Manuscript

Processing advantage for emotional words in bilingual speakers

Marta Ponari[1], Sara Rodríguez-Cuadrado[2,3], David Vinson[1], Neil Fox[1], Albert Costa[3,4], Gabriella Vigliocco[1]

[1]Experimental Psychology Department, Division of Psychology and Language Sciences, University College London, United Kingdom

[2]Hospital Sant Joan de Déu, Barcelona, Spain

[3]Departament de Tecnologies de Informació I les Comunicacions
Universitat Pompeu Fabra, Barcelona, Spain

[4]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Corresponding Author: Marta Ponari, Experimental Psychology Department, UCL, 26 Bedford Way, London WC1H 0AP, UK. email: m.ponari@kent.ac.uk

ABSTRACT

Effects of emotion on word processing are well established in monolingual speakers. However, studies that have assessed whether affective features of words undergo the same processing in a native and non-native language have provided mixed results: studies that have found differences between L1 and L2 processing, attributed it to the fact that a second language (L2) learned late in life would not be processed affectively, because affective associations are established during childhood. Other studies suggest that adult learners show similar effects of emotional features in L1 and L2. Differences in affective processing of L2 words can be linked to age and context of learning, proficiency, language dominance, and degree of similarity between the L2 and the L1. Here, in a lexical decision task on tightly matched negative, positive and neutral words, highly proficient English speakers from typologically different L1 showed the same facilitation in processing emotionally valenced words as native English speakers, regardless of their L1, the age of English acquisition or the frequency and context of English use.

Keywords: word processing; emotion; bilingualism

Language processing is affected by emotional content and this is the case even for single words (e.g., Altmann, Bohrn, Lubrich, Menninghaus & Jacobs, 2012). Studies with monolinguals have found that words with affective features (valenced words) are processed differently than neutral words. In particular, once factors such as length and familiarity are taken into account, both negative and positive words are processed faster than neutral words in lexical decision tasks (a relatively shallow task in which subjects are required to distinguish between words and nonwords; Kousta, Vinson & Vigliocco, 2009; see Vinson, Ponari & Vigliocco, 2014 for a discussion). Effects of valence on lexical decision reaction times are also modulated by frequency: positive and negative words are processed faster than neutral among low frequency words, while positive words are processed faster than neutral and negative among high frequency words (Kuchinke, Vö, Hofmann & Jacobs, 2007; Scott, O'Donnell, Leuthold & Sereno, 2009; see also Kuperman, Estes, Brysbaert & Warriner, 2014). Although an advantage for negative words has not been consistently reported, especially using tasks such as the Emotional Stroop task where negative words tend to lead to slower RTs (but see Larsen, Mercer, and Balota, 2006 for a review), a general difference in processing emotional vs. neutral words is well established.

The majority of work on emotional language has been carried out with monolingual (especially English speaking) participants, however, in the last decade the interest on emotion and bilingualism has increased (see Pavlenko, 2008, 2012 for reviews). The central question is whether and to what extent affect is associated to the second language, given that emotional associations of words might be established especially in childhood (e.g., in childhood reprimands/endearments; Pavlenko, 2012). In general, the first language (L1) is considered more emotional than the second (L2) or subsequent languages (Lx) (Altarriba, 2008; Dewaele, 2004; Pavlenko, 2008).

According to this view, in balanced early bilinguals both L1 and L2 may be closely attuned to the emotional content associated with each language, whereas in late sequential bilinguals the L2 is associated with greater emotional distance (Altarriba, 2008; Dewaele & Pavlenko, 2002), as it is typically learned in formal contexts (e.g. school, university, or workplace), where “the same opportunities for affective linguistic conditioning” are missing compared to childhood (Pavlenko, 2008, *p*.156). Therefore, particularly in late bilinguals, the emotional valence of L2 words is considered to be “disembodied” as L2 words would be processed semantically but not affectively (Pavlenko, 2012, p. 405).

Studies using subjective measures (questionnaires) have shown that bilinguals experience their L1 as more emotional than their L2 (e.g., Dewaele, 2004; Pavlenko, 2004, 2005): regardless of proficiency, bilinguals often report that although they *know* the emotional meaning of words in L2, they don’t *feel* it (Pavlenko, 2005). Self reports, however, can be related to confounding variables like cultural stereotypes (Degner, Doycheva & Wentura, 2012). Studies using objective measures (such as RTs) provide inconsistent results, showing in some cases an advantage in processing emotional words in L1 (Anooshian & Hertel, 1994), in other cases an advantage in L2 (Ayçiçegi & Harris, 2004) or no difference between languages (Ayçiçegi-Dinn & Caldwell-Harris, 2009; Eilola, Havelka & Sharma, 2007; Ferré, García, Fraga, Sánchez-Casas & Molero, 2010; Ferré, Sanchez-Casas & Fraga, 2012; Segalowitz, Trofimovich, Gatbonton & Sokolovskaya, 2008; Sutton, Altarriba, Gianico & Basnight-Brown, 2007). EEG studies using a lexical decision task suggest weaker (Conrad, Recio & Jacobs, 2011) or delayed (Opitz & Degner, 2012) automatic affective processing in L2.

Differences in affective processing of L2 words can be linked to age and context of learning (by instruction, in the classroom settings, or by immersion, as moving to the L2 speaking country), proficiency and language dominance. More specifically, high proficiency and high frequency of L2 use could lead to direct and strong associations between L2 words and corresponding semantic concepts (see Kroll & Stewart, 1994) and thus to higher accessibility of affective properties. Consistently, Degner et al. (2012) found affective priming in L2 only in participants who used their L2 frequently in everyday life.

Differences in results of previous studies may be related to the way in which these and other variables are taken into account. First, age of L2 acquisition and context of learning differ between and even within studies. Second, the type of task differs (self-reports, interviews, emotional Stroop and memory studies); most of the aforementioned studies used tasks in which participants are explicitly asked to focus on words' emotional content, which does not allow to disentangle whether the reduced affective processing is due to a general slower semantic processing in L2 (see Degner et al., 2012; Segalowitz et al., 2008). Third, the materials differ: some studies only used words referred to emotions (e.g., *happy, sad* etc.), others words with emotional associations (*promotion, crime* etc., that could be emotion words or not), other studies words like taboo, swearwords, reprimands and endearments. Finally, across studies, the degree of typological similarity between L1 and L2 also differs. This may be particularly important because processing cognates (words that share phonological and/or orthographic form although they are not always translation equivalent; Hall, 2002) is easier than processing words that have different phonological forms (e.g., Odlin, 1989).

Here, we assess L2 processing of valenced words controlling most of the confounding variables present in previous studies: first, we compare early and late learners of English to native speakers, thus taking into account effects of age of English acquisition. Second, we use a task that does not require to focus on emotion (i.e., lexical decision), to tap into automatic processing. Third, we test participants whose L1s vary across language families (including non-Latin-script languages and sign languages), taking into account the degree of similarity between English and individual L1s, and the presence of cognates. Finally, we control for effects of L2 immersion (duration of stay in a L2 speaking country, daily frequency and domain (personal/professional) of L2 use).

