

ORIGINAL ARTICLE

The processing advantage of emotion words in L1 speakers and L2 speakers: A meta-analysis

Yanlu Zhong , Lu Liu and Laurel Brehm

University of California Santa Barbara, USA

Corresponding author: Yanlu Zhong; Email: yanlu_zhong@ucsb.edu

(Received 26 April 2025; revised 12 January 2026; accepted 2 April 2026)

Abstract

This meta-analysis synthesized 88 studies to investigate the processing advantages of emotion words over neutral words. Additionally, we explored the moderating effects of emotional properties (valence, arousal, emotion word type), linguistic factors (concreteness, frequency, length, neighborhood size), and task type in L1, advanced L2, and intermediate L2 speakers. We found a significant valence effect, with positive words showing a greater processing advantage than negative words only in L1 speakers. For arousal, the interaction analysis revealed that high arousal reduced the processing advantage of emotion words to a greater extent in intermediate L2 speakers than in L1 speakers. Furthermore, only advanced L2 speakers showed a significant processing advantage for emotion-label words compared to emotion-laden words. Regarding linguistic factors, longer word length was associated with greater processing advantages compared to shorter word length, but only in advanced L2 speakers. The greater processing advantage for concrete over abstract emotion words was observed only in intermediate L2 speakers, indicating that this group was the most sensitive to concreteness among all language speaker groups. Finally, task type significantly influenced emotion word processing in interaction with language proficiency. Overall, our findings support theoretical frameworks in both L1 and L2 processing and cognition.

Keywords: Emotion; first language; meta-analysis; second language; word processing

Introduction

Words are the building blocks of language, and the brain is sensitive to diverse information encoded in individual words (Ellis & Ogden, 2017), including statistical (Brysbaert et al., 2018), phonological (Yates, 2005), semantic (Kissler et al., 2006), and orthographic cues (Jalbert et al., 2011) in first language (L1) processing. A key aspect of semantics is emotional content. Recent studies have examined the processing advantage of emotion words over neutral words, focusing on two affective dimensions: valence (positivity or negativity) and arousal (emotional

intensity) (Barrett & Russell, 1998; Citron *et al.*, 2014). In psycholinguistic research, debate remains over whether valence exerts similar (Balota *et al.*, 2007) or distinct effects (Estes & Adelman, 2008; Kousta *et al.*, 2009) on the word processing of L1 speakers. Findings are also mixed on whether processing is modulated by arousal (Kazanas & Altarriba, 2016), linguistic features (e.g., word length) (Larsen *et al.*, 2006), or task type (Liu *et al.*, 2023). For second language (L2) speakers, it is unclear whether their processing advantage is reduced (Iacozza *et al.*, 2017; Toivo & Scheepers, 2019) or comparable to L1 speakers (Ayçiçeği-Dinn & Caldwell-Harris, 2009). Moreover, how psycholinguistic factors affect emotion word processing in L2 speakers at different proficiency levels, and how these patterns diverge from L1 processing, remains underexplored.

To bridge these gaps, this meta-analysis compared the processing advantages of emotion words over neutral words in L1, intermediate L2, and advanced L2 speakers. We further examine whether language processing involves valence-sensitive mechanisms, as proposed by the Automatic Vigilance Theory (Pratto & John, 1991), or instead reflects general motivational engagement with both positive and negative stimuli, as suggested by the Model of Motivated Attention and Affective States (Lang *et al.*, 1997). The moderating effects of arousal, word type, linguistic variables (frequency, word length, phonological and orthographic neighborhood sizes), and task type were also investigated. The findings will shed light on how embodied experience shapes word representation across varying levels of language proficiency.

Literature review

Emotion word processing for L1 and L2 speakers

Emotion appears to be automatically activated during L1 processing (Kissler *et al.*, 2007; Kousta *et al.*, 2009), but the mechanisms and extent of such activation in L2 remain debated. Behavioral studies using emotional priming (Kazanas & Altarriba, 2016), word recall (Altarriba & Basnight-Brown, 2011), emotional Stroop (Eilola *et al.*, 2007), and lexical decision tasks (Conrad *et al.*, 2011; Ferré *et al.*, 2018) suggest that L2 speakers also automatically activate emotional information. For instance, Ayçiçeği-Dinn and Caldwell-Harris (2009) found comparable emotional activation in L1 and L2 speakers in a recall task. However, other studies have yielded contrasting findings. In a Stroop paradigm, Winskel (2013) observed significant emotion effects on word processing in L1 but not L2 speakers. Using eye-tracking, Tang and Ding (2024) found that, unlike L1 readers (e.g., Knickerbocker *et al.*, 2015), L2 readers showed no processing advantage for emotion words during sentence reading. Physiological data further revealed stronger pupil dilation in L1 than in L2 speakers during the processing of emotional information, consistently observed across both word recognition (Toivo & Scheepers, 2019) and sentence reading tasks (Iacozza *et al.*, 2017).

These divergent findings may be explained by the Emotional Contexts of Learning Theory (Caldwell-Harris *et al.*, 2006), which posits that L2 emotion word processing is shaped by learning context. Limited emotional engagement in L2 learning leads to weaker embodiment and reduced emotional activation compared

to L1 (Pavlenko, 2012), with activation further moderated by variation in language use and acquisition environment (Ahn & Jiang, 2023; Tang & Ding, 2024). Importantly, L1 and L2 speakers also differ in the cognitive resources available for emotional language processing. Evidence from the Ghent Eye-Tracking Corpus (Cop et al., 2017), where participants read entire novels in both L1 and L2, showed that processing emotionally rich narratives in L2 imposes greater cognitive demands than in L1. This aligns with multiple theoretical frameworks. According to the Shallow Structure Hypothesis (Clahsen & Felser, 2006), L2 speakers process linguistic information in a shallower and less detailed way due to limited working memory resources. Likewise, the Lexical Quality Hypothesis (Perfetti, 2007) suggests that L2 speakers' less robust and integrated lexical representations increase cognitive load and hinder activation of emotional, semantic, and syntactic information. Therefore, compared to L1 speakers, fewer cognitive resources are left for emotional information for L2 speakers and smaller emotion effects are expected in processing (Imbault et al., 2021). Finally, L1 and L2 speakers may differ in emotional representation. The Revised Hierarchical Model (Kroll & Stewart, 1994) posits that L2 speakers access concepts via L1 translation rather than direct lexical-conceptual links. This limits their ability to effectively activate emotional content during processing.

Emotional factors that may moderate the processing advantages of emotion words

Valence

Emotion words such as joy, sadness, and surprise are categorized by valence as positive or negative. Valence plays a central role in emotion word processing (Blackett & Harnish, 2022), as explained by two existing models.

The Automatic Vigilance Theory (Pratto & John, 1991) proposes that negative stimuli receive sustained attentional processing because of their survival relevance, resulting in humans' slower responses to negative relative to positive or neutral stimuli. Valence thus exerts an independent moderating effect beyond other psycholinguistic factors. Supporting this theory, studies with L1 speakers have shown processing disadvantages for negative words across Stroop (Winkel, 2013), lexical decision (Estes & Verges, 2008), and naming tasks (Algorn et al., 2004). A large-scale study by Kuperman et al. (2014) found a monotonic effect of valence on lexical decision times, with more negative words eliciting slower responses. Similar patterns have been observed in L2 speakers: negative words often show processing disadvantages in behavioral (Ferré et al., 2018), eye-tracking (Sheikh & Titone, 2016), and neuroimaging studies (Jończyk, 2016). However, vigilance effects may be weaker in L2 due to reduced emotional grounding or embodied experience (Ferré et al., 2018; Pavlenko, 2012). For instance, Jończyk (2016) reported delayed and attenuated neural responses to negative words in L2 compared to L1 speakers.

In contrast, the Model of Motivated Attention and Affective States (Lang et al., 1997) argues that attention is drawn to motivationally relevant stimuli regardless of valence, granting both positive and negative words processing advantages over neutral ones. Knickerbocker et al. (2015) found that both types facilitated L1 reading. The model attributes prior differences between positive and negative word

processing to uncontrolled psycholinguistic variables. For example, Vinson (2014), using the British Lexicon Project, showed that after controlling for variables like concreteness, frequency, and neighborhood size, emotion words—positive or negative—were processed faster than neutral ones. Similarly, Larsen *et al.* (2006) and Kousta *et al.* (2009) reported that the processing disadvantages of negative words disappeared when controlling for word length, frequency, and orthographic neighborhoods.

