



Subjective age of acquisition norms for 1604 English words by Spanish L2 speakers of English and their relationship with lexico-semantic, affective, sociolinguistic and proficiency variables

Sara Rodríguez-Cuadrado¹ · José Antonio Hinojosa^{2,3,4} · Marc Guasch⁵ · Carlos Romero-Rivas¹ · Lucía Sabater^{1,2} · Paz Suárez-Coalla⁶ · Pilar Ferré⁵

Accepted: 12 November 2022
© The Author(s) 2022

Abstract

Psycholinguistic studies have shown that there are many variables implicated in language comprehension and production. At the lexical level, subjective age of acquisition (AoA), the estimate of the age at which a word is acquired, is key for stimuli selection in psycholinguistic studies. AoA databases in English are often used when testing a variety of phenomena in second language (L2) speakers of English. However, these have limitations, as the norms are not provided by the target population (L2 speakers of English) but by native English speakers. In this study, we asked native Spanish L2 speakers of English to provide subjective AoA ratings for 1604 English words, and investigated whether factors related to 14 lexico-semantic and affective variables, both in Spanish and English, and to the speakers' profile (i.e., sociolinguistic variables and L2 proficiency), were related to the L2 AoA ratings. We used boosted regression trees, an advanced form of regression analysis based on machine learning and boosting algorithms, to analyse the data. Our results showed that the model accounted for a relevant proportion of deviance (58.56%), with the English AoA provided by native English speakers being the strongest predictor for L2 AoA. Additionally, L2 AoA correlated with L2 reaction times. Our database is a useful tool for the research community running psycholinguistic studies in L2 speakers of English. It adds knowledge about which factors—linked to the characteristics of both the linguistic stimuli and the speakers—affect L2 subjective AoA. The database and the data can be downloaded from: https://osf.io/gr8xd/?view_only=73b01dccbedb4d7897c8d104d3d68c46.

Keywords Age of acquisition · Second language speakers · Boosted regression trees

Introduction

Research on bilingualism has acquired a central role in psycholinguistics in recent decades. One of the main research objectives has been to establish the distinctive characteristics of language processing in bilinguals. To that end, many studies have examined bilinguals' performance in tasks including second language (L2) words, sentences or longer utterances, and compared them, in some cases, to first language (L1) units (e.g., Ferré et al., 2017; Whitford & Titone, 2017). To select the experimental materials for psycholinguistic studies, it is necessary to consider several lexico-semantic and affective variables which are known to influence word processing, such as word frequency, familiarity, concreteness or valence, among others. A common practice in bilingualism research is to obtain the values for these variables from word ratings provided by native speakers. That is, L2 speakers are presented in most studies with L2 materials

✉ Sara Rodríguez-Cuadrado
sara.rodriguezcuadrado@uam.es

¹ Departamento Interfacultativo de Psicología Evolutiva y de la Educación, Facultad de Formación del Profesorado y Educación, Universidad Autónoma de Madrid, C/ Francisco Tomás y Valiente, n° 3, 28049 Madrid, Spain

² Instituto Pluridisciplinar, Universidad Complutense de Madrid, Madrid, Spain

³ Departamento Psicología Experimental, Procesos Cognitivos y Logopedia, Universidad Complutense de Madrid, Madrid, Spain

⁴ Centro de Investigación Nebrija en Cognición (CINC), Universidad Nebrija, Madrid, Spain

⁵ Departamento de Psicología y CRAMC, Universitat Rovira i Virgili, Tarragona, Spain

⁶ Departamento de Psicología y Grupo de Investigación INCO, Universidad de Oviedo, Oviedo, Spain

rated by L1 speakers, which limits our understanding of the influence of those lexico-semantic and affective variables and potentially jeopardises the validity of the research findings. Using data provided by L2 speakers could make L2 research more ecologically valid and provide further insight into L2 acquisition and processing. However, only a few studies so far have gathered ratings from L2 speakers for distinct lexico-semantic and affective variables. For instance, Chen and Dong (2019) collected subjective frequency ratings for English words in a sample of Chinese L2 English learners and compared them to objective frequency ratings obtained from six English corpora. Interestingly, a superiority for subjective L2 frequency ratings over objective ratings in predicting L2 lexical processing was observed. Considering these findings, the authors recommended that researchers collect subjective frequency estimates from L2 learners in studies about L2 processing. A similar conclusion was reached by Wang and Chen (2020), who collected familiarity ratings for English words from Chinese-English bilinguals and found only a moderate correlation between these ratings and objective frequency measures obtained from English film subtitles (SUBTLEX-UK, van Heuven et al., 2014). Similarly, in the study of Hubers et al. (2020), German L2 learners of Dutch rated Dutch idioms on frequency of exposure and frequency of use, meaning familiarity, imageability and transparency, finding that non-native speakers' intuition regarding L2 idioms was a reliable source of information (see also Hasegawa, 2010, for L2 imageability).

Another set of studies, focused on affective features, collected valence ratings in L2 speakers of English who had a variety of L1s (Ferré et al., 2022a, Garrido & Prada, 2021, Imbault et al., 2021, Vélez-Urbe & Rosselli, 2019). All of them found more attenuated affective ratings in the L2 than in the L1. In addition, Imbault et al. (2021) showed that ratings were modulated by the characteristics of the words and the speakers. In particular, they found more native-like ratings for high-frequency words and by more proficient L2 speakers who had lived longer in the L2 country. Similarly, the L2 experience of the speakers (proficiency and age and context of L2 acquisition) was reported to modulate familiarity ratings of L2 words in the study by Garrido and Prada (2021).

Another variable that has attracted the attention of researchers on bilingualism is the age of acquisition (AoA) of words. AoA is the age at which a word was acquired, and there are two basic methods used to estimate it: an objective method, based on studies on children (e.g., asking children of different ages to name pictures; see Morrison et al., 1997), and a subjective method, based on adults' estimations of the age at which they acquired the words. Several studies have found a high correlation between AoA values obtained from both methods (e.g., Chalard et al., 2003; Liu et al., 2011; Morrison et al., 1997; Walley & Metsala, 1992).