If the emotional content of words is less available in L2, especially when L2 is learned late in life, then late English learners should not show the same emotional effect as native speakers (and early learners). However, models of L2 acquisition and processing (e.g. Kroll & Stewart, 1994) would suggest that proficient L2 learners process affective connotations of words just like native speakers, regardless of the age of English acquisition. If this is the case, we should not find any difference in the size of the emotion effect between L2 and native speakers. Language immersion, though, could modulate the emotion effect, in which case we would find differences related to whether the participants are resident or not in an English speaking country, the duration of stay, and the frequency of L2 use.

METHODS

Participants

Data from ninety-five native English speakers (Native: 60 females, mean age = 22.58, SD = 7.30) and 156 participants with English as second language (L2) were collected over the course of two years as part of a course requirement or for monetary compensation. As in Harris (2004), L2 participants were classified as Early L2 (N = 77, 65 females, mean Age of English Acquisition (AEA) = 4.07, SD = 1.89, mean age = 22.63, SD = 6.77) or Late L2 (N = 79, 60 females, mean AEA = 11.95, SD = 2.88, mean age = 31, SD = 9.03). A subset of L2 participants (N = 94) completed an adapted version of the Language History Questionnaire (LHQ; Li, Sepanski & Zhao, 2006). Proficiency was high (average self-rating of reading, writing, speaking and listening abilities for L2 participants was 5.98 (SD = 0.76) on a 1-to-7 Likert scale where 1 = "very poor" and 7 = "native-like"). The remaining L2 participants were undergraduate students of Psychology (N = 49), living in the UK for at least 2 years, who fulfilled the English language qualification requirements for enrolment in a UCL Faculty of Brain Sciences undergraduate program (advanced level at GCSE/IGCSE English language[1]); or BSL signers (N = 13), who were all born in the UK and lived there for the majority of their lives.

L2 participants' native languages were diverse, ranging from Indo-European languages (Dutch, German, Danish, Spanish, French, Swedish, Norwegian) to Sign languages (BSL) to non-Latin scripts (Cantonese, Chinese, Hindi, Japanese, Malay, Mandarin, Russian). All participants had normal or corrected to normal vision and gave informed written consent to the study.

[1] General Certificate of Secondary Education (GCSE) is an academic qualification in England, Wales and Northern Ireland, typically taken at age 16. International GCSE is a comparable international qualification.

Materials

Stimuli for the experiment were taken from Kousta et al. (2009), excluding two triplets that were also excluded from analyses of the original study, and one triplet due to repetition of the word "victory", leaving 111 words (37 positive, 37 negative and 37 neutral; stimuli characteristics are summarized in Supplemental Table 1) and 111 nonwords. Nonwords were created by changing a letter from additional positive, negative and neutral words; resulting nonwords were pronounceable, orthographically licit and matched pairwise with the experimental items in terms of length.

Procedure

Before carrying out the lexical decision task, all participants completed a brief questionnaire that included questions concerning age of English acquisition, native language, and other spoken languages. A subset of participants also completed the LHQ test, as described in the Participants section.

In the lexical decision task, stimuli were displayed in black Courier New 18 point font on a white background with a central fixation cross, viewed from comfortable distance. Stimulus presentation was controlled by E-Prime v.1.2 (Schneider, Eschman & Zuccolotto, 2002). On each trial, the fixation cross was displayed for 400ms, followed by the letter string, which remained on the screen until a response was given. Participants were instructed to respond as quickly and accurately as possible, by pressing the "j" or the "f" key on a standard UK keyboard with the right or left index finger respectively. A blank inter-trial interval of 1000ms followed each response. Each participant first completed 12 practice items with feedback on accuracy.

Exclusion criteria

Participants were excluded from the main analyses if they had overall accuracy under 80%, or mean RT over 1000ms, leaving 88 Native, 74 Early and 78 Late L2 participants. At a trial level, RTs faster than 250ms or slower than 2500ms (2.74% of the data) as well as error trials (overall mean accuracy = 0.95; SD = 0.22), were excluded from analysis of RTs.

RESULTS

Are the emotion effects modulated by Age of English Acquisition?

A mixed ANOVA with valence (negative, neutral, positive) as within subject variable and AEA group (Native, Early, Late) as between subject variable, showed a significant main effect of valence, $F(1.898, 449.7$[2]$) = 95.366$, $p <$ 001, $\eta^2_p = .207$, with neutral words (mean RT = 657.9) being recognised slower than negative (mean = 635.1; $t(239) = 9.608$, $p < .001$) and positive words (mean = 628.44; $t(239) = 12.247$, $p < .001$). Negative words were also slower than positive ($t(237) = 3.339$, $p = .001$). There was a main effect of AEA group, $F(2, 237) = 29.517$, $p < .001$, $\eta^2_p = .199$, with Late L2 participants (mean = 710.6) being slower than both Native (mean = 609.6) and Early L2 participants (mean = 603.2). The interaction between valence and AEA was also significant, $F(3.794, 449.725) = 2.716$, $p = .032$, $\eta^2_p = .022$ (see Figure 1). Paired sample t-tests by AEA group revealed that both Native and Early L2 participants were faster with negative and positive words compared to neutral (Native: negative vs neutral: $t(87) = 5.519$, $p < .001$; positive vs neutral: $t(87) = 6.762$, $p <$

[2] The significance levels of the ANOVAs reported here were, whenever necessary, adjusted according to the Greenhouse-Geisser procedure to compensate for violations of the sphericity assumption, as recommended by Maxwell and Delaney (1990).

.001; no difference between negative and positive (p = .457); Early L2: negative vs neutral: $t(73) = 4.760$, $p < .001$; positive vs neutral: $t(73) = 6.638$, $p < .001$; no difference between negative and positive (p = .095)). Late L2 participants were also faster with negative and positive words compared to neutral (negative vs neutral: $t(77) = 6.325$, $p < .001$; positive vs neutral: $t(77) = 8.075$, $p < .001$) but were also significantly faster with positive words compared to negative, $t(77) = 3.247$, $p = .002$).

In summary, we found that positive and negative words are processed faster than neutral words in a lexical decision task, thus replicating previous results on native speakers and extending them to non-native participants, regardless of AEA.

~ Please insert Figure 1 about here ~

As mentioned in the introduction, however, the effect of valence might be confounded with the effect of other variables, both related to specific characteristics of the words, like frequency or concreteness, that are known to exert a strong effect on reaction times[3]; or related to the subjects, such as the similarity between their L1 and English, or the degree of English immersion/dominance.