Arousal

Arousal refers to the intensity of emotional activation. The dimensional model posits that arousal is an independent emotional dimension, distinct from valence, and may engage separate cognitive mechanisms in language processing (Barrett & Russell, 1999). Norming studies have identified a U-shaped relationship between valence and arousal across modalities, including affective pictures (Kurdi *et al.*, 2017), single words (Stadthagen-Gonzalez *et al.*, 2017), and extended verbal stimuli like idioms and sentences (Citron *et al.*, 2016; Morid & Sabourin, 2024; Zhong *et al.*, 2025). This pattern suggests that stimuli with stronger emotional valence—whether positive or negative—are typically associated with higher arousal levels than neutral stimuli. Previous studies have also reported mixed findings regarding the arousal effects on behavioral performance and neural representation in language processing among L1 speakers (Estes & Adelman, 2008; Kousta *et al.*, 2009; Kuperman *et al.*, 2014; Larsen *et al.*, 2008). Estes and Adelman (2008), for instance, found that arousal positively predicted English word recognition speed, with high-arousal words recognized faster than calming ones. In contrast, Kuperman *et al.* (2014) found that high-arousal words were recognized more slowly than calming ones, though the effect was minimal (0.1% variance) in a dataset of 12,658 English words. Similarly, Kousta *et al.* (2009), analyzing 1,446 words, reported no significant arousal effect on reaction times when valence was controlled. Emotional arousal also affects L1 and L2 processing differently. Altarriba and Canary (2004) examined the arousal effects of emotion words in both English L1 speakers and Spanish L2 speakers of English using a lexical decision task. The results showed positive priming effects for both groups under arousal conditions compared to a baseline condition, but L2 speakers exhibited longer latencies compared to L1 speakers.

Emotion word type

Emotion words fall into two types: emotion-label words (e.g., *happy*, *fearful*), which directly denote emotional states, and emotion-laden words (e.g., *coffin*, *murder*), which imply emotions without explicit labeling (Pavlenko, 2008). The Emotion Duality Model (Imbir *et al.*, 2019; Tang & Ding, 2024) posits that emotion-label words trigger automatic, biologically rooted responses, while emotion-laden words require mediated conceptual access and greater cognitive effort (Altarriba & Basnight-Brown, 2011). Previous studies have reported varying findings on the processing of emotion-label and emotion-laden words in L1 speakers. Zhang *et al.* (2020) revealed that emotion-label and emotion-laden words activated different cortical responses and demonstrated neural dissociation,

while Martin and Altarriba (2017) and Vinson et al. (2014) found no significant word type effects. In L2 processing, Kazanas and Altarriba (2016) reported a larger priming effect for emotion-label than emotion-laden words in their L1, but not in L2, among Spanish-English bilinguals in a lexical decision task. However, in emotion categorization tasks (Tang et al., 2023), L2 speakers exhibited higher accuracy and shorter reaction times for emotion-label words compared to emotion-laden words. Additionally, Altarriba and Basnight-Brown (2011) found that L1 and L2 speakers differed in word type-valence interactions in an Affective Simon Task. L1 speakers showed valence-color congruency effects only for negative emotional-label words, while L2 speakers exhibited expected effects for both word types, regardless of valence.

Linguistic factors that may moderate the processing advantages of emotion words

Linguistic factors such as concreteness, frequency, length, and orthographic or phonological neighborhood size are known to influence general word processing (Ferré et al., 2018; Kuperman et al., 2014). These linguistic factors also moderate emotion word processing (e.g., Kousta et al., 2009; Larsen et al., 2006; Vinson et al., 2014), as reviewed below.

Concreteness

Concrete and abstract words differ in emotional connotation. According to Vigliocco et al.'s (2009) Embodied Theory, concrete words arise from external sensory-motor experiences, while abstract words relate to internal states. Norming studies have shown that more abstract words are generally rated as more positive (Hinojosa et al., 2016; Kousta et al., 2011) and more arousing (Guasch et al., 2016; Vigliocco et al., 2014). For L1 word processing, some studies reported that concrete emotion words were processed more quickly and accurately in recall and lexical decision tasks (Ferré et al., 2018) and elicited earlier ERP components (Palazova et al., 2013) compared to abstract ones. However, when attention focuses on emotional content, abstract words may show a processing advantage over concrete words. Yao and Wang (2014) found that abstract positive words elicited faster responses in emotional priming tasks. Similarly, Jin et al. (2023) reported that abstract words more effectively captured attention and enhanced emotional evaluation than concrete words in an emotion recognition task, as reflected in ERP data. Studies have also reported that concreteness differentially modulates emotion-related effects in L1 and L2 speakers. For instance, valence effects were found to be stronger for abstract words than for concrete words in L1 speakers (Palazova et al., 2013), but stronger for concrete words in L2 speakers (Ferré et al., 2018). This contrast may reflect disembodied cognition (Pavlenko, 2012) and acquisition order (Jones & Mewhort, 2007). L2 speakers tend to have less emotional grounding than L1 speakers. As a result, concrete words, acquired earlier and tied to sensory experience, may evoke stronger emotional responses than later-learned, abstract words in L2 speakers.

Frequency

Language acquisition is shaped by repeated exposure to language, and language speakers are sensitive to the frequency of linguistic units (Ellis & Ogden, 2017). High-frequency words are processed more efficiently (Balota et al., 2004; Ellis, 2002), which was also observed with emotion words (Larsen et al., 2006). Frequency also moderates the effects of other psycholinguistic factors such as valence and arousal (Kuperman et al., 2014). From a cognitive perspective, low-frequency words may allow deeper processing, thereby enabling psycholinguistic variables to exert their effects over the course of processing. In contrast, some studies reported stronger valence effects for high-frequency words, particularly in lexical decision and eye-tracking tasks (e.g., Scott et al., 2009).

Length

Few studies have directly examined word length effects on emotion word processing. Nevertheless, some norming studies have explored the link between word length and emotionality. Larsen et al. (2006), for instance, found that emotion words tend to be longer. Beyond emotion words, length effects are well-documented, with longer words generally processed more slowly than shorter ones (e.g., Baddeley et al., 1975). Yet, this relationship may not be strictly linear. Using English Lexicon Project data (Balota et al., 2002), New et al. (2006) identified a U-shaped effect: 5–8-letter words yielded faster responses compared to 3–4 letter or 9–13 letter words. Similar findings were reported in Balota et al. (2007). Yap and Balota (2009) later attributed this U-shaped trend to the optimal perceptual span for medium-length words (6–9 letters).

Orthographic and phonological neighborhood sizes

Language processing is influenced by the activation of similar lexical representations, measured by orthographic and phonological neighborhood sizes. Orthographic neighborhood size refers to the number of words formed by changing one letter while maintaining letter positions (Coltheart et al., 1977). Larger orthographic neighborhoods facilitate visual word recognition (Forster & Shen, 1996) and recall (Jalbert et al., 2011). A meta-analysis of 32 studies by Larsen et al. (2006) found that emotion words had significantly smaller orthographic neighborhoods than control words. In word processing, larger orthographic neighborhoods are generally associated with faster response latencies in behavioral tasks (Forster & Shen, 1996), possibly due to stronger semantic connectivity in memory (Larsen et al., 2006). Phonological neighborhood size—the number of words differing by one phoneme (Luce & Pisoni, 1998)—also affects language processing. Yates (2005) found a processing advantage for words with larger phonological neighborhoods across naming, lexical decision, and categorization tasks, whereas other studies reported inhibitory effects of larger phonological neighborhoods in lexical decision and naming tasks (Luce & Pisoni, 1998).

The role of task type in the processing advantages of emotion words

Over the past decades, various tasks—including Stroop color-naming, lexical decision (Estes & Verges, 2008), word naming (Algom et al., 2004), and eye-tracking

during reading—have been used to investigate emotion word processing. These tasks fall into three categories: (1) Emotional tasks (e.g., emotion judgment task) direct attention to emotional content, requiring participants to categorize words by valence (Liu et al., 2023); (2) Irrelevant tasks (e.g., Stroop color-naming task) focus on non-emotional, non-linguistic features like color, ignoring emotional or linguistic content (Algom et al., 2004; Larsen et al., 2008); (3) Linguistic tasks (e.g., lexical decision and semantic judgment tasks) focus on semantic, grammatical, or orthographic properties (Yao et al., 2016). The rationale for using irrelevant and linguistic tasks is that if emotional information is automatically activated, differences in processing between emotion and neutral words would be observed.

Some studies have examined how task type affects emotion word processing in L1 and L2 speakers. Winskel (2013) compared Stroop (irrelevant) and emotion judgment (emotional) tasks in Thai-English bilinguals. Emotional effects appeared in both tasks for L1 (Thai), but only in the emotion judgment task for L2 (English). Winskel explained that the Stroop task taps early, automatic processing, whereas the emotion judgment task involves later, conscious processing. Unlike L1 speakers, L2 speakers may not automatically activate emotional content but can do so at later processing stages. Task type also showed significant interactions with specific emotional variables. Estes and Verges (2008) reported an interaction between valence and task type in L1 speakers: while approach effects for positive words were consistent across tasks, freezing responses to negative words appeared only when emotion was not task-relevant, such as in irrelevant or linguistic tasks. Ferré et al. (2018) further examined this interaction in both the L1 and L2 of language speakers. In L1, negative words showed a processing disadvantage relative to neutral words in emotion judgment and lexical decision tasks, but not in free recall tasks. By contrast, positive words showed a processing advantage over neutral words in lexical decision and free recall tasks, but not in emotion judgment tasks. In L2, negative word effects appeared in both lexical decision and free recall tasks, while positive word effects emerged only in lexical decision tasks.