AoA is a variable that can help researchers understand the links between orthography, phonology and semantics in the lexicon (Juhász, 2005). In fact, the use of AoA (either manipulating it or controlling it across experimental conditions) is quite common in psycholinguistic research. The "age of acquisition effect" reflects how words with an early AoA are recognised faster and/or more accurately than those with a late AoA (e.g., Cortese & Khanna, 2007; Juhász & Rayner, 2006; Kuperman et al., 2012, 2014; Sereno & O'Donnell, 2009; for a review see Johnston & Barry, 2006). Most research investigating AoA has relied on data from normative studies to obtain the experimental materials. A few databases are available on both objective and subjective AoA in a variety of languages (e.g., Barca et al., 2002, in Italian; Cameirao & Vicente, 2010, in Portuguese; Ferrand et al., 2008, in French; Kuperman et al., 2012, in English; Liu et al., 2007, in Chinese; Moors et al., 2013, in Dutch; Schröder et al., 2012, in German; Alonso et al., 2015; Cuetos et al., 1999; Hinojosa et al., 2016b, and Piñeiro & Manzano, 2000, in Spanish; see also Łuniewska et al., 2016, 2019, for studies that have compared AoA on a total of 32 languages). All these normative studies provide AoA ratings collected from native speakers. However, the AoA of words in an L2 may also be a relevant variable in explaining bilingual language processing.

Thus, to explore L2 vocabulary acquisition, an essential stepping stone would be to elucidate the age at which L2 speakers acquire certain words, and the variables that influence this acquisition. To enable such research, normative AoA ratings for L2 words would be highly valuable. Izura and Ellis (2002) obtained L2 AoA ratings, but for a small set of items. To our knowledge, only two studies have collected this kind of rating for a large set of words, gathering subjective AoA ratings for a set of English words from L2 speakers (unbalanced late bilinguals): Dutch-English bilinguals in the study of Dirix and Duyck (2017), and Chinese-English bilinguals in the study of Wang and Chen (2020). These studies reported a relevant effect of the AoA of L2 words in language processing. Concretely, Dirix and Duyck (2017) found that L2 AoA ratings modulated eye movements during reading. Wang and Chen (2020), in turn, demonstrated that L2 AoA could account for an additional part of the variance on lexical decision times of L2 speakers of English (Berger et al., 2019) once other relevant variables had been controlled for. Finally, Izura and Ellis (2002) found that response times (RTs) in a lexical decision task performed in the participants' L2 were predicted by L2 AoA, but not by L1 AoA. These authors concluded that L2 AoA effects reflect the order in which the words were acquired in L2, rather than the order in which their translation equivalents were acquired in the native language. In turn, this would suggest that the AoA effect is not related to the acquisition of words' meanings, but rather to the acquisition of word forms

or the mappings between lexical and semantic representations (that is, in line with the *mapping hypothesis*, see also Cortese & Schock, 2013).

However, the aforementioned databases have not explored in depth the role of (and the relationship amongst) lexico-semantic and affective variables in subjective L2 AoA, nor have they carefully examined speaker-related variables such as the role of proficiency and linguistic history relative to these ratings. In addition, to our knowledge, no database has gathered English AoA ratings provided by late unbalanced Spanish L2 speakers of English, so the present database could be widely used by those researchers working with this population.

In the current study we examined the relationship between subjective AoA in an L2 (in this case, English) with different lexico-semantic and affective variables that impact word processing and are related to AoA. We also considered the participants' sociolinguistic background and measured their L2 proficiency. Regarding lexico-semantic and affective variables, we included concreteness, frequency, familiarity, prevalence, word length, number of orthographic neighbours, iconicity, imageability, sensory experience ratings, semantic size, cognate status, valence and arousal, along with AoA ratings for English and Spanish.

A number of these variables correlate with AoA in L1. Regarding *concreteness* (that is, the extent to which something can be experienced through our senses), concrete words are acquired earlier (e.g., Morris, 1981). As for *frequency* (meaning how often a word is found, mostly in print), high *frequency* words have an earlier acquisition (e.g., Citron et al., 2014), and words acquired earlier are also more *familiar* (familiarity being a measure of a person's experience with a word; see Hinojosa et al., 2016b).¹ Also, words with a low AoA are more *prevalent* ("prevalence" being a measure of a population's word knowledge; Brysbaert & New, 2009; Brysbaert et al., 2016). *Iconicity* (the resemblance between the form and the meaning of a word) helps language acquisition (Imai & Kita, 2014; Massaro & Perlman, 2017; Perniss & Vigliocco, 2014) and is related to AoA (Perry et al., 2015), so words with high iconicity scores are acquired earlier and are more frequent in infancy than words with low iconicity scores in both oral and sign languages (Caselli & Pyers, 2017; Hinojosa et al., 2021; Sidhu et al., 2021; Thompson et al., 2012; Vinson et al., 2008). In addition, *imageability* (a measure of how easy it is for a person to create a mental image of something) scores are higher for words with early AoA (Citron et al., 2014). Words with high *sensory*

experience ratings (SERs, a measure of how much a word generates a sensory experience in the mind of a person) are acquired earlier (see Hinojosa et al., 2016b²). *Semantic size* (a measure of magnitude; e.g., big, small) also correlates positively with AoA, as words with high semantic size are learned earlier (Scott et al., 2019). In addition, *cognate status* (i.e., the degree of orthographic overlap between an L2 word and its L1 translation equivalent) influences L2 word acquisition, where cognates are learned sooner than non-cognates (see Tonzar et al., 2009, and Comesaña et al., 2012 for data on children). Regarding *valence* (a measure of a word's hedonic positive or negative value) and *arousal* (a measure of the internal activation elicited by a word), words with an early AoA tend to be mostly pleasant and calm (Citron et al., 2014; Hinojosa et al., 2016b; Warriner et al., 2013). In addition, a variable that correlates with L2 AoA is L2 familiarity. The study of Wang and Chen (2020) shows an interaction between L2 familiarity and L2 AoA, where those words which were less familiar in the L2 were also acquired later in the L2. A positive correlation between L1 and L2 AoA was also found.