To take these into account, we conducted further analyses using linear mixed effect models (package lme4 v. 1.0-5: Bates & Maechler, 2009, running in R version 3.0.1 (R Core Team, 2013)). Mixed effect models are a variety of regression models,

[3] Although our items were matched for potentially confounding variables across valence categories, such variables may still play a role once error trials and outlying RTs are removed and thus the item-level matching is compromised.

commonly used in psycholinguistics, that take into account variation that is not generalizable to the independent variable (allowing therefore to consider variation across different subjects and different stimuli, specific to the data sample).

We started fitting a baseline model which included the following non-emotional factors: familiarity, imageability, concreteness (from Coltheart, 1981); age of acquisition (from Stadthagen-Gonzalez & Davis, 2006), and word frequency (SUBTLEX-UK; Van Heuven, Mandera, Keuleers & Brysbaert, in press), along with our variables of interest which were: valence and Age of English acquisition (AEA). Of particular interest was the interaction between valence and AEA (by items); we also included the interaction between valence and word frequency because Kuperman et al. (2014) reported that frequency (specifically SUBTLEX) modulates valence effects for L1. Words were classified as high frequency (Zipf-values = 4-7) or low frequency (Zipf-values = 0-3). Model fits included random intercepts as well as random slopes for both participants and items, thus starting with a maximal, "design-driven" random effects structure (Barr, Levy, Scheepers & Tily, 2013). As the model including random slopes for all predictors did not converge, we removed the slopes of the non-emotional variables by subject and item before proceeding further, leaving only the random slopes for variables of interest (by subject: valence and word frequency; by item: AEA group). Log-likelihood ratio tests were used to compare fitted models.

To test for effects of Age of English Acquisition (AEA), we first tested a model including the three-way interaction between valence, word frequency and AEA group (Native, Early L2, Late L2) against a model including the three combinations of two-way interactions and the main effects. Including the three-way interaction did not significantly improve the fit of the model (log-likelihood ratio for interaction model = -167040; log-likelihood ratio for two-way interactions model = -167043; $\chi^2(4)$ = 5.787, p = .216). Having established that a three-way interaction is not warranted, we then followed the same strategy to test whether two-way interactions were warranted; removing each one in turn and assessing whether the model including that interaction was significantly better than a comparable model without it.

These analyses did not provide evidence for a word frequency × valence interaction: keeping this interaction in the model did not significantly improve the fit (log-likelihood ratio for the model including the interaction = -167043; log-likelihood ratio for the model without it = -167044; $\chi^2(2)$ = 0.8564, p = .652).

However, we found a significant improvement gained by including the frequency × AEA interaction in the model, compared to the model without it (log-likelihood ratio for the model including the interaction = -167044; log-likelihood ratio for the model without it = -167064; $\chi^2(2)$ = 47.781, p < .001). Late L2 participants were particularly slower with low frequency words.

Crucially for the purpose of this study, we found no improvement by keeping the Valence × AEA interaction term against a model which did not include it (log-likelihood ratio for the model including the interaction = -167044; log-likelihood ratio for the model without it = -167045; $\chi^2(4)$ = 2.5604, p = .6339).

As a final check, we tested whether the effect of valence persisted in this model by removing the main effect of valence. In this case, keeping valence in significantly improved the fit of the model (log-likelihood ratio for the model including valence = -167045; log-likelihood ratio for the model without it = -167059; $\chi^2(9) = 27.642$, $p = .001$).

The final model at this stage thus included the main effects of the control variables as well as the main effects of valence, word frequency and AEA group, and the frequency × AEA group interaction. Partial effects of the control variables are shown in Figure 2.

~ Please insert Figure 2 about here ~

As this omnibus analysis does not reveal the precise nature of the valence effect, nor the group by frequency interaction, we conducted additional analyses, contrasting the performance of Early and Late L2 speakers with those of Native speakers, separately, and using linear mixed effects models. We used two-level factors (along with the same control variables as in the final model above) and collapsed together positive and negative words ("valenced"). Furthermore, we centred all continuous predictors and contrast-coded all of the fixed effects of interest to ensure that interactions and main effects were orthogonal, thus permitting clearer interpretation of the parameter estimates. Finally, although there was no evidence of valence × AEA interaction in the omnibus model reported above, we nonetheless included this interaction in our subsequent models to be able to capture any tendencies that might have been obscured.

Native vs Early L2 speakers

Valenced vs Neutral: the main effect of valence was significant (estimate of the Valenced-Neutral difference = -16.73, SE = 6.48, t = -2.58, p = .01[4]) as was the main effect of frequency (estimate of the High-Low frequency difference = -32.79, SE = 7.98, t = -4.11, p < .001). There was no significant main effect of AEA (t = 0.27, p = .787), and no significant valence × AEA interaction (coefficient estimate[5] = -1.69, SE = 5.08, t = -0.33, p = .741). However, the frequency × AEA interaction was significant (coefficient estimate = 14.39, SE = 6.83, t = -2.11, p = .035).

Positive vs Negative: there was no significant main effect of valence (estimate of the Positive-Negative difference = 0.78, SE = 7.53, t = 0.10, p = .920) or AEA group (t = 0.23, p = .818), and no valence × group interaction (coefficient estimate = 1.01, SE = 5.66, t = 0.18, p = .857). The main effect of frequency was significant (estimate of the High-Low frequency difference = -33.06, SE = 9.52, t = -3.47, p < .001) as was the frequency × AEA group interaction (coefficient estimate = 14.04, SE = 7.17, t = 1.96, p = 049)

Native vs Late L2 speakers

Valenced vs Neutral: there was a significant main effect of frequency (estimate of the High-Low frequency difference = -57.16, SE = 17.57, t = -3.25, p = .001) and a significant main effect of AEA (estimate of the Native-Late difference = -110.75, SE = 18.61, t = -5.95, p < .001), with Late L2 participants being overall slower than Native speakers. The valence × AEA (coefficient estimate = 2.99, SE = 26.57, t = 0.11, p = .912) and frequency × AEA group (coefficient estimate = 51.61,

---

[4] Here and in the following sections, we report p-values based on the normal approximation to the *t*-distribution, under the null hypothesis that a coefficient is drawn from a distribution with mean of zero. As degrees of freedom increase, the t-distribution converges to *z*.

[5] In the case of interactions, 'coefficient estimate' refers to the difference of difference scores. In this case it refers to the 'Native(Valenced-Neutral)-Early(Valenced-Neutral)' difference, thus testing whether the magnitude of the valence effect differs for Native and Early L2 participants.

SE = 28.79, t = 1.79, p = .074) were not significant. The main effect of valence was not significant (estimate of the Valenced-Neutral difference = -19.22, SE = 15.94, t = -1.21, p = .226), probably due to slowest and more variable reaction times in Late L2 speakers, as when we contrasted positive to neutral and negative to neutral words in separate models, both of these valence effects were again reliable, (estimate of the Positive-Neutral difference = -19.84, SE = 8.89, t = -2.23, p = .026; estimate of the Negative-Neutral difference = -20.73, SE = 10.08, t = -2.06, p = .039), consistent with all the other analyses in which emotional words elicit faster responses than neutral words regardless of frequency and AEA group.