The current study

To synthesize divergent findings, we conducted a meta-analysis on the processing advantages of emotion words. We examined how emotional, linguistic factors and task type influence emotion word processing¹ across studies involving L1 and L2 speakers. Given differences in language acquisition, we further assessed whether L1, advanced, and intermediate L2 speakers² differ in sensitivity to these variables. Three research questions (RQs) guided this investigation:

- RQ₁: To what extent do emotion words enjoy a processing advantage over neutral words?
- RQ₂: To what extent do emotional factors (i.e., valence, arousal, emotion word type), linguistic factors (i.e., concreteness, word frequency, word length, orthographic and phonological neighborhood sizes), and task type moderate the processing advantage of emotion words?
- RQ₃: To what extent are the main effects of the above-mentioned psycholinguistic factors on the processing advantage of emotion words moderated by language proficiency?

Methodology

Inclusion and exclusion criteria

We searched Google Scholar, Web of Science, PsycINFO, ERIC, Linguistics and Language Behavior Abstracts, and ProQuest Global Dissertations to identify studies on emotion word processing. Keywords including “emotions”, “emotion-label”, “emotion-laden”, “negative words”, “positive words”, “neutral words,” and “word processing” were used in the searches for studies published before March 1, 2025. This yielded 3,436 reports. A backward citation search (Yanagisawa & Webb, 2021) identified 241 additional reports, resulting in 3,677 studies screened using the six inclusion criteria below.

1. We only included studies written in English and excluded those written in other languages.
2. We included studies examining the emotion word processing advantage across all languages represented in the dataset (e.g., English, Chinese, German).
3. This meta-analysis focused on the processing advantage of emotion words over neutral words. Therefore, we only included studies that involved neutral words as control groups.
4. The included studies fell into four speaker group types: (a) L1-only, (b) L2-only, (c) mixed L1 and L2 speakers of the same language, and (d) bilingual individuals with both an L1 and an L2.³
5. We used response times or reaction times (both referred to as RTs hereafter) as the measures of effect sizes. Studies without reported RTs were excluded.
6. We included studies that investigated the processing of emotion words in adults and excluded those exclusively focused on children.
7. We only included data from healthy, neurotypical participants.
8. We included studies reporting sufficient data (e.g., RTs, standard deviations, or t-values) to calculate effect sizes. Studies lacking such data were excluded.

We examined the abstracts and introductions of potentially relevant studies, identifying 307 potential studies. Figure 1 presents the screening process (see Appendix S1 for more details about the PRISMA flow diagram). Through a full-text review, we then narrowed these down to 88 studies with 3,280 participants that met all our criteria, reporting 391 effect sizes. These studies comprised 85 journal papers, 2 doctoral theses, and 1 master’s thesis (see Appendix S2 for included studies for this meta-analysis).

Coding

For all included studies, we coded outcome variables (e.g., descriptive statistics or metrics for effect size calculation), predictor variables (e.g., valence, arousal, frequency, emotion word type, length, concreteness, orthographic and phonological neighborhood sizes, task type, language proficiency, language type, language background), and study identifiers (author, year). All predictors except neighborhood size and frequency were treated as categorical, using scale midpoints when needed. If predictor values were unreported but full stimuli were available ($N = 43$),

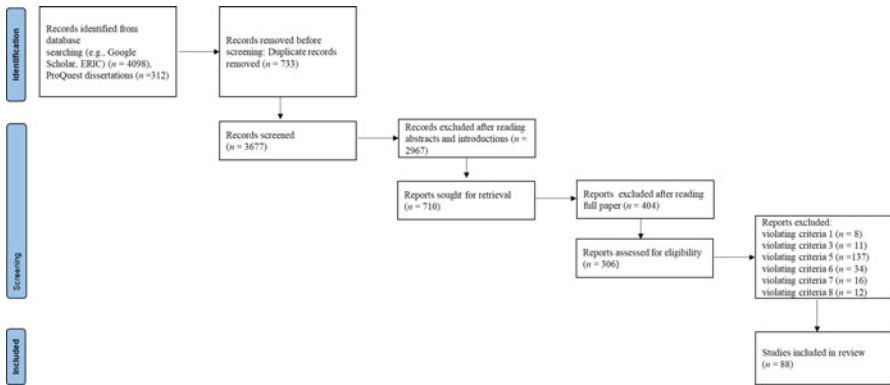


Figure 1. PRISMA flow diagram for meta-analysis.

we retrieved values from reference databases; otherwise, values were coded as missing (see Appendix S3 for details of the coding scheme in Table 1 and Appendix S4 for the coding sheet).

Effect size calculation

When studies reported complete RTs and standard deviations, Cohen's d was computed as the difference between emotion word (positive or negative) and neutral word RTs, divided by the pooled standard deviation. For 12 studies reporting t -values or correlations, we converted them directly into Cohen's d . To correct small-sample bias, we applied a correction factor J to obtain Hedges' g (see Appendix S6 for details).

Coding procedure

Two researchers with expertise in psycholinguistics and statistics collaboratively refined the coding scheme. One conducted the initial coding, and the other independently verified it based on the finalized scheme. Discrepancies were resolved through discussion. Inter-rater reliability was high (Cohen's $\kappa = 0.998$). See Appendix S7 for details.

Analysis procedure

Analyses were conducted in R (v4.3.1; R Core Team, 2023) using the *metafor* (v3.0.2; Viechtbauer, 2010) and *clubSandwich* (v0.5.10; Pustejovsky, 2023) packages. To address effect size dependencies and account for language type and learner variables, we employed a five-level multilevel meta-regression model (e.g., Yanagisawa & Webb, 2021), with random effects for sampling variance (Level 1), within-study variance (Level 2), between-study variance (Level 3), language background (Level 4), and language type (Level 5). Cluster-robust variance estimation (Hedges et al., 2010), with small-sample corrections (Tipton & Pustejovsky, 2015), was used to account for dependency among effect sizes and ensure valid inference under unbalanced conditions. The alpha level was set at .05.⁴

Table 1. Coding scheme for predictor variables

Predictor	Code	Reference
Language proficiency ^a	L1 speakers: 0; advanced L2 speakers: 1; intermediate L2 speakers: 2	Data reported in the paper
Valence ^b	Positive words:0; negative words:1	Data reported in the paper
Emotional predictors		
Arousal ^b	Values Lower than the midpoint of the scale (low arousal): 0 Values higher than the midpoint of the scale (high arousal): 1	1. Data reported in the paper 2. National Research Council Canada (NRC) Valence, Arousal, and Dominance (VAD) Lexicon (Mohammad, 2018)
Word type	Emotion-laden words: 0; emotion-label words: 1, mixed: 2	1. Data reported in the paper 2. The stimuli in the appendix
Linguistic predictors		
Concreteness ^b	Values higher than the midpoint of the scale (concrete): 0 Values lower than the midpoint of the scale (abstract): 1	1. Data reported in the paper 2. Norm studies for concreteness (e.g., Brysbaert et al., 2014)
Frequency	Continuous variable	1. Subtlex database 2. Data reported in the paper
Length ^c	Short (3.92–6.23 letters):0; medium (6.24–7.00):1; long (7.01–10.30):2	1. Data reported in the paper 2. Calculating based on the stimuli in the appendix
Phonological neighborhood size	Continuous variable	CLEARPOND database
Orthographic neighborhood size	Continuous variable	CLEARPOND database
Task type		
Task type ^d	Emotional tasks (tasks focusing on emotional information) (e.g., emotion judgment task): 0 Irrelevant tasks (tasks focusing on non-linguistic and non-emotional information) (e.g., Emotion Stroop task): 1 Linguistic tasks (tasks focusing on linguistic information) (e.g., lexical decision tasks, phrase judgment tasks): 2	Data reported in the paper
Random effects		
Language type	English:0, German:1, Chinese:2, Spanish:3, Polish:4, Catalan: 5, Finnish:6, Korean:7, Hungarian:8, Italian:9, Turkish:10, French:11, Uyghur:12, mixed: 13	Data reported in the paper

(Continued)

Table 1. (Continued)

Predictor	Code	Reference
Language background ^e	English:0, German:1, Chinese:2, Catalan:3, Spanish:4, French: 5, Finnish:6, Korean:7, Polish:8, Italian:9, Hungarian:10, Greek:11, Cantonese:12, Tai:13, Malaysian:14, Turkish:15, Uyghur:16, mixed:17	Data reported in the paper

Note:

^aEach unique participant group and language spoken was treated as a separate entry. Language proficiency was coded based on self-rated proficiency, length of language learning, and age of acquisition. Participants who acquired the language from birth and reported near-ceiling proficiency or were identified as native speakers were coded as L1. Those who rated themselves in the top third and began learning before age 10 ^[4] (Hartshorne et al., 2018) were classified as advanced L2 speakers, and the rest were labeled as intermediate speakers. This yielded 23 advanced and 29 intermediate L2 speaker groups—a roughly balanced distribution that facilitated subsequent analyses. Of the 88 included studies, 71 involved only L1 speakers, 2 only advanced L2, and 6 only intermediate L2. Additionally, 5 studies included both L1 and advanced L2 speakers, and 4 included both L1 and intermediate L2 speakers. Language proficiency was coded categorically rather than continuously due to the absence of consistent, standardized, and psychometrically valid test scores across studies. Self-report scales also varied (e.g., 10-point in Chen et al., 2015 vs. 7-point in Naranowicz et al., 2023), limiting comparability. Categorical coding further allowed detection of non-linear effects that linear models might obscure. In addition, we used multilevel meta-regression with robust variance estimation to assess minimum detectable effects (MDE) under the unbalanced L2–L1 data structure. Results showed sufficient sensitivity to detect small effects for advanced L2–L1 comparisons (MDE = 0.13) and small-to-moderate effects for intermediate L2–L1 comparisons (MDE = 0.23), indicating that group imbalance did not preclude meaningful analyses. To further mitigate bias from sample imbalance, we adopted several strategies: (1) cluster-robust variance estimation with small-sample corrections, which was used to ensure valid inference under unbalanced conditions; (2) sensitivity analyses excluding outliers, unpublished studies, or collapsing L2 groups; (3) disaggregating studies with multiple contrasts to increase usable data for L2 groups; and (4) applying a five-level meta-regression model to account for effect size variance, within- and between-study dependence, and language-level heterogeneity. This approach ensured accurate modeling and reduced the influence of group size imbalance on statistical inference.