Many of the aforementioned lexico-semantic variables modulate word processing, as AoA does (e.g., Brysbaert et al., 2016; Dijksterhuis & Aarts, 2003; Ferré et al., 2018; Hinojosa et al., 2010, 2020; Imai et al., 2008; Juhasz & Rayner, 2003; Kantartzis et al., 2011; Kousta et al., 2009; Kuperman et al., 2012; Peeters, 2016; Scott et al., 2009, 2019; Sereno et al., 2009; Sidhu et al., 2020; Winter et al., 2017; Yao et al., 2013). In addition, it is important to consider other relevant variables in word processing such as *word length* (which positively influences processing time; see Davies et al., 2013 and Kuperman et al., 2012), or *orthographic neighbourhood size* (the number of similarly spelled words), which, in turn, is related to psycholinguistic variables such as word frequency (e.g., Grainger, 1990; Yarkoni et al., 2008). Words from sparse neighbourhoods are benefitted in recognition, naming and lexical decision tasks (for developmental differences, see Garlock et al., 2001). Finally, the studies that have examined the effect of *semantic size* show that words referring to bigger things benefit from faster recognition (Scott et al., 2019; Sereno et al., 2009; Yao et al., 2013).

The influence of all these variables on both L1 and L2 processing has been notorious (e.g., van Heuven et al., 1998), where cognate status greatly modulates these effects, facilitating L2 processing (e.g., Costa et al., 2000). Given the relevance of these variables on the study of L1 and L2

¹ We performed a correlational analysis on the mean AoA and familiarity ratings provided in the supplementary materials by Hinojosa et al. (2016b) showing that words acquired earlier were also more familiar ($r = -.665, p < .001$).

² We performed a correlational analysis on the mean AoA and SERs provided in the supplementary materials by Hinojosa et al. (2016b) showing that words acquired earlier have higher SERs ($r = -.255, p < .001$).

processing, and their relationship with AoA in the L1, we include them in our study to better describe and understand subjective AoA ratings in L2, specifically those of late unbalanced L2 Spanish-English bilinguals. In addition, we examine the relationship between the above-mentioned variables and ratings of AoA in the L2 (English) considering, when available, the ratings of these variables both in L1 (Spanish) and in L2 (English).

It should be noted that, although there are two previous studies gathering L2 AoA ratings (Dirix & Duyck, 2017; Wang & Chen, 2020), they did not take into consideration the sociolinguistic characteristics of their participants. In this study, however, we examine several subjective and objective measures of the participants' linguistic history and proficiency. Specifically, we have included a sociolinguistic survey (see Materials and Procedure) and have objectively measured the participants' English proficiency with the LexTALE English test (Lexical Test for Advanced Learners of English; Lemhöfer & Broersma, 2012), given that variables related with the speakers such as proficiency and language use, among others, can influence how lexico-semantic (Guasch et al., 2008; Perani & Abutalebi, 2005) and affective (Degner et al., 2011) variables impact L1/L2 language processing.

In sum, knowing the L2 AoA of words would allow researchers to select those stimuli that participants are likely to know (or not know, depending on the research purposes). Having an appropriately normed dataset on AoA *by* and *for* L2 speakers fills a gap in the field. Also, exploring how L2 AoA is influenced by lexico-semantic and affective variables, and sociolinguistic and proficiency variables, is not only informative by itself, but can have relevant educational implications in the context of English as a foreign language (EFL) and bilingual education programmes (BEP). For instance, in the 2019/2020 academic year, the Spanish-English BEP was present in 50% of public schools, 59.2% of high schools and 54.7% of charter schools (Mañas Antón, 2019).

In addition, Izura and Ellis (2002) investigated (with a limited set of words) the locus of the AoA effect, by looking at whether L2 reaction times (RTs) could be predicted by AoA values in the first or second language of bilingual speakers. We aim at expanding this investigation with our L2 AoA data and L2 RTs from a previously published study (Berger et al., 2019).

Objectives and hypotheses

Our first objective is to gather subjective AoA data for 1604 English words provided by Spanish L2 speakers of English. Our second objective is to examine if, and how, several variables related to the words, both lexico-semantic and affective, influence the ratings provided by Spanish L2

speakers of English. Thirdly, we will explore if, and how, sociolinguistic and proficiency variables related to the speakers influence these AoA L2 ratings. Finally, we will look at the relationship between AoA and L2 processing with L2 RTs. Accordingly, and based on previous literature, we derive the following hypotheses linked to our second and third objectives: Regarding the second objective, we expect a positive relationship between L1 (Spanish) AoA and L2 AoA scores (Dirix & Duyck, 2017; Wang & Chen, 2020). In addition, and in line with Wang and Chen (2020), we expect to find a negative relationship between L2 familiarity (that is, familiarity scores in English) and L2 AoA, so more familiar words would be acquired earlier. With respect to the third objective, we are not knowledgeable of any study that has looked at sociolinguistic variables in detail when examining AoA. However, previous studies have shown a relationship between several sociolinguistic factors and lexico-semantic and/or affective variables in L2. For instance, Imbault et al. (2021) found that L2 speakers had greater word knowledge and provided more native-like ratings for valence and arousal when they learned English at a younger age, for more years, in an immersive context, had a high self-reported proficiency, and used English frequently in their day-to-day lives. Therefore, we expect an effect of those variables on L2 AoA ratings. Finally, following Izura and Ellis (2002), we hypothesise that what matters for RTs in an L2 lexical decision is the L2 AoA of the L2 words, but not the AoA in Spanish (the participants' L1) of the translation equivalents, which will speak in favour of the *mapping hypothesis* and against a semantic locus in AoA effects.

Methods

Participants

A total of 309 participants took part in the study. After data trimming (see "Results"), the final group consisted of 292 participants (242 women, 50 men) with a mean age of 20.91 years ($SD = 3.18$; range = 18–39), and who were recruited from several Spanish universities, including Universidad de Oviedo (41%), Universitat Rovira i Virgili (26%), Universidad Complutense de Madrid (23%) and other universities (10%).