Positive vs Negative: there was no significant main effect of valence (estimate of the Positive-Negative difference = -2.64, SE = 11.31, t = -0.23, p = .818). Again, the main effects of frequency (estimate of the High-Low frequency difference = -59.24, SE = 11.31, t = -4.41, $p < .001$) and AEA group (estimate of the Native-Late difference = -110.70, SE = 16.66, t = -6.64, $p < .001$) were significant. There was no significant valence × AEA group interaction (coefficient estimate = 6.68, SE = 12.11, t = 0.55, p = .582), but the frequency × AEA group interaction was significant (coefficient estimate = 58.58, SE = 13.93, t = 4.20, $p < .001$).

Are the emotion effects modulated by language similarities?

To assess whether the degree of similarity between languages and in particular the presence of cognates could have had a role in the processing advantage of emotionally valenced over neutral words, we conducted additional analyses testing for the role of language family.

Early L2 participants' native languages were classified as Germanic (Dutch, German, Danish, Swedish; N = 18), Romance (Spanish, French; N = 20), Non-Latin-Script languages (Cantonese, Chinese, Hindi, Japanese, Malay, Mandarin, Russian; N = 23) or Sign languages (British Sign Language, N = 13); Late L2 participants' native languages were classified as Germanic (Dutch, German, Norwegian; N = 42) or Romance languages (Spanish, French; N = 36). The models also included the same control variables as in the main analyses (familiarity, imageability, word frequence, concrenetess, age of acquisition).

We started testing a model that included the two-way interaction between valence and language family, keeping AEA group (Early L2, Late L2) as a main effect (as we did not find a significant AEA × valence interaction in the main mixed models analysis), against a model that did not include the valence × language family interaction. We found a marginally significant improvement gained by adding the interaction (log-likelihood ratio for interaction model = -80327; log-likelihood ratio for the model without it = -80333; $\chi^2(6) = 12.024$, $p = .06$). Participants with Germanic (coefficient estimate = 101.81, SE = 40.66, $t = 2.50$, $p = .012$) and Romance (coefficient estimate = 124.54, SE = 41.79, $t = 2.98$, $p = .003$) were significantly slower. As this analysis does not show sufficient evidence to suggest a valence × language family interaction, and this could be modulated by AEA, we then conducted follow up analyses formally testing for interactions between language family, valence and AEA using only Romance and Germanic language families in order to permit a fully factorial design. As in the main analyses above, we used two-level factors and contrasted emotional versus neutral words. Continuous predictors (the same control variables as in the final model above) were centred and fixed effects were contrast-coded. We tested a model including the three-way interaction between valence, language family (Romance, Germanic) and AEA group (Early L2, Late L2) against a model including the three combinations of two-way interactions and the main effects. There was no significant improvement gained by adding the three-way interaction (log-likelihood ratio for interaction model = -56364; log-likelihood ratio for two-way interactions model = -56364; $\chi^2(1) = 0.854$, $p = .356$). There was no reliable main effect of language family (coefficient estimate = 45.19, SE = 29.50, $t = 1.54$) or AEA (coefficient estimate = -33.33, SE = 30.07, $t = -1.11$, $p = .227$). Most crucially, the main effect of valence was still reliable (coefficient estimate = -24.77, SE = 12.07, $t = -2.05$, $p = .040$), but there was no evidence for a valence × language family interaction (coefficient estimate = -11.18, SE = 8.33, $t = -1.34$, $p = .018$).

We also looked at the role of language similarities in terms of presence of cognates. Words were classified as cognates or noncognates for each individual L1 (again focusing only on Romance and Germanic L1 families, that were more likely to be similar to English; words were classified as cognates by 2 native speakers of each language). Cognate status was included as a fixed effect and as a random slope by participants in a model contrasting valenced and neutral words for Early L2 and Late L2 participants. The model also included the same control variables as in the main analyses.

Cognate status was a significant predictor of RTs (coefficient estimate = -20.78, SE = 6.87, $t = -3.02$, $p = .003$), but the effect of valence persisted (coefficient estimate = -33.89, SE = 11.02, $t = -3.08$, $p = .002$), with no difference between Early and Late L2. A model including the three-way interaction between cognate status, valence and AEA group did not add significant improvement (log-likelihood ratio for interaction model = -45487; log-likelihood ratio for main effects model = -45487; $\chi^2(3) = 0.893$, $p = .827$).

Are the emotion effects modulated by frequency and context of L2 use?

The processing advantage of valenced words that we found in both Early and Late L2 learners was not modulated by language similarities or the presence of cognates. However, as suggested by Degner et al. (2012), the processing of words' valence in L2 is largely determined by the frequency and context of L2 use. We performed additional analyses in order to investigate the role of frequency of L2 use in the valence effect found in Late L2 learners. From the participants who completed the LHQ (see methods), we first contrasted a subgroup of Late L2 learners who were resident in the UK or another English speaking country at the time of testing (N = 36) with a subgroup of Late L2 learners who had never lived in an English speaking country (N = 20). Second, only for the participants living in the UK or another English speaking country, we tested a model that included as continuous predictors the duration of stay in the English speaking country (in years), and the frequency of daily English use in the private-personal (partner/family/friends) or in the professional (study/job) context. Participants in this analysis had lived in an English speaking country for an average of 3.7 years (SD = 3.1, range = 0.2-11) and used English in around 60% (range = 0%-100%) of their personal daily interactions, and around 90% (range = 50%-100%) of professional daily interactions. Both models again included the same control variables as in the previous analyses.

The main effect of country of residence (English speaking vs non-English speaking) was significant (coefficient estimate = -109.50, SE = 29.87, t = -3.67, $p <$

.001), with participants resident in an English speaking country being overall faster than participants who lived in a non-English speaking country. The main effect of valence (coefficient estimate = -25.29, SE = 12.48, t = -2.03, p = .042) was still significant, with, crucially, no interaction between valence and country of residence (coefficient estimate = 5.68, SE = 10.51, t = 0.54, p = .589) (see Figure 3).

In addition, none of the measures of language immersion were significant predictors of reaction times (duration of stay: coefficient estimate = 1.14, SE = 4.32, t = 0.26, p = .794; frequency of private/personal use: coefficient estimate = 5.72, SE = 52.23, t = 0.11, p = .912; frequency of professional use: coefficient estimate = -70.13, SE = 81.09, t = -0.87, p = .384).

~ Please insert Figure 3 about here ~

DISCUSSION

Using a lexical decision task with positive, negative and neutral words, we found no difference in the type and direction of the emotion effect between native speakers and both late and early high proficient L2 speakers. Valenced words were recognised faster than neutral words by both Native speakers and L2 participants, despite Late L2 being overall slower than Native speakers in performing the task. This valence effect was not modulated by language similarity or language immersion: we found comparable emotion effects across all language families, regardless of AEA, with no evidence for an effect of cognates and, crucially, no evidence for an effect of country of residence, duration of stay, frequency and domain of language use.