^bValence, arousal, and concreteness were treated as categorical variables for several reasons. First, cross-linguistic variation in semantic interpretation complicates fine-grained numerical comparisons. Second, rating scales varied across studies—for valence and arousal (e.g., 9-point in Sheikh & Titone, 2013; 7-point in Chen et al., 2015; 5-point in Pauligk et al., 2019) and concreteness (e.g., binary in Spadacenta et al., 2014; 5-point in Kaltwasser et al., 2013; 7-point in Ferré et al., 2017; 9-point in Chen et al., 2015). Third, the Model of Motivated Attention and Affective States (Lang et al., 1997) predicts non-linear valence effects that are better captured through categorical rather than continuous modeling. Finally, numerical concreteness ratings were often missing (in 76.1% of studies), and categorical coding enabled broader inclusion and theory-driven contrasts.

^cLength was measured in letters. To ensure consistency, we excluded languages with non-Latin orthographies (e.g., Chinese and Korean), which rely on logographic characters or syllabic block units rather than linear sequences of Latin alphabetic characters. Given the U-shaped patterns of length effects, where words of medium length have the shortest RTs as found in previous studies (Balota et al., 2007; New et al., 2006), we categorized word length into short, medium, and long bins based on tertiles.

^dSee Appendix S5 for the descriptions and coding of each type of task.

^eLanguage background refers to the language speakers' first language.

To examine the processing advantage of emotion words, as indexed by effect sizes (RQ₁), we computed weighted average effect sizes by multiplying each estimate by the inverse of its variance, summing the weighted values, and dividing by the total weight (See Appendix S6 for variance calculations and Appendix S8 for the forest plot of effect sizes and variances). To explore the moderating effects of emotional, linguistic, and task-related factors on effect sizes (RQ₂), each factor was entered individually into separate models as a predictor.⁵ To assess whether these effects vary by language proficiency (RQ₃), we added language proficiency and its interactions with each of the above-mentioned psycholinguistic variables as model predictors.⁶

To assess potential publication bias, we generated a funnel plot, conducted Egger's sandwich test, and applied fail-safe N , Orwin's fail-safe N , and the trim-and-fill method (Borenstein *et al.*, 2009). All results indicated no publication bias (see Appendix S9). We also ran multiple sensitivity analyses, confirming the robustness of the identified effects (see Appendix S10).

Results

No significant overall processing advantage was found for emotion words (combined negative and positive) compared to neutral words when analyzing data pooled across all language groups (Hedge's $g = 0.030$, 95% CI = $[-0.097, 0.157]$, $P = .612$). Only L1 speakers showed a significant processing advantage of positive words ($b = 0.198$, $P = .008$), while negative words did not differ significantly from neutral words in any language speaker group. L1 speakers also showed a significant processing advantage for positive over negative words ($b = -0.179$, $P = .005$). A significant interaction with proficiency ($P = .021$) further confirmed that this valence effect was stronger in L1 than in intermediate L2 speakers, in which no significant effects were observed (see Table 2).

A significant moderating effect of language proficiency (intermediate L2 vs. L1) on arousal was observed ($P = .036$), such that high arousal reduced the processing speed advantage of emotion words measured by RTs to a greater extent in intermediate L2 speakers than in L1 speakers. A significant effect of word type was observed only in advanced L2 speakers ($b = 0.450$, $P = .036$), indicating that emotion-label words were associated with a larger processing advantage compared to emotion-laden words. Interaction analysis with proficiency ($P = .041$) revealed that this word type effect was significantly stronger in advanced L2 speakers than in L1 speakers, who showed no significant effect.

When considering linguistic factors (see Table 3), a significantly greater processing advantage for longer words compared to shorter words (medium vs. short: $b = 0.148$, $P = .015$; long vs. short: $b = 0.325$, $P = .011$) was found only among advanced L2 speakers. We also observed a significant interaction between language proficiency and word length (long vs. short: $P = .048$), indicating that the length effects were more pronounced in advanced L2 speakers than in L1 speakers, who showed no significant word length effect. In addition, a significant processing advantage for concrete emotion words compared to abstract emotion words was found only in intermediate L2 speakers ($b = 0.370$, $P = .030$). This concreteness effect was stronger than that observed in advanced L2 speakers ($P = .049$) and L1 speakers ($P = .042$), neither of whom showed significant concreteness effects.

Significant task effects were observed only in intermediate L2 speakers, with smaller processing speed advantages measured by RTs for emotion words in both the irrelevant task ($b = -0.725$, $P = .007$) and the linguistic task ($b = -0.588$, $P = .007$), compared to the emotional task. Interaction analyses revealed that task effects (linguistic task vs. emotional task: $P = .024$) were greater in intermediate L2 speakers than in L1 speakers, in which no significant task effects were found (see Table 3).

Table 2. Results of the moderator analyses including emotional factors as predictors

Valence			Arousal			Word type		
Variable	<i>b</i> [95% CI]	<i>P</i>	Variable	<i>b</i> [95% CI]	<i>P</i>	Variable	<i>b</i> [95% CI]	<i>P</i>
intercept	0.177 [0.061, 0.293]	0.008	intercept	0.032 [−0.098, 0.162]	0.568	intercept	0.050 [−0.107, 0.207]	0.457
valence	−0.179 [−0.255, −0.103]	0.005	arousal	0.107 [−0.075, 0.289]	0.116	word_type	−0.031 [−0.374, 0.313]	0.656
language proficiency (advanced L2)	−0.121 [−0.679, 0.437]	0.320	language proficiency (advanced L2)	−0.118 [−0.511, 0.275]	0.279	language proficiency (advanced L2)	−0.147 [−0.510, 0.217]	0.146
language proficiency (intermediate L2)	0.037 [−0.387, 0.461]	0.664	language proficiency (intermediate L2)	0.369 [0.112, 0.625]	0.024	language proficiency (intermediate L2)	0.100 [−0.118, 0.317]	0.189
valence: language proficiency (advanced L2)	0.116 [−0.214, 0.446]	0.156	^a arousal: language proficiency (advanced L2)	–	–	word_type: language proficiency (advanced L2)	0.480 [0.080, 0.880]	0.041
valence: language proficiency (intermediate L2)	0.186 [0.070, 0.302]	0.021	arousal: language proficiency (intermediate L2)	−0.267 [−0.475, −0.060]	0.036	word_type: language proficiency (intermediate L2)	0.054 [−0.586, 0.695]	0.631
<i>R</i> -square: Total: 0.156; Between effect sizes:0.891; Between studies: 0.060; Between Language Background: 0.000; Between Language Type: 1.000			<i>R</i> -square: Total: 0.134; Between effect sizes:1.000; Between studies: 0.353; Between Language Background: 0.000; Between Language Type: 1.000			<i>R</i> -square: Total: 0.375; Between effect sizes:1.000; Between studies: 0.200; Between Language Background: 0.000; Between Language Type: 1.000		

Note: Valence (reference group): positive. Arousal (reference): low arousal. Word_type (reference group): emotion-laden word. Language proficiency (reference group): L1 speaker

^aRedundant predictors were dropped from the model due to the absence of high-arousal cases among advanced L2 speakers.