After completing the AoA task (see below), participants took the LexTALE English test (Lemhöfer & Broersma, 2012) and filled out a sociolinguistic survey (see "Materials and procedure"). Their mean score in the LexTALE test was 67.02 ($SD = 9.22$; range = 50–90). Following the criteria established by Lemhöfer and Broersma with a Dutch sample (2012), advanced users (C1–C2) would score between 80 and 100, upper intermediate users (B2) would score between 60 and 79, and lower intermediate (B1) and basic (A2) users

Table 1 Mean and standard deviation (in parentheses) of age of acquisition and self-rated proficiency in English of the participants

Age of acquisition of English (in years)				
Speaking	Reading	Writing	Listening	Mean
8.00 (4.11)	7.72 (3.05)	8.13 (3.25)	7.30 (3.88)	7.79 (3.61)
Self-rated English proficiency (1–7 scale; 1 = very bad, 7 = very good)				
Speaking	Reading	Writing	Listening	Mean
4.51 (1.46)	5.36 (1.27)	4.62 (1.32)	4.98 (1.50)	4.87 (1.43)

Bold indicates mean of four linguistic skills

would score 59 or less (Lemhöfer & Broersma, 2012, p. 341). Therefore, based on Lemhöfer and Broersma (2012), the average English level of our participants could be classified as upper intermediate.

Responses to the sociolinguistic questionnaire revealed that all participants were native Spanish speakers or bilingual in Spanish and one of the other official languages in Spain (i.e., Basque, Catalan or Galician); 99.31% of the participants had completed at least a baccalaureate level of studies, professional training or higher, and 38.01% had completed a university degree or higher (e.g., masters, PhD). The contexts in which participants reported having learned English were mainly at school (94.86%) and informal settings (41.10%; i.e., with friends, films or music). Only 5.82% of the participants reported having learned English in an English-speaking country, where they had lived for less than 1 year on average. Participants reported spending an average of 13.92 years ($SD = 3.66$; range = 3–25) learning English. They reported using English 23.95% ($SD = 21.23$; range = 0–90) of the time in a week. Table 1 shows their mean estimated AoA of English and their self-rated proficiency in English in the four basic linguistic skills.

All participants accepted an informed consent form in Spanish before starting the study, which was conducted with the approval of the Ethics Committee of the Principality of Asturias (reference 153/19). The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki.

Materials and procedure

Words were selected considering the AoA values of normative studies developed in English (Scott et al., 2019) and Spanish (Sabater et al., 2020). We firstly selected a large set of words ($n = 1258$) with an AoA equal to or below 7 years of age, to be sure that the participants were likely to know them. The selection was completed with 346 additional words with an AoA above 7, resulting in a total of 1604 words.

The words were randomly divided into 11 lists, nine containing 146 words and two containing 145 words. Each list was rated by a minimum of 25 participants ($M = 26.55$, $SD = 1.21$, range = 25–28).

A single questionnaire was designed in web format and was administered online with a unique URL that randomly fed from one of the 11 lists at each access. The questionnaire started with an information screen in which participants had to provide their explicit consent to participate by ticking a box. Next, participants were asked to provide their home university, age and gender. These first two screens were presented in Spanish to ensure understandability. Then, the task instructions were displayed (in English) as follows: “A word’s age of acquisition is the age at which that word was initially learned. Please estimate when in your life you think you first acquired or learned each of the presented words. That is, try to remember how old you were when you learned each word either in its spoken or written form (whichever came first). We mean the age at which you would have understood that word if somebody had used it in front of you, **EVEN IF YOU DID NOT** use, read, or write it at the time. In order to indicate the estimated age of acquisition, please click in one of the following boxes:”. Next, an image of the response scale was presented (see Fig. 1), followed by the reminder: “Please, remember that you are not being asked about the age at which you acquired the word in Spanish, but the age at which you think you acquired the word in English.”

After displaying the instructions, the words from the selected list were presented randomly, one at a time, following the layout of Fig. 1 with a progress bar at the bottom. A response to each word was required to advance throughout the task, but participants had the option to mark a word as unknown.

After the ratings, the English version of the LexTALE vocabulary test (Lemhöfer & Broersma, 2012) was administered, closely following the implementation and scoring system proposed on the LexTALE website (www.lextale.com). It is composed of 60 trials and takes less than 5 minutes to complete. Participants received feedback on their scores immediately after completing the test.

In the final part of the questionnaire, participants were asked in Spanish to answer some questions about their sociolinguistic background. This section was divided into three screens. In the first one, they were asked about their level of education, their mother tongue (where they could report a maximum of two languages), and if they have or had any type of difficulty affecting reading and/or writing (and which one). On the next screen they were asked about the age at which they had acquired four basic skills in English, namely speaking, reading, writing and listening, and the number of years they had been studying English and the context(s) in which they had acquired English (i.e., at home, at school,

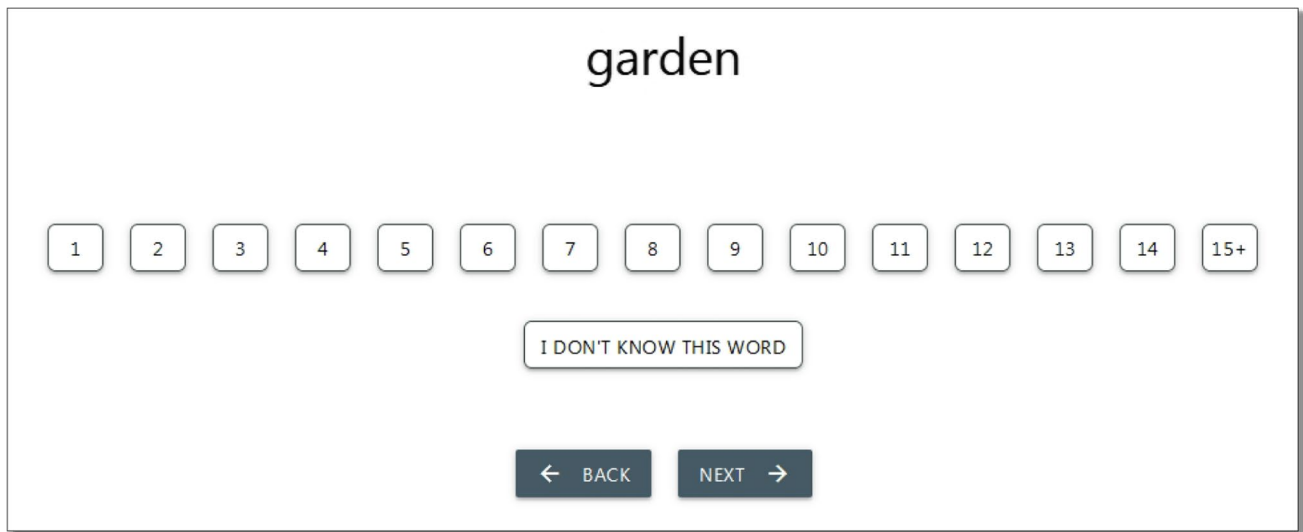


Fig. 1 Layout of a word and the rating scale in the age of acquisition rating task

