008), strict length and frequency control is particularly important for reaction-time-based analyses dependent on processing speed, but, possibly, plays a lesser role when the measure is the peak SCR value in a relatively generous 7.6-second time window. In terms of complexity, the critical stimuli in this study consisted of a combination of single emotion-laden words and more complex emotion-laden phrases. Although not rare in this domain (e.g., Harris et al., 2003) such mixing might problematise interpretations because of unmatched contextual richness, greater in phrases than it is in single words. The extent to which word and phrase mixing affects the processing of emotionality is still poorly understood. Future work comparing the same emotion-laden words with and without context will be in a good position to explicitly test this assumption. With respect to word/phrase choices in this study, it is also important to acknowledge that all critical stimuli come from the same semantic category, but the control stimuli span across different categories. The exact extent to which such a discrepancy may have influenced processing during SCR recordings and emotionality ratings is an epistemological issue that remains open for future investigations.

## Similarities and differences in emotional arousal in response to L1 and L2

Analyses revealed no significant differences between English and Russian *neutral* words. These results were consistent across both SCR measures and emotionality ratings, providing assurance of a valid comparative baseline. In the rating task, the average ratings of the English neutral words and of the Russian neutral words were in fact the same value (1.12). Consequently, it can be concluded that the Russian equivalents of English neutral words were appropriate choices to serve as neutral stimuli together with the English neutral words adopted from Bradley & Lang's (1999) Repository of the affective norms for English words. A potentially useful quality booster for the design would be to perform L1 and L2 valence value matching (e.g., Caldwell-Harris *et al.*, 2011; Winskel, 2013). However, as Russian affective norms were not available at the time of testing, an L1 stimulus validity check was performed through 100% inter-translator agreement on English-Russian translation equivalents.

With regard to emotionality ratings, our results contradict the observations in Caldwell-Harris and Ayçiçeği-Dinn (2009), who found that lies were rated as more emotionally charged in the native language. They also misalign with ratings that Chinese-English bilinguals assigned for insults, reprimands and taboo words in

Caldwell-Harris *et al.* (2011). These differences may be attributed to the participant base comprising balanced bilinguals with equal proficiency in both L1 an L2, unlike the present study with unbalanced bilinguals. Further cross-study comparisons of emotionality ratings also highlight numerous overlaps. For instance, the findings here are consistent with those in Harris *et al.* (2003), where late Turkish-English bilinguals' ratings of aversive words (e.g., "kill", "murder") were similar in L1 and L2. Similarly, Winskel (2013) found that late Thai-English bilinguals assigned roughly equal valences for negative words in their L1 and L2.

The major difference between languages was the finding that skin conductance responses in L1 Russian clearly distinguished between emotionally charged and neutral words while in L2 English they did not. One possible way to interpret this pattern mismatch between electro-dermal reactivity towards English university-related vocabulary and their Russian analogues is that even though there is no pattern match, emotionally charged expressions did not elicit reduced emotional reactivity in the second language compared to the first (unlike e.g., in Jankowiak & Korpal, 2018). Similar SCR levels in the first and second language contradict the findings of studies pointing to an increased emotional weight of the mother tongue. Among these studies is e.g., Harris *et al.* (2003), reporting greater autonomic reactivity towards taboos and reprimands in the first than in the second language. Another possible explanation for no significant difference found between SCRs to critical and control items in L2 English is that usable data in this study only came from 20 participants. Such a reduced sample size presents a considerable limit to generalisability. A future extension of this work with more participants will be in a good position to address this power issue.

In relation to its closest empirical context, the SCR results of this study are in line with those of Caldwell-Harris *et al.* (2011), who observed comparable electrodermal activity of Chinese-English balanced bilinguals in reaction to insults, reprimands and taboo words. They are also consistent with the findings of Harris *et al.* (2003), where comparable arousal of the autonomic nervous system was found for aversive words such as *danger* and *fight* in the L1 and L2. The discrepancy in the results for taboo words and reprimands found between the studies of Harris et al. (2003) and Caldwell-Harris *et al.* (2011) was explained by the difference in the level of bilingual participants' L2 proficiency (unbalanced and balanced bilinguals, respectively). The same explanation does not hold for the present study as the participants were late unbalanced bilinguals, learning English as an L2 in an instructed context rather than a naturalistic setting for most of their lives. It is therefore unlikely that an arguably lower L2 proficiency compared to the L1 would be the driving factor to have equalised the SCR levels in the L2 compared to the native language. Developing context-dependent L2-to-concept associations are an alternative candidate to which one could attribute the similarity of SCRs found across languages.

## Context as a booster of emotive associations with L2 expressions

Strengthened associations between L2 expressions and emotive concepts were reported in previous work for instance for endearments (Caldwell-Harris et al, 2011), highlighting the importance of an emotionally salient context of acquisition and L2 use for a shift from mother tongue dominance to an increased emotional power of the second language. In this study, strong associations between emotive university-related expressions in the L2 and the related concepts can be explained by secondary affective socialization at a UK university. Immersion in the L2 context could be the ideal trigger for strengthening links between the emotive concepts in question and their expression in the L2 English because of the high relevance of such associations for participants in their recent past and at the time of testing. It is plausible that the elicitation of the two greatest SCRs by L2 (rather than L1) expressions is attributable to their highest relevance in the participants' recent academic past. Although intuitively appealing, the assumption of SCR variation changing as a function of recent relevance remains open for future inquiry.