Table 3. Results of meta-regressions including valence, language proficiency, and linguistic factors or task type as predictors

Variable	Regression	<i>b</i> [95% CI]	<i>P</i>	Variable	Regression	<i>b</i> [95% CI]	<i>P</i>
Concreteness	intercept	0.085 [−0.246, 0.292]	0.136	Frequency	intercept	0.091 [−0.059, 0.241]	0.193
	concreteness	−0.062 [−0.476, 0.352]	0.602		frequency	0.0001 [−0.001, 0.001]	0.721
	language proficiency (advanced L2)	−0.031 [−0.573, 0.512]	0.862		language proficiency (advanced L2)	−0.022 [−0.427, 0.383]	0.871
	language proficiency (intermediate L2)	−0.023 [−0.255, 0.210]	0.701		language proficiency (intermediate L2)	−0.004 [−0.463, 0.456]	0.977
	concreteness: language proficiency (advanced L2)	−0.051 [−0.622, 0.519]	0.702		frequency: language proficiency (advanced L2)	−0.001 [−0.003, 0.001]	0.205
	concreteness: language proficiency (intermediate L2)	0.432 [0.006, 0.858]	0.049		frequency: language proficiency (intermediate L2)	0.004[−0.027, 0.035]	0.367
<i>R</i> -square Total: 0.259; Between effect sizes:1.000; Between studies: 0.320; Between Language Background:0.000; Between Language Type: 1.000				<i>R</i> -square Total: 0.055; Between effect sizes:0.910; Between studies: 0.166; Between Language Background:0.000; Between Language Type: 1.000			
Orthographic Neighborhood Sizes (OTAN)	intercept	0.175 [−0.437, 0.786]	0.410	Phonological Neighborhood Sizes (PTAN)	intercept	0.174 [−0.446, 0.795]	0.396
	OTAN	−0.019 [−0.168, 0.130]	0.598		PTAN	−0.016 [−0.06, 0.028]	0.230
	language proficiency (advanced L2)	−0.107 [−0.718, 0.504]	0.598		language proficiency (advanced L2)	−0.156 [−0.776, 0.464]	0.441
	language proficiency (intermediate L2)	−0.148 [−0.759, 0.463]	0.476		language proficiency (intermediate L2)	0.228 [−0.392, 0.848]	0.292
	OTAN: language proficiency (advanced L2)	0.007 [−0.142, 0.156]	0.836		PTAN: language proficiency (advanced L2)	0.013 [−0.031, 0.057]	0.305
	OTAN: language proficiency (intermediate L2)	0.087 [−0.062, 0.236]	0.118		PTAN: language proficiency (intermediate L2)	0.006 [−0.038, 0.05]	0.606

(Continued)

Table 3. (Continued)

Variable	Regression	<i>b</i> [95% CI]	<i>P</i>	Variable	Regression	<i>b</i> [95% CI]	<i>P</i>
<i>R</i> -square Total: 0.000; Between effect sizes: 1.000; Between studies:0.000; Between Language Background:0.000; Between Language Type: 1.000				<i>R</i> -square Total: 0.000; Between effect sizes: 1.000; Between studies: 0.350; Between Language Background: 0.000; Between Language Type: 1.000			
Length	intercept	0.011 [−0.152, 0.174]	0.861	Task type	intercept	0.203 [−0.179, 0.586]	0.222
	length (medium)	0.066 [−0.134, 0.265]	0.237		task (irrelevant)	−0.250 [−0.613, 0.112]	0.144
	length (long)	0.083 [−0.112, 0.277]	0.232		task (linguistic)	−0.119 [−0.490, 0.252]	0.422
	language proficiency (advanced L2)	−0.164 [−0.703, 0.375]	0.255		language proficiency (advanced L2)	0.155 [−0.207, 0.517]	0.315
	language proficiency (intermediate L2)	0.165 [−0.354, 0.684]	0.226		language proficiency (intermediate L2)	0.525[0.152, 0.898]	0.016
	length (medium): language proficiency (advanced L2)	0.082 [−0.124, 0.289]	0.185		task (irrelevant): language proficiency (advanced L2)	−0.295 [−1.121, 0.532]	0.244
	length (long): language proficiency (advanced L2)	0.242 [0.004, 0.480]	0.048		task (linguistic): language proficiency (advanced L2)	−0.177 [−1.112, 0.758]	0.412
	length (medium): language proficiency (intermediate L2)	0.053 [−0.257, 0.363]	0.376		task (irrelevant): language proficiency (intermediate L2)	−0.474 [−1.089, 0.141]	0.082
	length (long): language proficiency (intermediate L2)	−0.117 [−0.992, 0.758]	0.536		task (linguistic): language proficiency (intermediate L2)	−0.468 [−0.836, −0.101]	0.024
<i>R</i> -square Total: 0.070; Between effect sizes:0.153; Between studies: 0.010; Between Language Background: 0.000; Between Language Type: 1.000				<i>R</i> -square Total: 0.259; Between effect sizes: 0.000; Between studies:0.176; Between Language Background:0.000; Between Language Type: 1.000			

Note: Concreteness (reference: low concreteness); length (reference: short); task (reference: emotional task); valence (reference: positive words); language proficiency (reference: L1 speakers).

Discussion

No overall processing advantage was found for emotion words (combined positive and negative) over neutral words across language groups, consistent with mixed findings in L1 and L2 research. Some studies report advantages for emotion words regardless of valence (e.g., Kousta *et al.*, 2009; Ferré *et al.*, 2018), while others suggest that emotional content—especially negative—may impair processing (e.g., Palazova *et al.*, 2013). These inconsistencies may reflect the influence of various psycholinguistic moderators, as we discuss below.

Valence effects emerged only in L1 speakers, who showed a significant processing advantage for positive over negative words, suggesting distinct cognitive mechanisms for different valences. In contrast, among L2 speakers, emotion word processing was significantly moderated by arousal, emotion word type, word length, frequency, and task type, indicating a reliance on integrated encoding of emotional, linguistic, and cognitive factors, while such effects were largely absent in L1 speakers. These differences may stem from emotional grounding and lexical representation quality (Ellis, 2002). First, according to the Emotional Contexts of Learning Theory (Caldwell-Harris *et al.*, 2006), L1 speakers acquire linguistic forms, sensory-motor experiences, and emotional content concurrently in emotionally rich childhood environments (Pavlenko, 2012), enabling automatic emotional activation. In contrast, L2 is often learned through explicit instruction in emotionally neutral contexts, offering fewer opportunities to link word meanings with autobiographical experiences (Tenderini *et al.*, 2022; Jończyk, 2016), resulting in weaker emotional grounding—particularly for intermediate learners. Second, emotional representations in the L2 mental lexicon may be less robust (the Lexical Quality Hypothesis, Perfetti, 2007), and emotional language processing may be shallower and more structurally constrained (the Shallow Structure Hypothesis, Clahsen & Felser, 2006). Due to weaker emotional grounding and representation, emotional activation in L2 is less automatic and demands greater cognitive effort, reducing sensitivity to emotional content compared to L1 speakers (Caldwell-Harris *et al.*, 2006; Pavlenko, 2012).

Valence has significant effects that vary between L1 and L2 speakers

The processing advantage of positive over negative words was significant among L1 speakers. This valence effect aligns with prior findings (e.g., Algom *et al.*, 2004; Estes & Verges, 2008; Kazanas & Altarriba, 2016; Kuperman *et al.*, 2014) and supports the Automatic Vigilance Theory (Pratto & John, 1991), which posits that negative emotions trigger a “freeze” response in language processing (Kousta *et al.*, 2009). However, the lack of a significant disadvantage for negative versus neutral words suggests the findings do not fully contradict the Stimulated Attention and the Emotion State Model (Lang *et al.*, 1997). Rather, both models may coexist: while negative words can also enhance processing relative to neutral words, similar to positive words, avoidance mechanisms may offset this advantage. In either case, the results indicate that, at least for L1 speakers, language processing is deeply intertwined with emotional content (Sheikh & Titone, 2016).

However, no significant valence effects were found in either advanced or intermediate L2 speakers. Interaction analysis confirmed that valence effects were significantly stronger in L1 than in intermediate L2 speakers, who showed no effect. Our findings align with prior evidence of reduced emotional responses in L2 speakers compared to L1 speakers, as seen in pupil dilation (Toivo & Scheepers, 2019), eye-tracking (Tang & Ding, 2024), and skin conductance (Eilola & Havelka, 2011). This highlights potential heterogeneity in emotion word processing between L1 and L2 speakers. The absence of valence effects in L2 speakers may stem from reduced emotional grounding (Caldwell-Harris et al., 2006), lower-quality emotional representations (Perfetti, 2007), and shallower syntactic processing (Clahsen & Felser, 2006), all of which may limit cognitive resources for emotional processing. Notably, although advanced L2 speakers showed no significant valence effects, the interaction analysis revealed no significant group differences between advanced L2 and L1 speakers. This suggests that with increased proficiency, cognitive resources, and language experience, L2 speakers' sensitivity to emotional valence gradually aligns with that of L1 speakers.

L2 speakers showed stronger arousal effects and word type sensitivity

The reduction in processing speed for highly arousing emotion words, as measured by RTs, was found only in intermediate L2 speakers. A significant interaction between arousal and language proficiency confirmed that the inhibitory effect of high arousal was stronger in intermediate L2 speakers than in L1 speakers, who showed no significant effect. While studies on L1 speakers have reported mixed arousal effects—some showing facilitation (e.g., Kousta et al., 2009), others reporting null or slight inhibitory effects (e.g., Kuperman et al., 2014)—our results suggest that intermediate L2 speakers consistently exhibit a qualitative difference in their processing of high-arousal stimuli compared to L1 speakers, aligning with prior findings (Altarriba & Canary, 2004). The stronger inhibitory effect of high arousal in intermediate L2 speakers may result from less emotionally grounded access to emotion word meanings and shallower emotional representations. First, L1 speakers typically acquire language in emotionally rich contexts, forming strong links between words and emotional experiences (Caldwell-Harris et al., 2006). In contrast, L2 speakers—particularly at the intermediate level—often learn in emotionally neutral settings, limiting such associations. Therefore, they exhibit disembodied emotion processing, with weaker emotional engagement and reduced sensorimotor grounding compared to L1 speakers (Dylman & Bjärtå, 2019; Kühne & Gianelli, 2019; Pavlenko, 2008). Second, intermediate L2 speakers may have difficulty forming precise, automatized lexical representations (Perfetti, 2007) and tend to develop shallower emotional representations (Clahsen & Felser, 2006). Reduced automatic access to emotion word meanings and less deeply emotional grounding increase the cognitive demands of processing high-arousal information, thereby amplifying its inhibitory effects on word processing in less proficient L2 speakers.