informally [i.e., with friends, through films or music], living in an English-speaking country, or other). If they ticked the option of having lived in an English-speaking country, they were asked about which country and the time they had spent living there. They were also asked to rate their proficiency level on a 7-point scale (1 = very bad, 7 = very good) for each of the four basic skills—speaking, reading, writing and listening—in English. Finally, on the last screen before submitting the information, they were asked to estimate the percentage of time they used English over the course of a week and could add clarifying comments if they so wished.

The average time to complete the whole task was around 20 minutes. Each participant rated only one list of words and did not complete the questionnaire more than once.

Finally, to explore the relationship between L2 AoA values and L2 RTs in native Spanish L2 speakers of English, we extracted RT data from Berger et al. (2019). Out of their 1315 non-native speakers, 532 (40%) were dominant in Spanish, where we shared 1089 words between databases.

Results

Availability of the norms

The ratings are available at https://osf.io/gr8xd/?view_only=73b01dccbedb4d7897c8d104d3d68c46 under the name “Subjective age of acquisition norms for 1604 English words by Spanish L2 speakers of English”. The datasheet contains the list of 1604 words rated in English together with their translation to Spanish, and the mean AoA (*Mean_AoA_Eng*) and their standard deviation (*SD_AoA_Eng*). For the calculation of other indicators (e.g., standard error), the number of

people who rated each word is also given (*N*). However, not all words were known by all participants, so a column with the percentage of participants who knew each word is also included (*Perc_knowledge*).

A second file entitled “Lexico-semantic and affective values of the items” is also available at https://osf.io/gr8xd/?view_only=73b01dccbedb4d7897c8d104d3d68c46. This file has two sheets. The first sheet (entitled “Summary”) lists all the variables with their abbreviation (*Variable*), the *predictor* (e.g., “concreteness in English”), the number of words out of the 1,604 where the values associated to the variables were available (*Availability*) and the *source* (i.e., which database) from which we extracted the values of the variables. There is also a legend of sources with the abbreviated and full reference of the source (this is also available in Table 2). On the second sheet (entitled “Data”) the values for each variable associated to the word (when available) are listed. Please note that headings correspond to the variable name used on the first sheet.

Data cleaning

An initial sample of 309 participants took part in the study, but 292 remained after data trimming. Specifically, correlations between the responses of each participant and the mean responses of the other participants in the same questionnaire were computed to detect anomalous response patterns. Participants with a personal correlation with the mean of the group of less than 0.1 were discarded. This led to the exclusion of 17 participants (5.5% of the data). As observed in Fig. 2, the L2 AoA ratings data resemble those reported by Dirix and Duyck (2017) and Wang and Chen (2020).

Table 2 Predictors included in the study, classified by type of variable (lexico-semantic, affective, sociolinguistic or proficiency). Descriptions of the predictors and the abbreviations used in the analysis and graphs are provided, as well as the sources of the values and their relative contribution (RC) to the model (please note that if the predictor did not enter the simplified model, this is marked as n/a). Highest RCs are highlighted in bold.

Abbreviation	Predictor	Description	Source(s)	RC
Lexico-semantic variables				
eng_con	Concreteness in English	The extent to which an English word can be experienced by our senses	Scott et al., 2019	n/a
spa_con	Concreteness in Spanish	The extent to which the Spanish translation of the English word can be experienced by our senses	Duchon et al., 2013; Ferré et al., 2012; Guasch et al., 2016; Hinojosa et al., 2016b	1.4%
eng_zipf	Frequency in English	Word frequency in Zipf scale for the English word	Brybaert & New, 2009	2.9%
spa_zipf	Frequency in Spanish	Word frequency in Zipf scale for the Spanish translation of the English word	Duchon et al., 2013	1.8%
eng_fam	Familiarity in English	Familiarity of the English word	Scott et al., 2019	7.7%
spa_fam	Familiarity in Spanish	Familiarity of the Spanish translation of the English word	Duchon et al., 2013; Ferré et al., 2012; Guasch et al., 2016; Hinojosa et al., 2016b	1.8%
eng_prevalence	Prevalence in English	A population's knowledge of a word in English	Brybaert et al., 2019	1.3%
spa_prevalence	Prevalence in Spanish	A population's knowledge of the Spanish translation of the English word	Aguasvivas et al., 2018	1.6%
eng_length	Word length in English	Length in number of letters of the English word	n/a	n/a
spa_length	Word length in Spanish	Length in number of letters of the Spanish translation of the English word	n/a	n/a
eng_old20	Orthographic neighbours in English	Average Levenshtein distance of the 20 nearest neighbours in English	Balota et al., 2007	1.1%
spa_old20	Orthographic neighbours in Spanish	Average Levenshtein distance of the 20 nearest neighbours for the Spanish translation of the English word	Duchon et al., 2013	n/a
eng_ico	Iconicity in English	The resemblance between an English word and its meaning	Winter et al., 2022	1.1%
spa_ico	Iconicity in Spanish	The resemblance between the Spanish translation of the English word and its meaning	Hinojosa et al., 2021	n/a
eng_ima	Imageability in English	How easy it is to create a mental image of an English word	Scott et al., 2019	1.6%
spa_ima	Imageability in Spanish	How easy it is to create a mental image of the Spanish translation of the English word	Duchon et al., 2013; Guasch et al., 2016	1.6%
eng_ser	Sensory experience ratings in English	The extent to which an English word generates a sensory experience in the mind of a person	Juhász & Yap, 2013	1.5%
spa_ser	Sensory experience ratings in Spanish	The extent to which the Spanish translation of the English word generates a sensory experience in the mind of a person	Díez-Álamo et al., 2019; Ferré et al., 2022b; Hinojosa et al., 2016b	1.4%
eng_size	Semantic size in English	Magnitude (big, small) of an English word	Scott et al., 2019	1.1%
Nld	Cognate status	Normalised Levenshtein distance between the English word and its translation in Spanish	Guasch et al., 2013	5.1%
eng_aoa	Age of acquisition in English	The age at which an English word is acquired	Scott et al., 2019	14.3%
spa_aoa	Age of acquisition in Spanish	The age at which the Spanish translation of the English word is acquired	Alonso et al., 2015	4.7%