An important point to emphasize is that secondary socialization might have increased emotional reactivity towards English university-related words to the level of those in native Russian, but overall reactivity to L2 still did not exceed reactivity to L1. Comparable levels of reactivity across languages suggest that the tested emotion concepts are not exclusively L2-based. Although immersion in the L2 context and secondary affective socialization could have played a great role in boosting the force of L2 expressions, L1 reactivity found on a par with L2 reactivity signals a direct link of similar strength also between the L1 expressions and the related concepts. An alternative view is that in the dynamics of their L2 development, the participants of this study were late unbalanced bilinguals whose L2-to-concept mapping can be characterised as largely L1-mediated (Kroll & Stewart, 1994). Continued and reinforced context-dependent learning (Marian & Kaushanskaya, 2008) may be able to pack a more powerful emotional punch in the L2 than in the L1 later in their L2 development, but for participants in this study, self-identified as L1-dominant, this stage has not (yet) been reached.

Not to exclude the possibility that L2-based concepts with high contextual relevance could elicit stronger reactivity to L2 expressions than to their L1 analogues, it is also important to consider the idea that second language processing could have added cognitive cost reflected in the SCRs for English items. In other words, reactivity towards the L2 may have reduced and matched the level of reactivity to the native language stimuli not because of their equal emotional arousal, but rather due to higher mental demands needed to process L2 English. This explanation builds on Caldwell-Harris and Ayçiçeği-Dinn (2009) and their discussion of the discrepancy between SCR levels and emotionality ratings. No such discrepancy emerged in the results of this study, but the claim of increased cognitive demands in L2 processing, especially online, may still apply. If the L2 was cognitively more demanding to

process, there must have been another facilitative factor to balance out the magnitude of autonomic responses to the L2 and the L1. We attribute such facilitation to language-specific autobiographic memory (Marian & Kaushanskaya, 2008), namely to a high saliency of university-related expressions habitually encountered in an emotionally resonant L2 context.

# References

Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition & Emotion, 8(6)*, 503-514.

Baumeister, J. C., Foroni, F., Conrad, M., Rumiati, R. I., & Winkielman, P. (2017). Embodiment and emotional memory in first vs. second language. *Frontiers in Psychology, 8*, 394.

Beck, A.T., Epstein, N., Brown, G., & Steer, R.A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893-897.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: theoretical implications. *Frontiers in Psychology, 5,* 1055.

Caldwell-Harris, C. L., Tong, J., Lung, W., & Poo, S. (2011). Physiological reactivity to emotional phrases in Mandarin—English bilinguals. *International Journal of Bilingualism, 15(3),* 329-352.

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology, 71(3),* 193-204.

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.

Dewaele, J. M. (2008). The emotional weight of *I love you* in multilinguals' languages. *Journal of Pragmatics, 40(10)*, 1753-1780.

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.

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.

Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion, 21(5)*, 1064-1076.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28(1)*, 1-11.

Fan, L., Xu, Q., Wang, X., Xu, F., Yang, Y., & Lu, Z. (2018). The automatic activation of emotion words measured using the emotional face-word Stroop task in late Chinese–English bilinguals. *Cognition and Emotion, 32(2)*, 315-324.

Ferré, P., Anglada-Tort, M., & Guasch, M. (2018). Processing of emotional words in bilinguals: Testing the effects of word concreteness, task type and language status. *Second Language Research, 34(3)*, 371-394.

Frie, R. (2011). Irreducible cultural contexts: German–Jewish experience, identity, and trauma in a bilingual analysis. *International Journal of Psychoanalytic Self Psychology, 6(2)*, 136-158.

Grabovac, B., & Pléh, C. (2014). Emotional activation measured using the emotional Stroop task in early Hungarian-Serbian bilinguals from Serbia. *Acta Linguistica Hungarica, 61(4)*, 423-441.

Grosjean, F. (2010). *Bilingual. Life and Reality.* Cambridge, MA: Harvard university press.

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.

Harris, C. L., Gleason, J. B., & Ayçíçeğí, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (ed.) *Bilingual Minds: Emotional Experience, Expression, and Representation*, pp. 257-283. Clevedon: Multilingual Matters.

Harris, C. L., Ayçíçeğí, 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.

Hugdahl, K. (2001). *Psychophysiology: The mind–body perspective*. Cambridge, MA: Harvard University Press.

Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PloS One, 12(10)*, e0186027.

Jankowiak, K., & Korpal, P. (2018). On modality effects in bilingual emotional language processing: Evidence from galvanic skin response. *Journal of Psycholinguistic Research, 47(3)*, 663-677.

Jończyk, R., Boutonnet, B., Musiał, K., Hoemann, K., & Thierry, G. (2016). The bilingual brain turns a blind eye to negative statements in the second language. *Cognitive, Affective, & Behavioral Neuroscience, 16(3)*, 527-540.