Regarding word type, a processing advantage for emotion-label over emotion-laden words was found only in advanced L2 speakers. They also showed greater sensitivity to word type than L1 speakers, who showed no significant effect. This

pattern, consistent with prior research (Altarriba & Basnight-Brown, 2011; Kazanas & Altarriba, 2015), suggests that representational differences between the two word types are most pronounced in advanced L2 speakers. L1 speakers process both types efficiently due to fully developed emotion-semantic mappings, while intermediate L2 speakers show generally reduced effects due to weaker access to emotion meanings. In contrast, advanced L2 speakers—whose emotional language processing is partially developed—are sufficient for effective engagement with emotion-label words but still limited for the more context-dependent emotion-laden words. To be more specific, L1 speakers' reduced sensitivity to word type, relative to L2 speakers, may reflect differences in emotional representation (the Emotion Duality Model; Imbir *et al.*, 2019; Tang & Ding, 2024) and distinct lexical structures in L1 and L2 systems (the Revised Hierarchical Model; Kroll & Stewart, 1994). Emotion-label words denote discrete emotional states, while emotion-laden words convey affect through contextual associations. For L1 speakers, emotionally grounded experience supports efficient extraction of affective content from both types. In contrast, L2 speakers, lacking embodied experience, find emotion-laden words—whose meanings depend on real-world associations—less accessible. They process these with reduced automaticity and greater reliance on context, leading to stronger differentiation between word types. Notably, the word type effect was absent in intermediate L2 speakers, possibly due to limited emotional language experience. This may hinder even the processing of emotion-label words, diminishing their advantage over emotion-laden words and weakening the overall word type effect.

Taken together, our findings suggest that distinct emotional dimensions are independently represented in word processing. Their interaction with language proficiency indicates that L2 speakers, especially intermediate ones, process emotional language differently from L1 speakers. In general, L2 speakers' emotional language processing is constrained by cognitive resources and language experience.

Multifaceted moderating effects of linguistic factors

A processing advantage for concrete over abstract emotion words was found only in intermediate L2 speakers. This concreteness effect was also stronger in intermediate L2 speakers than in both advanced L2 and L1 speakers, where no significant concreteness effect was found. These results suggest a hierarchical pattern in the acquisition of emotional lexical information, with emotional content in an L2 developing from concrete to abstract concepts. Concrete words are more easily learned, as they connect to L1-based experiences and tangible referents. In contrast, abstract words require higher-order conceptual representations and are typically acquired later, as language proficiency moves from intermediate to advanced and eventually native-like levels. Thus, emotion word processing in intermediate L2 speakers depends more heavily on concreteness than in the other language speaker groups.

Furthermore, word length mattered, but only among advanced L2 speakers, who showed greater advantages for longer emotion words. These words may be more salient in their language experience, leading to facilitation. The facilitatory effect of longer word length was more pronounced in advanced L2 speakers than in L1 speakers, who did not exhibit a significant word length effect. This may reflect L2

speakers' greater reliance on surface-level linguistic features like word length (the Shallow Structure Hypothesis; Clahsen & Felser, 2006), rather than deeper emotional or pragmatic content, compared to L1 speakers (the Lexical Quality Hypothesis; Perfetti, 2007). Interestingly, positive length effects were absent in intermediate L2 speakers. This may be because intermediate L2 speakers are additionally constrained by limited cognitive resources compared to advanced L2 speakers. With richer language experience, advanced L2 speakers have more stable lexical representations of longer words, allowing them to process surface-level orthographic, phonological, and grammatical features with less effort. In contrast, intermediate L2 speakers, with less stable representations, may face greater cognitive load and resource depletion as word length increases (Liu et al., 2021).

In contrast, no frequency effects were found for the processing advantage of emotion words. This is surprising, given well-established frequency effects in word processing (Brysbaert et al., 2018), including for emotion words (Larsen et al., 2006). However, unlike prior studies, our focus was not on baseline word processing but on the processing advantage of emotion words over neutral words. Hence, our results suggest that frequency did not interact with the general effects of emotional information. One explanation is that over 90% of the emotion words in the meta-analysis were low-frequency (i.e., below 100 per million) and narrowly distributed, limiting the detectability of frequency effects. Another possibility is that emotional activation tied to word frequency is absent or unstable. As emotional processing is embodied, emotional activation requires real-world experiential grounding. High linguistic frequency alone—especially from decontextualized contexts like classrooms—may not strengthen emotional associations, and thus may not enhance processing advantages for emotion words (Pavlenko, 2012; Jończyk, 2016).

Neither orthographic nor phonological neighborhood size significantly influenced word processing in any group. Although prior studies reported neighborhood effects on word processing (e.g., Forster & Shen, 1996; Jalbert et al., 2011; Luce & Pisoni, 1998; Yates, 2005), our meta-analysis did not replicate these findings. As discussed above, this discrepancy likely reflects our distinct focus on emotional processing advantages (i.e., emotion vs. neutral words) rather than baseline word processing. Thus, our results suggest that the richness of orthographic or phonological representations is not closely intertwined with emotional activation.

Task type showed moderating effects that differed between L1 and L2 speakers

Greater processing speed advantages measured by RTs for emotion words in emotional tasks, compared to both linguistic and irrelevant tasks, were observed only in intermediate L2 speakers. Moreover, the interaction revealed that these task effects were more pronounced in intermediate L2 speakers than in L1 speakers, where no significant effects were observed. With limited language experience, intermediate L2 speakers may struggle to activate emotional content when it is not the focus. Similarly, Winkler (2013) found that L2 speakers do not automatically evoke emotions during early processing stages but can consciously activate them later. Specifically, emotional tasks tend to elicit stronger emotional responses by directing attention to emotional content and facilitating its retrieval.

By contrast, linguistic and non-emotional tasks offer fewer cues for emotional processing, resulting in more implicit activation. Both the nature of L2 learning environments and the structure of emotional representation in the bilingual lexicon may explain intermediate L2 speakers' differential performance across task types. First, intermediate L2 learners often acquire language through explicit instruction lacking emotionally rich, real-world experiences (the Emotional Contexts of Learning Theory; Caldwell-Harris *et al.*, 2006), limiting their ability to automatically activate emotional content in tasks without an emotional focus. Second, learning and representation are likely interconnected: for intermediate L2 speakers, linguistic symbols and embodied emotional experiences may be processed through separate channels (the Revised Hierarchical Model; Kroll & Stewart, 1994), such that emotional activation is less efficient when it is not explicitly required by the task in intermediate L2 speakers. This may further hinder automatic access to emotional meaning in non-emotional tasks, unlike more proficient L2 or L1 speakers.

Furthermore, the absence of task effects in advanced L2 speakers, compared to their presence in intermediate L2 speakers, highlights the role of language experience in the implicit activation of emotional content. With increased language proficiency, L2 speakers become better at implicitly processing emotional information and suppressing interference. Consequently, task-related differences in emotion word processing diminish as language proficiency increases.

Limitations and future directions

As in most language processing research (Collart, 2024), studies on emotion word processing have predominantly focused on English and German. Future work should include a broader range of languages to explore cross-linguistic and cross-cultural variation. Second, the structural imbalance in the current data—marked by fewer L2 studies—indicates the need for more research on L2 speakers to enhance group balance and generalizability. Third, the L2 speakers in the included studies came from varied learning environments—some with extensive immersion in the target language (e.g., Conrad *et al.*, 2011), others living in bilingual contexts (Cieśllicka & Guerrero, 2023). According to the Emotional Contexts of Learning Theory (Caldwell-Harris *et al.*, 2006), cultural and social differences critically influence emotion word processing. Future research should report and control for L2 learning environments to better understand emotion word processing in L2 speakers. Furthermore, the absence of consistent, standardized, and fine-grained proficiency measures across studies limited the modeling of L2 proficiency as a continuous variable. Future research should adopt uniform, validated assessments (e.g., IELTS or TOEFL) to improve comparability and allow for more flexible modeling. Fifth, inconsistent norming scales and cross-linguistic variation in emotional semantics hindered continuous modeling of valence, arousal, and concreteness. Future studies should use consistent rating scales and examine cross-linguistic differences to support fine-grained modeling.

Finally and most importantly, most emotion word processing research has largely overlooked psycholinguistic features such as imageability, due to (1) limited

data availability (none of the L2 studies included in this meta-analysis reported imageability ratings, and large-scale norms remain scarce), and (2) high inter-individual (Su et al., 2023) and cross-linguistic (Rofes et al., 2018) variability in ratings, which undermines effect robustness and interpretability. Likewise, AoA was excluded due to its absence in L1 studies, limited reporting in L2 research and because this meta-analysis focuses on lexical properties rather than proxies of cumulative learning history. Future research should improve standardization and reporting of key psycholinguistic variables across languages, reduce variability and include features reflecting developmental trajectories.