Table 2 (continued)

Abbreviation	Predictor	Description	Source(s)	RC
Affective variables				
eng_val	Valence in English	An English word's positive/negative value	Scott et al., 2019	1.1%
spa_val	Valence in Spanish	The Spanish translation of the English word's positive/negative value	Ferré et al., 2012; Guasch et al., 2016; Hinojosa et al., 2016a; Stadthagen-Gonzalez et al., 2017	2%
eng_aro	Arousal in English	The internal activation provoked by an English word	Scott et al., 2019	1.1%
spa_aro	Arousal in Spanish	The internal activation provoked by the Spanish translation of the English word	Ferré et al., 2012; Guasch et al., 2016; Hinojosa et al., 2016a; Stadthagen-Gonzalez et al., 2017	n/a
Sociolinguistic variables				
age	Age	Age of the participant	n/a	2.3%
sex	Sex	Sex of the participant	n/a	n/a
education	Level of education	Level of education of the participant	n/a	n/a
difficulty	Reading or writing difficulty	If the participant had or has any reading or writing difficulty	n/a	n/a
age_s	Age of speaking in English	Age at which speaking in English was acquired	n/a	4.2%
age_r	Age of reading in English	Age at which reading in English was acquired	n/a	4.2%
age_w	Age of writing in English	Age at which writing in English was acquired	n/a	7.2%
age_l	Age of listening in English	Age at which listening in English was acquired	n/a	2.2%
years_e	Years studying English	Number of years that the participant has been studying English	n/a	6.2%
home	Context: home	If the participant acquired English at home	n/a	n/a
school	Context: school	If the participant acquired English at school	n/a	n/a
english_country	Context: English-speaking country	If the participant acquired English in an English-speaking country	n/a	n/a
informal	Context: Informal setting	If the participant acquired English informally (i.e., with friends, through films or music)	n/a	n/a
other_context	Context: other context	If the participant acquired English in a context that is not home, school, English-speaking country or informal	n/a	n/a
rep_s	Self-rated English proficiency: speaking	Self-rated English proficiency (scale 1 to 7) in speaking	n/a	1.3%
rep_r	Self-rated English proficiency: reading	Self-rated English proficiency (scale 1 to 7) in reading	n/a	2.7%
rep_w	Self-rated English proficiency: writing	Self-rated English proficiency (scale 1 to 7) in writing	n/a	1.4%
rep_l	Self-rated English proficiency: listening	Self-rated English proficiency (scale 1 to 7) in listening	n/a	1.3%
perc_eng	Use of English during the week	Percentage of time the participant used English over the course of a week	n/a	4.4%
Proficiency variable				
LexTALE_score	Objective English proficiency	Score obtained in the LexTALE test	n/a	5.4%