For native speakers, these results replicate previous findings showing that valence, regardless of polarity, facilitates processing of words (Kousta et al., 2009). Note that despite using the same set of stimuli as in Kousta et al. (2009), here we did not only rely on matching lexical and sublexical variables between valence categories, but instead we explicitly modelled variation among these control variables (in addition to simultaneously treating participants and items as random effects), in order to adequately control potentially confounding variables. Moreover, the present study further tested for the role of frequency, which has been argued to be important in modulating valence effects especially for negative words (e.g., Kuperman et al., 2014). It remains to be seen whether the interaction observed by Kuperman et al. depends upon specific characteristics of the item set tested, but here we find no indication of such an interaction.

Regarding processing of L2, our findings contrasting early, late and native speakers provide novel and strong evidence that valenced English words are processed in a similar way by Native, Early and Late L2 highly proficient speakers with a range of different L1s. Thus, once AEA is taken into account, and when comparing the same words and the same task across groups, we see that L1 and L2 speakers process the emotional valence of words similarly. The valence effect in L2 is not modulated by etymological similarities across languages: although, unsurprisingly, cognate status was a significant predictor of reaction times, the valence effect persisted with no difference between early and late learners, suggesting that the effect is genuinely conceptually mediated rather than linked to more surface properties of the words.

The valence effect observed in L2 participants is consistent with results of Ferré et al. (2010) who found an advantage in recalling valenced words in both L1 and L2, regardless of age and context of L2 acquisition, and with recent neuroimaging evidence showing enhanced ERPs components for valenced words in both L1 and L2 (Conrad et al., 2011; Opitz & Degner, 2012).

This is not to say that word processing is the same in the two languages. In fact, we found a number of differences across groups. First, Late L2 responded slower than Native speakers, although no RT difference was observed between Early L2 and Native speakers, compatible with findings from Sandoval, Gollan, Ferreira, and Salmon (2010; see Hanulová, Davidson, & Indefrey, 2011 for review). This is consistent with the idea that adults learn a new language by establishing translation equivalents between L1 and L2, especially in early stages (MacWhinney, 2008); is compatible with ERP findings showing delayed automatic processing in L2 (Opitz & Degner, 2012); and in line with current models of the organization of the bilingual lexicon (e.g., Kroll & Tokowicz, 2005), claiming that the mental lexicon in L2 users becomes more integrated across the L1 and L2 as proficiency develops. These models imply that in highly proficient L2 speakers, a word presented in one language automatically activates, in parallel, the mental representation of the corresponding translation in the other language. This is consistent with the significant effect of cognate status we found, although crucially we still found a valence effect after taking cognates into account, with no differences of AEA. Late L2 learners were also more sensitive to the effect of word frequency, in line with evidence suggesting that the frequency effect might be larger in L2 as compared with L1 (e.g., van Wijnendaele & Brysbaert, 2002).

Surprisingly, however, we did not find any modulation of language immersion on the valence effect, that would be suggested by those same models of L2 acquisition. This is also in contrast with Degner et al. (2012), who found that only participants with a high daily frequency of L2 use showed a significant automatic processing of valence in L2, despite the participants they tested did not differ in terms of proficiency or duration of stay in a L2 speaking country. However, valence effects in L2 have been found by Eilola et al. (2007) in a group of Finnish-English bilinguals living in Finland, who most likely did not use English frequently in everyday life. It remains to be established if differences in terms of design, stimuli and task (affective priming in Degner et al., 2012; emotional Stroop in Eilola et al., 2007; lexical decision in the present study) are responsible for the differences in results. In any case, finding a valence effect in both L1 and L2 speakers regardless of AEA, language similarity, and also degree of language immersion, argue against a 'disembodied' account of L2 processing, that would imply that even highly proficient bilingual speakers can only process their L2 semantically but not emotionally (see Pavlenko, 2012).

In summary, we showed for the first time in a large-scale study including highly proficient L2 speakers with a range of different L1s and different degrees of language immersion that valenced words, regardless of polarity, are processed faster than neutral words and therefore strongly suggest conceptually-mediated processing even for individuals who acquired their L2 late and/or do not use English in everyday life.

Acknowledgements

Supported by UK Economic and Social Research Council grant RES-062-23-2012 to Gabriella Vigliocco. Trial-level data from the experiments reported here will be made available via the UK Data Service upon publication of this work.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*(3), 466–478.

Altarriba, J. (2008). Expressions of emotion as mediated by context. *Bilingualism: Language and Cognition*, *11*(02), 165–167.

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., & Jacobs, A. M. (2012). The power of emotional valence-from cognitive to affective processes in reading. *Frontiers in human neuroscience*, *6*, 192.

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

Aycicegi, A., & Harris, C. (2004). Bilinguals' recall and recognition of emotion words. *Cognition & Emotion*, *18*(7), 977–987.

Ayçiçegi-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, *12*(03), 291.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., & Maechler, M. (2009). Ime4: linear mixed-effects models using S4 classes R package version 0.999375-32.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

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*, 351.

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition*, *15*(01), 181–189.

Dewaele, J.-M. (2004). Perceived language dominance and language preference for emotional speech: The implications for attrition research. In M. Schmid (Ed.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 81–104).

Dewaele, J.-M., & Pavlenko, A. (2002). Emotion Vocabulary in Interlanguage. *Language Learning*, *52*(2), 263–322.

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

Ellis, R. (1997). Second Language Acquitision. In *Oxford Introduction to Linguistics*. Oxford: Oxford University Press.

Ferré, P., García, T., Fraga, I., Sánchez-Casas, R., & Molero, M. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition & Emotion*, *24*(5), 760–785.

Ferré, P., Sánchez-Casas, R., & Fraga, I. (2012). Memory for emotional words in the first and the second language: Effects of the encoding task. *Bilingualism: Language and Cognition*, *16*(03), 495–507.

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*, *26*(7), 902–934.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL - International Review of Applied Linguistics in Language Teaching*, *40*(2), 69–87.

Harris, C.L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 2*, 223-247.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–81.

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.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 531–553). New York: Oxford University Press.

Kuchinke, L., Võ, M. L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *65*(2), 132–40.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A.B. (2014). Emotion and Language: Valence and Arousal Affect Word Recognition. *Journal of Experimental Psychology*. *General*.

Lang, P. J., Bradley, M. M., & Cuthbert, B. (1997). Motivated attention: Affect, activation, and action. In P. Lang, R. Simons, & M. Balaban (Eds.), *Attention and Orienting: Sensory and Motivational Processes* (pp. 97–135). Erlbaum, Hillsdale NJ.

Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical Characteristics of Words Used in Emotional Stroop Experiments. *Emotion*, *6*(1), 62–72.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaires: A web-based interface for bilingual research. *Behavioral Research Methods, 38*, 202-210.