Conclusion

This meta-analysis investigated the processing advantage of emotion words. The results suggest that, like most cognitive activities, language processing is dynamically influenced by valence and largely aligns with the principles of the Automatic Vigilance Theory (Pratto & John, 1991), which posits preferential attention to affectively positive stimuli. Critically, our findings reveal that emotion word processing is modulated by multiple dimensions—including arousal, lexical category (e.g., emotion-label vs. emotion-laden words), linguistic features (e.g., word length, frequency), and task type—all of which interact with the cognitive mechanisms underlying affective language comprehension. Last, the meta-analysis revealed significant heterogeneity across L1, advanced L2, and intermediate L2 speakers regarding the moderating effects of emotional polarity (valence), intensity (arousal), the way emotions are represented in words (word type), length, concreteness, and language speakers' attention (task type) on emotion word processing. This suggests that the processing of emotional information in language is embodied and shaped by language experiences and cognitive resources. The findings are consistent with several theoretical frameworks, including the Emotional Contexts of Learning Theory (Caldwell-Harris et al., 2006), the Shallow Structure Hypothesis (Clahsen & Felser, 2006), the Revised Hierarchical Model (Kroll & Stewart, 1994), and the Lexical Quality Hypothesis (Perfetti, 2007). Our study supports the practice of considering L1 speakers and L2 speakers as heterogeneous groups in language processing in empirical research. The comparison between intermediate and advanced L2 speakers further suggests that emotional information acquisition in L2 is a dynamic process that progressively approximates native-like levels. Therefore, emotional information should be regarded as unique and additional content in language learning. Future language teaching and research for L2 speakers should place greater emphasis on the acquisition of emotional meaning.

Supplementary material. The supplementary material for this article can be found <https://doi.org/10.1017/S0142716426100630>.

Replication package. The data that support the findings of this study are openly available in Open Science Framework at <https://osf.io/hteg4/>.

Competing interests. The authors declare none.

Notes

1 Given the heterogeneity between negative and positive words as suggested by the Automatic Vigilance Theory (Pratto & John, 1991), we examined not only the main effect of valence but also whether other psycholinguistic moderators showed different patterns or processing mechanisms for negative versus positive words. However, our analysis revealed no significant interaction effects between valence and other psycholinguistic variables. Moreover, model fit (R^2) improved after removing valence as a moderator. Therefore, we did not include interactions between valence and other psycholinguistic variables in subsequent analyses.

2 We also combined the data from intermediate and advanced L2 speakers and compared L1 speakers with all L2 speakers (see Appendix S11 for details). The results largely replicated those from separate analyses of advanced and intermediate L2 speakers, except that the concreteness and word type effect did not show a clear difference between L1 and combined L2 speakers. This suggests that the amount of L2 language experience influences the mental representation of both emotional and linguistic information.

3 The meta-analysis integrates all four types of studies, allowing for the examination of both within-study and between-study differences in emotion word processing across L1, advanced L2, and intermediate L2 speakers. The method of coding language proficiency was consistent across the four types of studies. In the meta-analytic dataset, the unit of analysis was not the study as a whole, but each unique combination of language and speaker group. Specifically, each language $\times$ proficiency group pairing was treated as an independent data point. For studies involving multiple languages and participants with varied L1 and/or L2 backgrounds, data were disaggregated such that each language–proficiency pairing constituted a separate entry in the dataset. For example, a study examining both German and English emotion words in a group of English L1 speakers and a group of German L1 speakers would yield four data points: English L1 speakers, German L2 speakers, German L1 speakers, and English L2 speakers. This coding strategy allowed for a fine-grained analysis of the effects of language proficiency and language context on emotion word processing.

4 As suggested by some of the previous research, we also tested using age 12 as the cutoff for the critical period (Long, 1990), which did not affect the classification results.

5 We initially planned to include all variables simultaneously to construct a multivariable meta-regression model. However, due to the varying missing values across different variables, and the substantial missing data for some variables (such as word type), including all variables resulted in very little usable data for regression analysis. Therefore, following previous meta-analyses (e.g., Yanagisawa & Webb, 2021; Kim & Webb, 2022), we included only one key variable at a time, along with language proficiency and valence.

6 In preliminary exploratory analyses, we also examined meta-regression models excluding interactions with language proficiency to assess the main effects of each variable independently. Please see Appendix S12 for full details.

References

- Ahn, S., & Jiang, N. (2023). Can adult speakers' sense L2 emotion words automatically? The role of L2 use on the emotional Stroop effect. *Second Language Research*, 39(4), 1265–1278. <https://doi.org/10.1177/02676583221131256>
- Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, 133(3), 323. <https://doi.org/10.1037/0096-3445.133.3.323>
- Altarriba, J., & Basnight-Brown, D. M. (2011). The representation of emotion vs. emotion-laden words in English and Spanish in the Affective Simon Task. *International Journal of Bilingualism*, 15(3), 310–328. <https://doi.org/10.1177/13670069103792>
- Altarriba, J., & Canary, T. M. (2004). The influence of emotional arousal on affective priming in monolingual and bilingual speakers. *Journal of Multilingual and Multicultural Development*, 25(2–3), 248–265. <https://doi.org/10.1080/01434630408666531>
- Ayçiçeği-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, 12(3), 291–303. <https://doi.org/10.1017/S1366728909990125>
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, 14(6), 575–589. [https://doi.org/10.1016/S0022-5371\(75\)80045-4](https://doi.org/10.1016/S0022-5371(75)80045-4)

- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, 46(1), 199–226. <https://doi.org/10.1006/jmla.2001.2803>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133(2), 283. <https://doi.org/10.1037/0096-3445.133.2.283>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . , & Treiman, R. (2007). The English Lexicon Project. *Behavioral Research Methods*, 39, 445–459.
- Barrett, F. L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, 74(4), 967. <https://doi.org/10.1037/0022-3514.74.4.967>
- Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current directions in psychological science*, 8(1), 10–14. <https://doi.org/10.1111/1467-8721.00003>
- Blackett, D. S., & Harnish, S. M. (2022). A scoping review on the effects of emotional stimuli on language processing in people with aphasia. *Journal of Speech, Language, and Hearing Research*, 65(11), 4327–4345. https://doi.org/10.1044/2022_JSLHR-22-00104
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 46, 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Caldwell-Harris, C. L., Gleason, J. B., & Aycicegi, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. *Bilingual education and bilingualism*, 56, 257. <https://doi.org/10.21832/9781853598746-012>
- Cieśllicka, A. B., & Guerrero, B. L. (2023). Emotion word processing in immersed Spanish-English/English-Spanish bilinguals: An ERP study. *Languages*, 8(1), 42. <https://doi.org/10.3390/languages8010042>
- Citron, F. M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A. M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior Research Methods*, 48(1), 91–111.
- Citron, F. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, 56, 79–89. <https://doi.org/10.1016/j.neuropsychologia.2014.01.002>
- Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27(1), 107–126. <https://doi.org/10.1017/S0142716406060206>
- Collart, A. (2024). A decade of language processing research: Which place for linguistic diversity? *Glossa Psycholinguistics*, 3(1). <https://doi.org/10.5070/G60111432>
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornick (Ed.), *Attention and performance* (Vol. VI, pp. 535–556).
- 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. <https://doi.org/10.3389/fpsyg.2011.00351>
- Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, 49(2), 602–615. <https://doi.org/10.3758/s13428-016-0734-0>
- Dylman, A. S., & Bjärtå, A. (2019). When your heart is in your mouth: The effect of second language use on negative emotions. *Cognition and Emotion*. <https://doi.org/10.1080/02699931.2018.1540403>
- Eilola, T. M., & Havelka, J. (2011). Behavioral and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, 15(3), 353–369. <https://doi.org/10.1177/1367006910379263>
- Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion*, 21(5), 1064–1076. <https://doi.org/10.1080/02699930601054109>
- Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 297–339. <https://doi.org/10.1017/S0272263102002140>