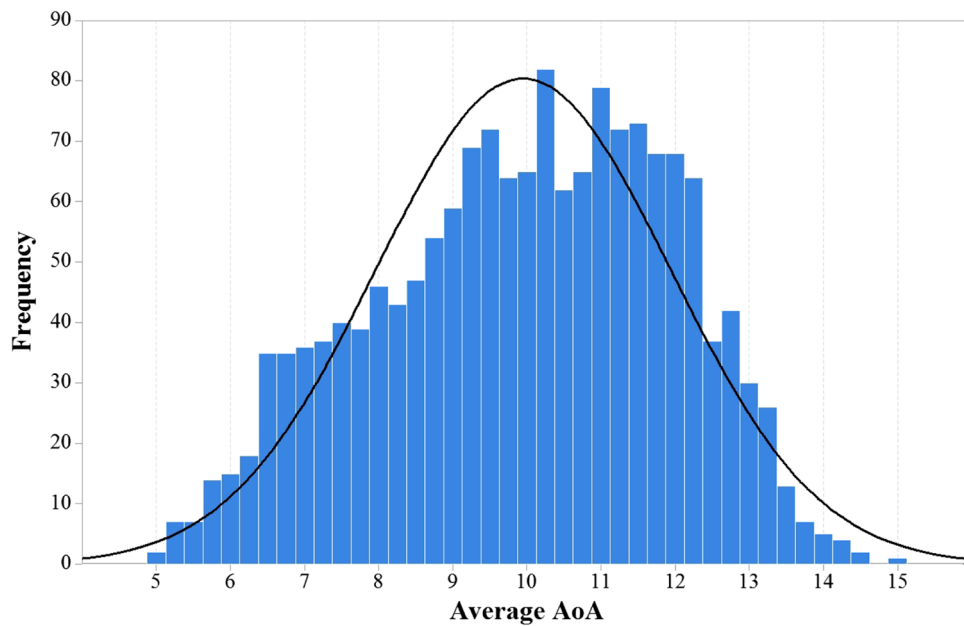


Fig. 2 Frequency distribution of L2 AoA ratings

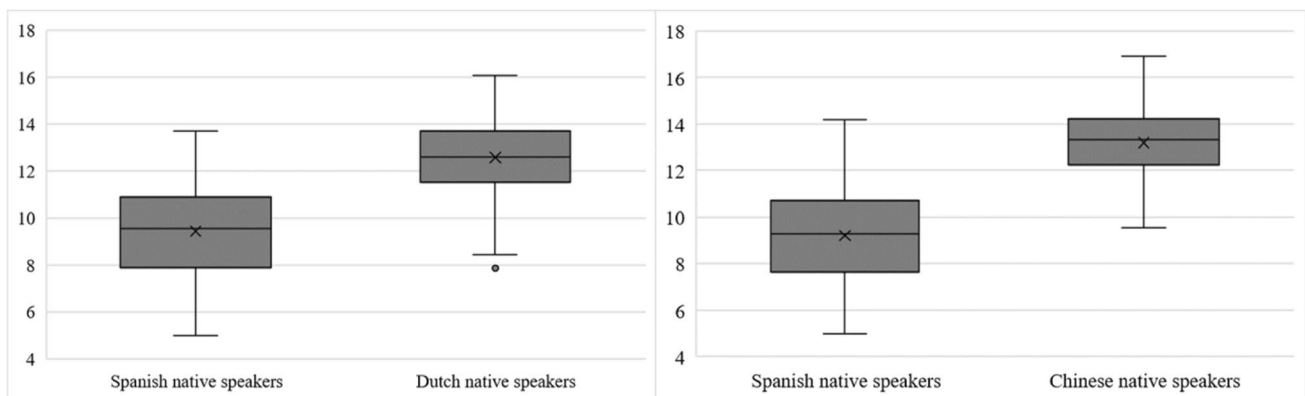


Fig. 3 Left: boxplot displaying L2 AoA values in our database and Dirix and Duyck (2017), including the 549 words available in both databases. Right: boxplot displaying L2 AoA values in our database and Wang and Chen (2020), including the 731 words available in both databases

Reliability and validity

The inter-rater reliability was explored by calculating the intra-class correlation coefficient (ICC) for each AoA questionnaire with the *psych* package in R (Revelle, 2021), using the two-way random effects based on the absolute agreement of multiple raters (2, k). Data were strongly reliable ($M = .93$, $SD = .017$, range = .90–.96), even more so considering that L2 AoA data could bear more variability as L2 learning onset typically differs more between speakers than in L1 acquisition (see Dirix & Duyck, 2017).

The validity of our ratings was assessed by performing Pearson's correlations between our L2 AoA ratings and those provided by Chinese-English (Wang & Chen,

2020) and Dutch-English (Dirix & Duyck, 2017) speakers. Out of the 1835 words included in Wang and Chen (2020) and the 4900 words in Dirix and Duyck (2017), 731 and 549 were available in our database, respectively. The AoA values in our database showed a correlation of .78 ($p < .001$) with those from Wang and Chen (2020) and a correlation of .75 ($p < .001$) with the values of Dirix and Duyck (2017). Furthermore, we wanted to explore whether L2 AoA ratings followed a similar pattern in our database as in Dirix and Duyck (2017) and Wang and Chen (2020), by performing boxplots which indicate how the data are distributed. Data are shown in Fig. 3.

Our pattern of results differs from both Dirix and Duyck (2017) and Wang and Chen (2020). For the shared words in

both databases, our average L2 AoA is lower ($M=9.43$) and the dispersion in our sample is larger ($SD=1.96$) than for the data obtained with native speakers of Dutch ($M=12.57$; $SD=1.5$). Something similar happens in comparison to the data of the native speakers of Chinese ($M=9.21$, $SD=1.91$ in our sample; $M=13.19$, $SD=1.33$ for the Chinese data). The results of two one-way analyses of variance (ANOVAs) (one comparing our data with those of Dirix & Duyck, 2017, and the other comparing our data with those of Wang & Chen, 2020) showed that those differences are significant. That is, the average L2 AoA estimate made by our participants (considering only the shared words between databases) is significantly lower than that estimated by native Dutch speakers, $F(1, 1096)=891.30$, $MSE=2711.96$, $p<.001$, $\eta_p^2=.45$, and by native Chinese speakers, $F(1, 1460)=2130.02$, $MSE=5786.04$, $p<.001$, $\eta_p^2=.593$.

We also correlated our L2 AoA ratings with L1 AoA ratings in English (Scott et al., 2019) and in Spanish (Alonso et al., 2015). All 1604 words were shared between our sample and Scott et al.'s (2019). In the case of Alonso et al. (2015), 1538 words (i.e., the Spanish translations of the English words) were shared between their database and ours. Our AoA values showed a correlation of .64 ($p<.001$) with the values from Scott et al. (2019), and a correlation of .58 ($p<.001$) with the values from Alonso et al. (2015) for Spanish-L1.

We also compared our L2 average AoA with the English L1 AoA values of Scott et al. (2019) and the Spanish L1 AoA values of Alonso et al. (2015, considering only the shared words between databases). The average AoA in our L2 sample is higher ($M=9.95$) and more disperse ($SD=1.99$) than the English L1 AoA data ($M=3.65$; $SD=1.18$). In comparison to the Spanish L1 AoA data, the average L2 AoA is higher ($M=9.91$) but less disperse ($SD=1.99$) than the Spanish L1 AoA data ($M=6.36$; $SD=2.04$). The average L2 AoA for English words was significantly higher in comparison to the English L1 data, $F(1, 3206)=11,887.82$, $MSE=31,838.62$, $p<.001$, $\eta_p^2=.788$, and to the Spanish L1 data, $F(1, 3074)=2388.05$, $MSE=9700.06$, $p<.001$, $\eta_p^2=.437$.

Relationships between L2 AoA and lexico-semantic, affective, sociolinguistic and proficiency variables

Data were analysed using boosted regression trees (BRTs), which are an advanced form of regression analysis based on machine learning and boosting algorithms (Elith et al., 2008). It combines large numbers of simple tree models to optimise predictive performance by fitting many models and combining them for predicting the dependent variable. Also, and unlike other regression methods, BRTs use an algorithm to learn the relationship between predictors and responses, instead of setting up a model first and then estimating

parameters for the model from the data. This implies that BRTs considers that the relationships between predictors and responses are unknown and tries to learn about this relationship by processing inputs and responses and finding dominant patterns. BRTs also identify linear and non-linear relationships, which can be observed on the generated graphs (see Fig. 4). Raw data are available at https://osf.io/gr8xd/?view_only=73b01dccbedb4d7897c8d104d3d68c46

Following the guidelines established by Elith et al. (2008) and Elith and Leathwick (n.d.), we first fitted a model including all predictors, using a tree complexity of 5, a learning rate of 0.01 and a bag fraction of 0.5. This model accounted for a relevant proportion of deviance (58.75%). Then, we simplified the model by dropping predictors until the average change in predictive deviance exceeded their original standard error. The simplified model eliminated 14 out of the 46 predictors, given their low predictive value: English_country, eng_length, difficulty, school, spa_length, home, sex, education, informal, spa_old20, eng_con, other_context, spa_aro and spa_ico. Also, the simplified model accounted for a relevant proportion of deviance, virtually identical to that of the complete model (58.56%, see Fig. 4). The predictors included in the simplified model and their relative contributions (RCs) can be found in Table 2. Below, data will be presented in three subsections: first, variables related to the words (lexico-semantic and affective), then, variables related to the speakers (sociolinguistic and proficiency), and finally, L2 AoA values and their relationship with L2 reaction times.