MacWhinney, B. (2008). A unified model. In N. Ellis & P. Robinson (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 341–371). New York/London: Routledge.

Odlin, T. (1989). *Language Transfer: Cross-linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.

Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, *50*(8), 1961–1967.

Pavlenko, A. (2004). "Stop Doing That, Ia Komu Skazala !": Language Choice and Emotions in Parent—Child Communication. *Journal of Multilingual and Multicultural Development*, *25*(2-3), 179–203.

Pavlenko, A. (2005). *Emotions and multilingualism*. (Vol. 1). Cambridge: Cambridge University Press.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, *11*(02), 147–164.

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

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(02), 231.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime Reference Guide. Pittsburgh: Psychology Software Tool, Inc.

Scott, G. G., O'Donnell, P.J., Leuthold, H., & Sereno, S.C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology, 80,* 95-104.

Segalowitz, N., Trofimovich, P., Gatbonton, E., & Sokolovskaya, A. (2008). Feeling affect in a second language: The role of word recognition automaticity. *The Mental Lexicon*, *3*(1), 47–71.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability and familiarity. *Behavior Research Methods, 38*, 598–605.

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 & Emotion*, *21*(5), 1077–1090.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (in press). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*.

van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 616–627.

Vinson, D., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition & Emotion, 28*(4), 737-746.

Figure 1 - Mean trimmed correct reaction times of Native English speakers, Early L2 participants and Late L2 participants as a function of Valence. Error bars represent standard error of the mean by subject.

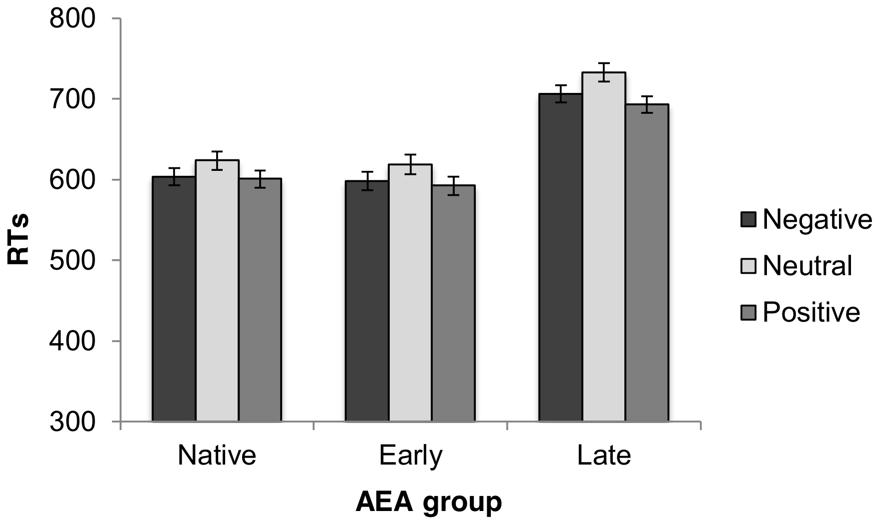


Figure 2 - Partial effects of the continuous predictors on trimmed correct RTs (msec) in the final linear mixed model, including main effects of the control variables as well as the main effects of valence, word frequency and AEA group, and the frequency × AEA group interaction. Upper-left: familiarity; upper-right: imageability; lower-left: concreteness; lower-right: age of acquisition.

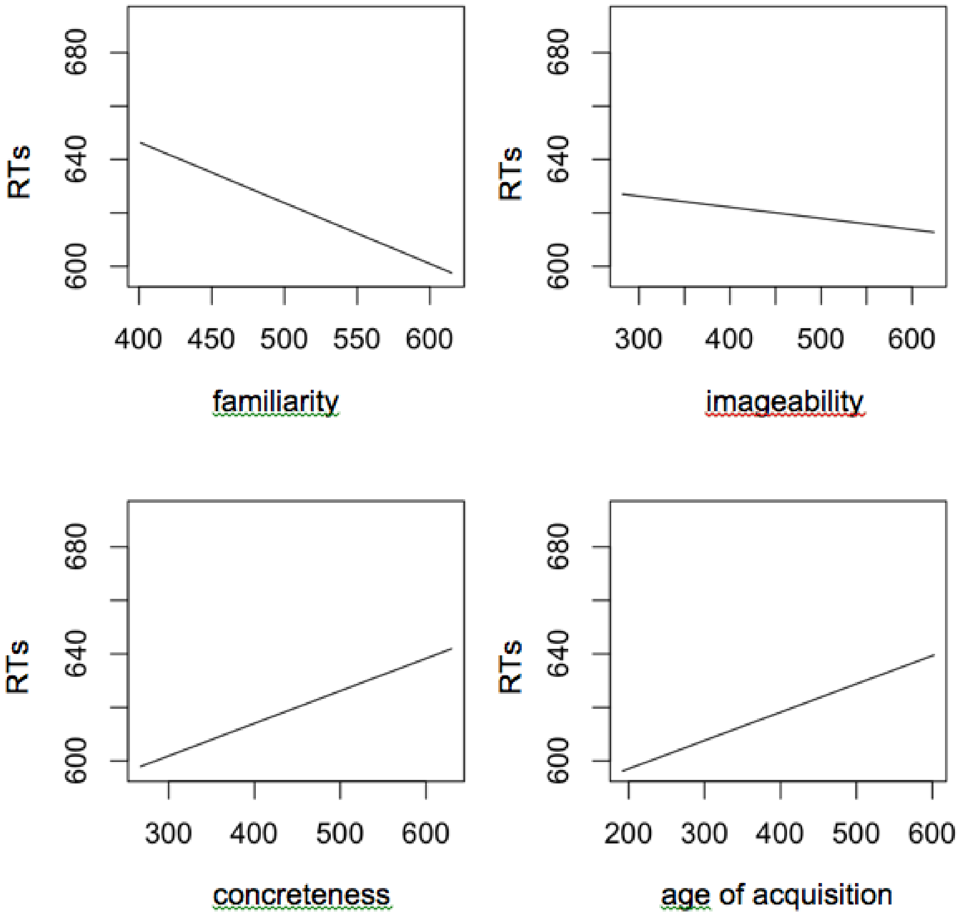


Figure 3 - Mean trimmed correct reaction times (msec) of Late L2 residents in an English speaking country and residents in a Non-English speaking country as a function of Valence. Error bars represent 95% upper confidence bound based on fixed-effects uncertainty and random-effects variance.