- Ellis, N. C., & Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, 9(3), 604–620. <https://doi.org/10.1111/tops.12256>
- Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words is categorical and general. *Emotion*, 8(4), 453–457. <https://doi.org/10.1037/a0012887>
- Estes, Z., & Verges, M. (2008). Freeze or flee? Negative stimuli elicit selective responding. *Cognition*, 108(2), 557–565. <https://doi.org/10.1016/j.cognition.2008.03.003>
- Ferré, P., Anglada-Tort, M., & Guasch, M. (2018). Processing of emotion words in bilinguals: Testing the effects of word concreteness, task type and language status. *Second Language Research*, 34(3), 371–394. <https://doi.org/10.1177/0267658317744008>
- Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 696. <https://doi.org/10.1037/0278-7393.22.3.696>
- Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48, 1358–1369. <https://doi.org/10.3758/s13428-015-0684-y>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods*, 48, 272–284. <https://doi.org/10.3758/s13428-015-0572-5>
- 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. <https://doi.org/10.1371/journal.pone.0186027>
- Imbault, C., Titone, D., Warriner, A. B., & Kuperman, V. (2021). How are words felt in a second language: Norms for 2,628 English words for valence and arousal by L2 speakers. *Bilingualism: Language and Cognition*, 24(2), 281–292. <https://doi.org/10.1017/s1366728920000474>
- Imbir, K. K., Jurkiewicz, G., Duda-Gólawska, J., & Żygierewicz, J. (2019). The role of valence and origin of emotions in emotional categorization task for words. *Journal of Neurolinguistics*, 52, 100854. <https://doi.org/10.1016/j.jneuroling.2019.100854>
- Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, 39, 1198–1210. <https://doi.org/10.3758/s13421-011-0094-z>
- Jin, Y., Ma, Y., Li, M., & Zheng, X. (2023). The influence of word concreteness on acquired positive emotion association: an event-related potential study. *Acta Psychologica*, 240, 104052. <https://doi.org/10.1016/j.actpsy.2023.104052>
- Jończyk, R. (2016). *Affect-language interactions in native and non-native English speakers*. Springer International Publishing AG. <https://doi.org/10.1007/978-3-319-47635-3>
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114(1), 1–37. <https://doi.org/10.1037/0033-295X.114.1.1>
- Kazanas, S. A., & Altarriba, J. (2015). The automatic activation of emotion and emotion-laden words: Evidence from a masked and unmasked priming paradigm. *The American Journal of Psychology*, 128(3), 323–336. <https://doi.org/10.5406/amerjpsyc.128.3.0323>
- Kazanas, S. A., & Altarriba, J. (2016). Emotion word processing: Effects of word type and valence in Spanish-English bilinguals. *Journal of Psycholinguistic Research*, 45, 395–406. <https://doi.org/10.1007/s10936-015-9357-3>
- Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, 72(1), 269–319. <https://doi.org/10.1111/lang.12479>
- Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Progress in Brain Research*, 156, 147–183. [https://doi.org/10.1016/S0079-6123\(06\)56008-X](https://doi.org/10.1016/S0079-6123(06)56008-X)
- Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: Early cortical responses to emotional words during reading. *Psychological Science*, 18(6), 475–480. <https://doi.org/10.1111/j.1467-9280.2007.01924.x>
- Knickerbocker, H., Johnson, R. L., & Altarriba, J. (2015). Emotion effects during reading: Influence of an emotion target word on eye movements and processing. *Cognition and Emotion*, 29(5), 784–806. <https://doi.org/10.1080/02699931.2014.938023>

- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, **140**(1), 14. <https://doi.org/10.1037/a0021446>
- Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, **112**(3), 473–481. <https://doi.org/10.1016/j.cognition.2009.06.007>
- 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. <https://doi.org/10.1006/jmla.1994.1008>
- Kühne, K., & Gianelli, C. (2019). Is embodied cognition bilingual? Current evidence and perspectives of the embodied cognition approach to bilingual language processing. *Frontiers in Psychology*, **10**, 108. <https://doi.org/10.3389/fpsyg.2019.00108>
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, **143**(3), 1065. <https://doi.org/10.1037/a0035669>
- Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior Research Methods*, **49**(2), 457–470.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International Affective Picture System (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, **1**(39–58), 3.
- Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion*, **6**(1), 62. <https://doi.org/10.1037/1528-3542.6.1.62>
- 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. <https://doi.org/10.1037/1528-3542.8.4.445>
- Liu, J., Fan, L., Tian, L., Li, C., & Feng, W. (2023). The neural mechanisms of explicit and implicit processing of Chinese emotion-label and emotion-laden words: Evidence from emotional categorization and emotional Stroop tasks. *Language, Cognition and Neuroscience*, **38**(10), 1412–1429. <https://doi.org/10.1080/23273798.2022.2093389>
- Liu, L., Margoni, F., He, Y., & Liu, H. (2021). Neural substrates of the interplay between cognitive load and emotional involvement in bilingual decision making. *Neuropsychologia*, **151**, 107721. <https://doi.org/10.1016/j.neuropsychologia.2020.107721>
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, **19**(1), 1.
- Martin, J. M., & Altarriba, J. (2017). Effects of valence on hemispheric specialization for emotion word processing. *Language and Speech*, **60**(4), 597–613. <https://doi.org/10.1177/0023830916686128>
- Mohammad, S. (2018, July). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In Proceedings of the 56th annual meeting of the Association for Computational Linguistics (Volume 1: Long papers) (pp. 174–184).
- Morid, M., & Sabourin, L. (2024). Affective and sensory–motor norms for idioms by L1 and L2 English speakers. *Applied Psycholinguistics*, **45**(1), 138–155. <https://doi.org/10.1017/S0142716423000504>
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, **13**, 45–52. <https://doi.org/10.3758/BF03193811>
- Palazova, M., Sommer, W., & Schacht, A. (2013). Interplay of emotional valence and concreteness in word processing: An event-related potential study with verbs. *Brain and language*, **125**(3), 264–271. <https://doi.org/10.1016/j.bandl.2013.02.008>
- Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, **11**(2), 147–164. <https://doi.org/10.1017/S1366728908003283>
- Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, **47**(6), 405–428. <https://doi.org/10.1080/00207594.2012.743665>
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, **11**(4), 357–383. <https://doi.org/10.1080/1088430701530730>
- Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of personality and social psychology*, **61**(3), 380. <https://doi.org/10.1037/0022-3514.61.3.380>

- Pustejovsky J (2023). *_clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections_*. R package version 0.5.10, <<https://CRAN.R-project.org/package=clubSandwich>>.
- R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.
- Rofes, A., Zakariás, L., Ceder, K., Lind, M., Johansson, M. B., De Aguiar, V., ... & Howard, D. (2018). Imageability ratings across languages. *Behavior research methods*, 50(3), 1187–1197. <https://doi.org/10.3758/s13428-017-0936-0>
- 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(1), 95–104. <https://doi.org/10.1016/j.bio psycho.2008.03.010>
- Sheikh, N. A., & Titone, D. (2016). The embodiment of emotion words in a second language: An eye-movement study. *Cognition and Emotion*, 30(3), 488–500
- Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, 49(1), 111–123.
- Su, I. F., Yum, Y. N., & Lau, D. K. Y. (2023). Hong Kong Chinese character psycholinguistic norms: Ratings of 4376 single Chinese characters on semantic radical transparency, age-of-acquisition, familiarity, imageability, and concreteness. *Behavior Research Methods*, 55(6), 2989–3008. <https://doi.org/10.3758/s13428-022-01928-y>
- Tang, D., Fu, Y., Wang, H., Liu, B., Zang, A., & Kärkkäinen, T. (2023). The embodiment of emotion-label words and emotion-laden words: Evidence from late Chinese–English bilinguals. *Frontiers in Psychology*, 14, 1143064. <https://doi.org/10.3389/fpsyg.2023.1143064>
- Tang, E., & Ding, H. (2024). Emotion effects in second language processing: Evidence from eye movements in natural sentence reading. *Bilingualism: Language and Cognition*, 27(3), 460–479. <https://doi.org/10.1017/S1366728923000718>
- Tenderini, M. S., de Leeuw, E., Eilola, T. M., & Pearce, M. T. (2022). Reduced cross-modal affective priming in the L2 of late bilinguals depends on L2 exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(2), 284. <https://doi.org/10.1037/xlm0000889>
- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, 40(6), 604–634. <https://doi.org/10.3102/107699861560609>
- Toivo, W., & Scheepers, C. (2019). Pupillary responses to affective words in bilinguals' first versus second language. *Plos one*, 14(4), e0210450. <https://doi.org/10.1371/journal.pone.0210450>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.
- Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, 24(7), 1767–1777. <https://doi.org/10.1093/cercor/bht025>
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, 1(2), 219–247. <https://doi.org/10.1515/LANGCOG.2009.011>
- Vinson, D., Ponari, M., & Vigliocco, G. (2014). How does emotional content affect lexical processing? *Cognition & emotion*, 28(4), 737–746. <https://doi.org/10.1007/s10936-019-09647-w>
- Winkel, 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. <https://doi.org/10.1080/00207594.2013.793800>
- Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, 71(2), 487–536. <https://doi.org/10.1111/lang.12444>
- Yao, Z., & Wang, Z. (2014). Concreteness of positive word contributions to affective priming: An ERP study. *International Journal of Psychophysiology*, 93(3), 275–282. <https://doi.org/10.1016/j.ijpsycho.2014.06.005>
- Yao, Z., Yu, D., Wang, L., Zhu, X., Guo, J., & Wang, Z. (2016). Effects of valence and arousal on emotion word processing are modulated by concreteness: Behavioral and ERP evidence from a lexical decision task. *International Journal of Psychophysiology*, 110, 231–242. <https://doi.org/10.1016/j.ijpsycho.2016.07.499>
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60(4), 502–529. <https://doi.org/10.1016/j.jml.2009.02.001>

- Yates, M.** (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**(6), 1385. <https://doi.org/10.1037/0278-7393.31.6.1385>
- Zhang, J., Wu, C., Yuan, Z., & Meng, Y.** (2020). Different early and late processing of emotion-label words and emotion-laden words in a second language: An ERP study. *Second Language Research*, **36**(3), 399–412. <https://doi.org/10.1177/0267658318804850>
- Zhong, Y., Shao, Y., & Yi, W.** (2025). Affective and non-affective psycholinguistic norms for 500 Chinese three-character idiomatic expressions. *Behavior Research Methods*, **57**(4), 116. <https://doi.org/10.3758/s13428-025-02633-2>

Cite this article: Zhong, Y., Liu, L., & Brehm, L. (2026). The processing advantage of emotion words in L1 speakers and L2 speakers: A meta-analysis. *Applied Psycholinguistics*. <https://doi.org/10.1017/S0142716426100630>