Kokaliari, E., Catanzarite, G., & Berzoff, J. (2013). It is called a mother tongue for a reason: A qualitative study of therapists' perspectives on bilingual psychotherapy - Treatment implications. *Smith College Studies in Social Work, 83(1)*, 97-118.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33(2)*, 149-174.

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion, 8(4)*, 445-452.

Marian, V., & Kaushanskaya, M. (2004). Self-construal and emotion in bicultural bilinguals. *Journal of Memory and Language, 51(2)*, 190-201.

Ożańska-Ponikwia, K. (2019). Expression and perception of emotions by Polish–English bilinguals *I love you* vs. *Kocham Cię*. *International Journal of Bilingual Education and Bilingualism, 22(4)*, 493-504.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and cognition, 11(2),* 147-164.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47(6)*, 405-428.

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.

West, R., & Frumina, E. (2012). European standards in Russian higher education and the role of English: a case study of the National University of Science and Technology, Moscow (MISiS). *European Journal of Education, 47(1),* 50-63.

Winskel, H. (2013). The emotional Stroop task and emotionality rating of negative and neutral words in late Thai–English bilinguals. *International Journal of Psychology, 48(6)*, 1090-1098.

Appendix: Average ratings and skin conductance responses per item presented in L2 English and L1 Russian

| Number | Item | Type | Rating M(SD) | SRC M(SD) |
|---|---|---|---|---|
| 1 | Deadline is tomorrow<br>*Завтра крайний срок* | Critical | 2.79 (0.71)<br>2.47 (0.81) | 1.83 (0.83)<br>1.77 (0.96) |
| 2 | You failed the exam<br>*Вы провалили экзамен* | Critical | 2.74 (0.99)<br>2.79 (0.92) | 1.69 (0.80)<br>1.73 (0.77) |
| 3 | You are expelled<br>*Вы отчислены из университета* | Critical | 2.68 (0.82)<br>2.79 (1.03) | 1.73 (0.79)<br>1.79 (0.75) |
| 4 | Submit your dissertation<br>*Пора сдавать диссертацию* | Critical | 2.58 (0.84)<br>2.37 (0.90) | 1.67 (0.81)<br>1.72 (0.83) |
| 5 | You did not pass<br>*Вы не сдали* | Critical | 2.53 (1.07)<br>2.63 (0.88) | 1.82 (0.78)<br>1.78 (0.96) |
| 6 | Assignment<br>*Задание* | Critical | 2.00 (0.92)<br>1.58 (0.69) | 1.73 (0.78)<br>1.72 (0.79) |
| 7 | Presentation<br>*Презентация* | Critical | 1.95 (0.90)<br>1.68 (0.67) | 1.71 (0.77)<br>1.72 (0.75) |
| 8 | Fees payment<br>*Оплата расходов* | Critical | 1.95 (0.85)<br>1.89 (0.88) | 1.63 (0.75)<br>1.73 (0.88) |
| 9 | Write an essay<br>*Напишите эссе* | Critical | 1.89 (0.66)<br>1.89 (0.81) | 1.66 (0.75)<br>1.70 (0.79) |
| 10 | Plagiarism<br>*Плагиат* | Critical | 1.89 (0.78)<br>2.11 (0.94) | 1.64 (0.76)<br>1.70 (0.75) |
| 11 | Market<br>*Рынок* | Control | 1.74 (0.41)<br>1.15 (0.37) | 1.69 (0.77)<br>1.64 (0.40) |
| 12 | News<br>*Новости* | Control | 1.50 (0.51)<br>1.40 (0.60) | 1.66 (0.76)<br>1.65 (0.42) |
| 13 | Office<br>*Офис* | Control | 1.45 (0.69)<br>1.40 (0.60) | 1.69 (0.76)<br>1.68 (0.43) |
| 14 | Black<br>*Черный* | Control | 1.20 (0.56)<br>1.15 (0.37) | 1.66 (0.75)<br>1.63 (0.37) |
| 15 | Manner<br>*Характер* | Control | 1.15 (0.36)<br>1.10 (0.32) | 1.63 (0.79)<br>1.62 (0.43) |
| 16 | Lion<br>*Лев* | Control | 1.15 (0.37)<br>1.10 (0.32) | 1.60 (0.77)<br>1.59 (0.42) |
| 17 | Key<br>*Ключ* | Control | 1.10 (0.31)<br>1.10 (0.32) | 1.62 (0.79)<br>1.64 (0.44) |
| 18 | Plant<br>*Растение* | Control | 1.05 (0.22)<br>1.05 (0.23) | 1.73 (0.79)<br>1.72 (0.40) |
| 19 | Lamp<br>*Лампа* | Control | 1.05 (0.22)<br>1.05 (0.23) | 1.58 (0.76)<br>1.64 (0.45) |
| 20 | *Museum*<br>Музей | *Control* | *1.00 (0.00)*<br>*1.10 (0.23)* | *1.63 (0.76)*<br>*1.59 (0.40)* |