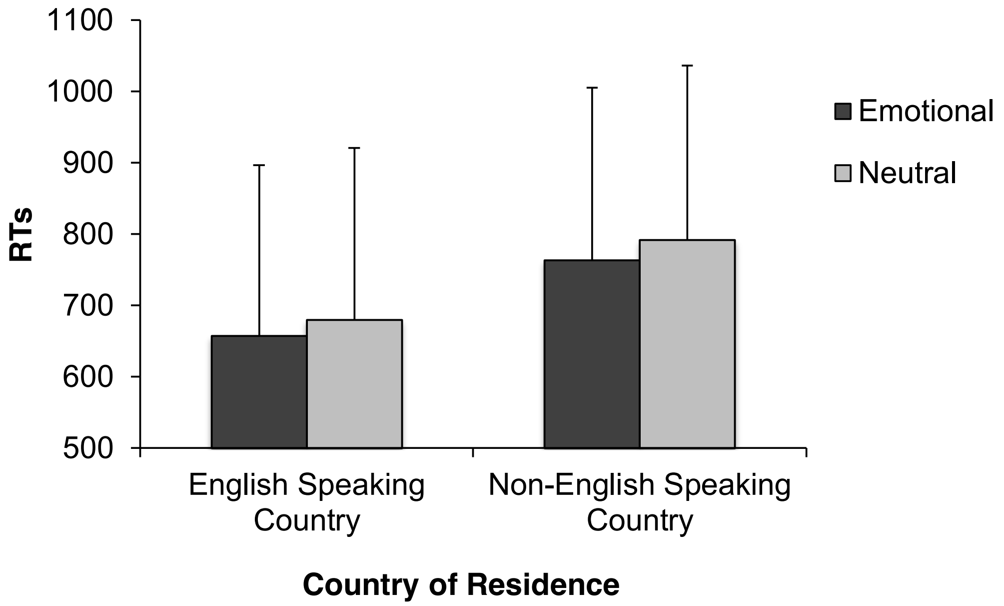


Figure 1

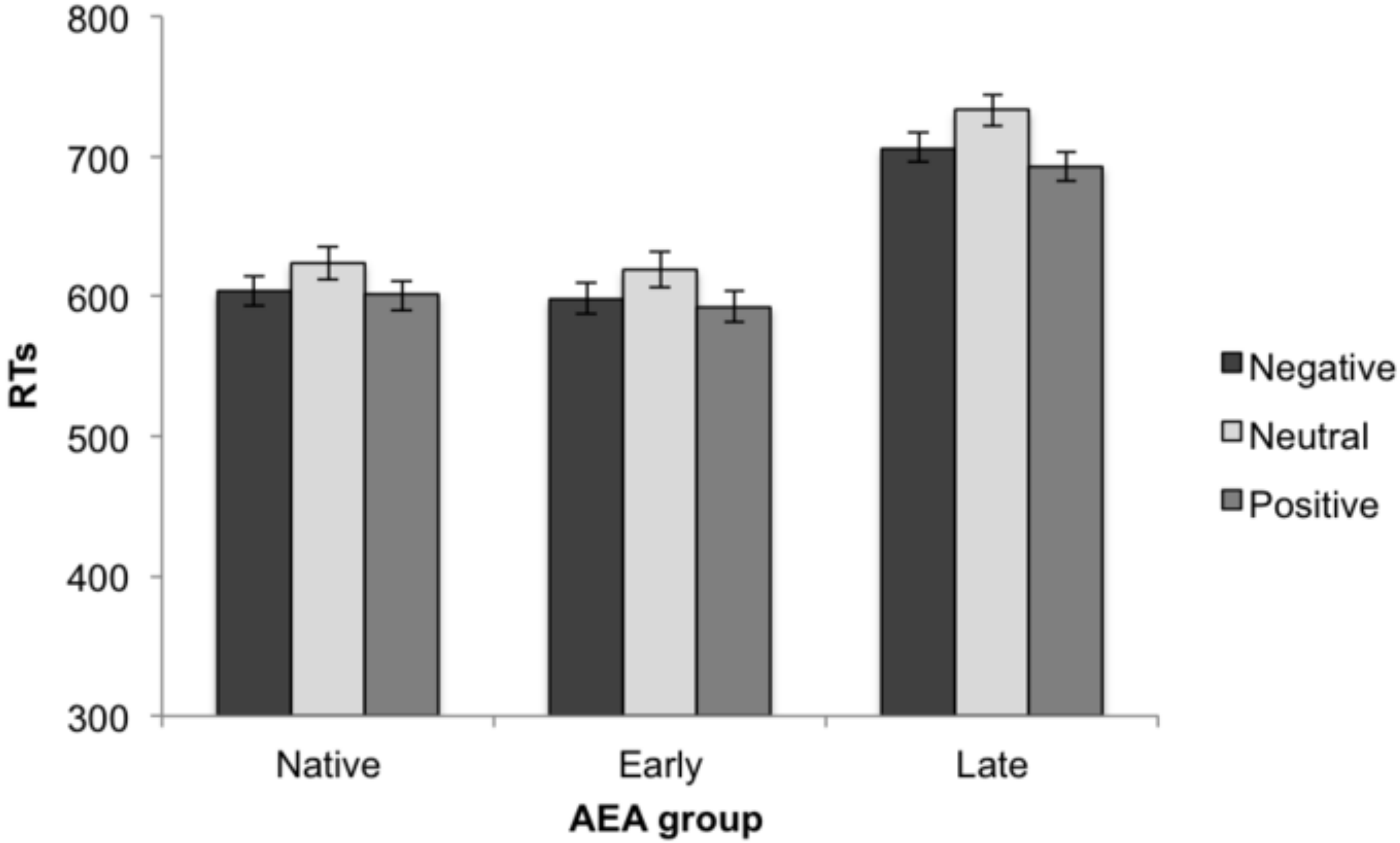

800
700
600
500
400
300
RTs
Native
Early
Late
AEA group
Negative
Neutral
Positive

Figure 2

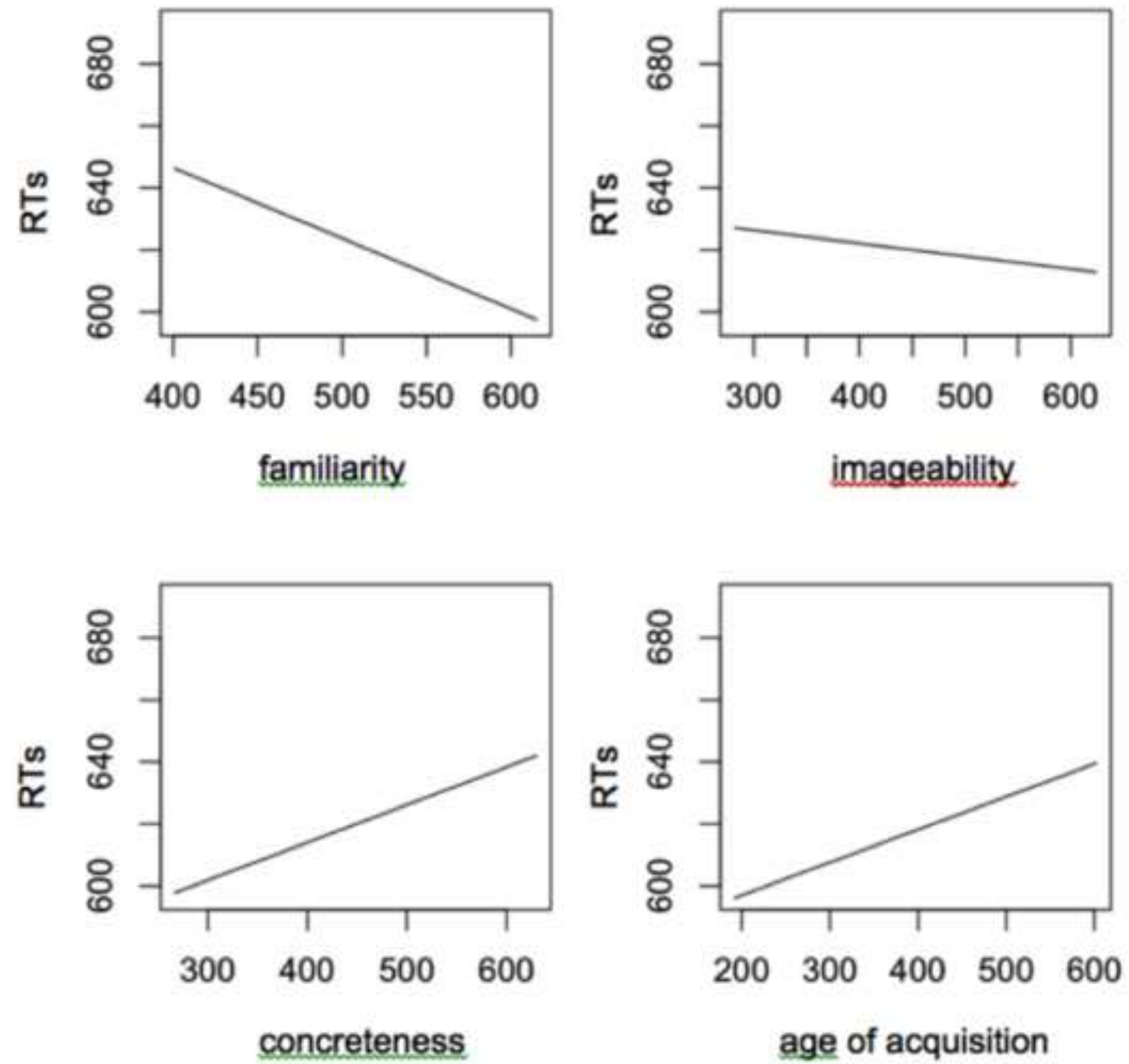
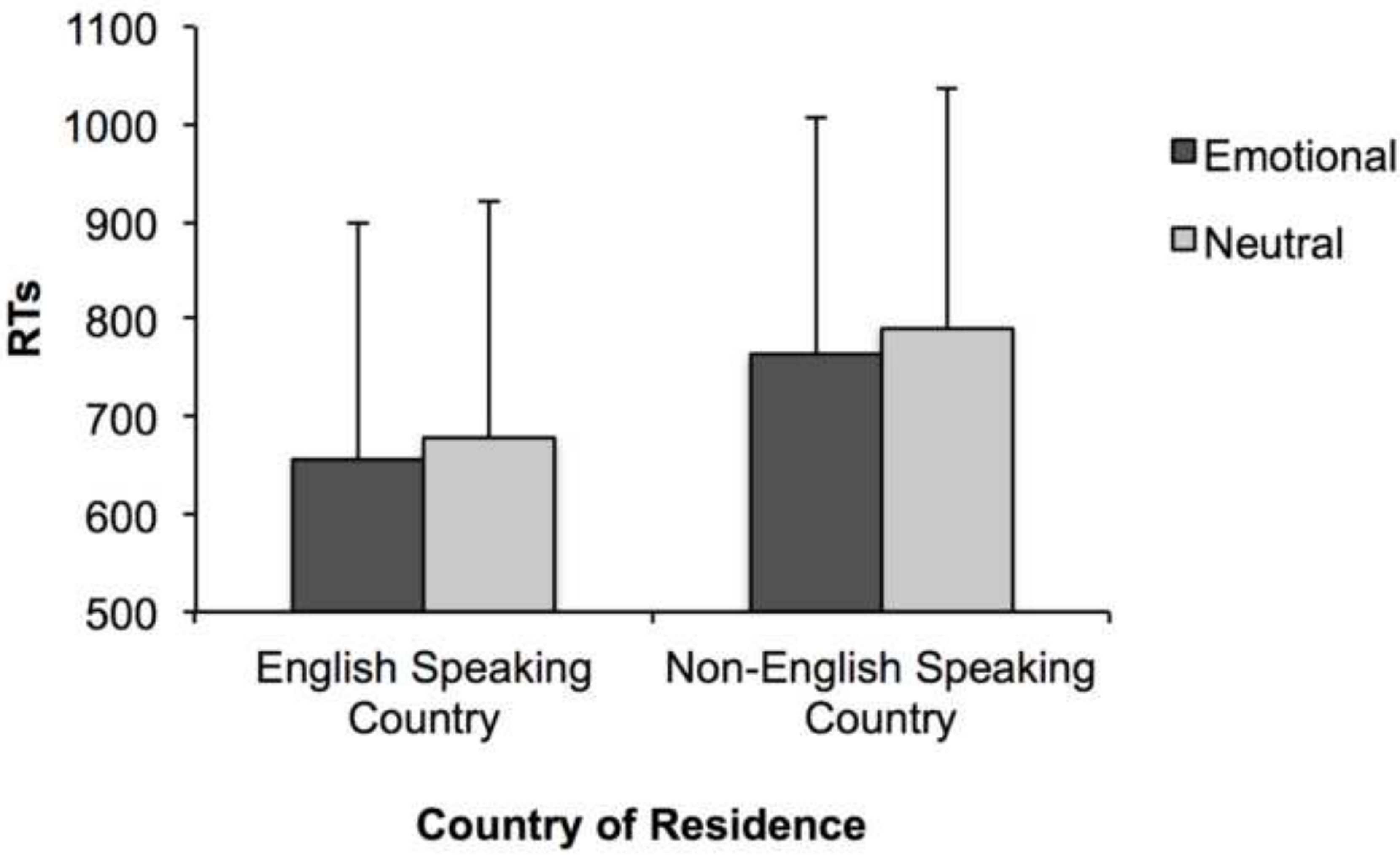

RTs
600
640
680
400 450 500 550 600
familiarity
RTs
600
640
680
300 400 500 600
imageability
RTs
600
640
680
300 400 500 600
concreteness
RTs
600
640
680
200 300 400 500 600
age of acquisition

Figure 3



RTs
1100
1000
900
800
700
600
500
English Speaking Country
Non-English Speaking Country
Country of Residence
Emotional
Neutral

Supplemental Table 1
Click here to download Supplemental Material - Integral: emotion processing in L2_SP1.docx