L2 AoA and lexico-semantic and affective variables

All the RCs of the predictors can be found in Table 2. The predictor with the highest RC was eng_aoa (14.3%), showing that the earlier the English L1 AoA, the earlier the L2 AoA. Then, other predictors showed a moderate RC: eng_fam (7.7%, the higher the familiarity, the earlier the L2 AoA), nld (5.1%; the smaller the Levenshtein distance between English and Spanish words, the earlier the L2 AoA) and spa_aoa (4.7%; the earlier the Spanish AoA, the earlier the L2 AoA). The other predictors had a smaller RC (<3%, see Fig. 4 for more information).

L2 AoA and sociolinguistic and proficiency variables

All the RCs of the predictors can be found in Table 2. The predictor with the highest RC was age_w (7.2%), showing that the earlier the age of writing in English, the earlier the L2 AoA. Other predictors showed a moderate RC: years_e (6.2%; non-linear relationship, although, in general, the more time participants spent learning English, the later their L2 AoA), LexTALE_score (5.4%; the higher the proficiency, the earlier the L2 AoA, becoming more acute

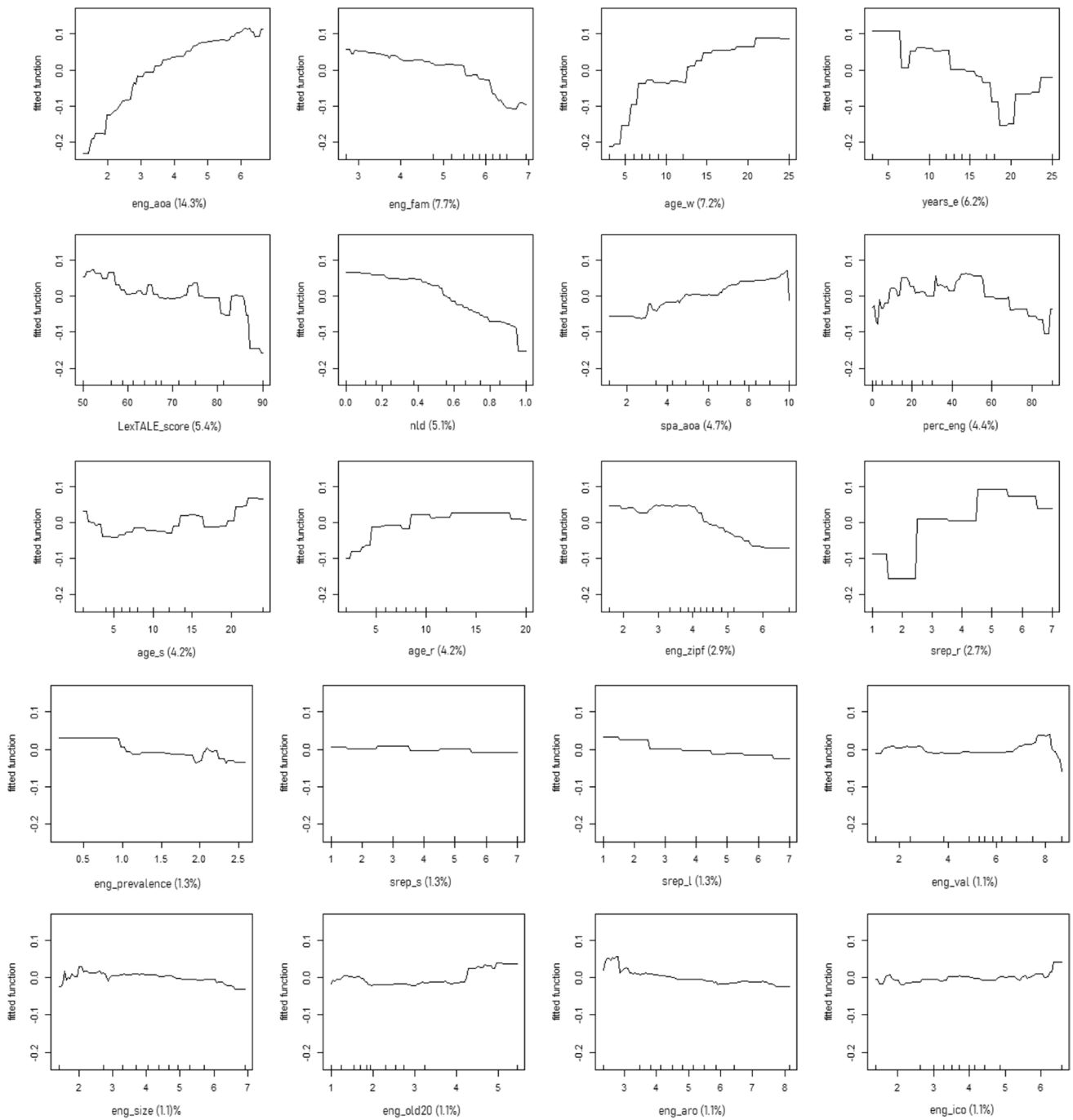


Fig. 4 Partial dependence plots for the variables included in the simplified BRTs model predicting responses to the L2 AoA ratings, ordered by decreasing RC

at high proficiency levels), perc_eng (4.4%; non-linear relationship), age_s (4.2%; non-linear relationship) and age_r (4.2%; the earlier the participants started reading in English, the earlier the L2 AoA, although this is more pronounced for early ages). The other predictors had a smaller RC (<3%, see Fig. 4 for more information).

L2 AoA and L2 reaction times

The relationship between L2 AoA values and L2 RTs in native Spanish L2 speakers of English was explored by retrieving 1089 words from Berger et al. (2019). We correlated our L2 AoA data with the L2 RT data from Berger et al. (2019). Results

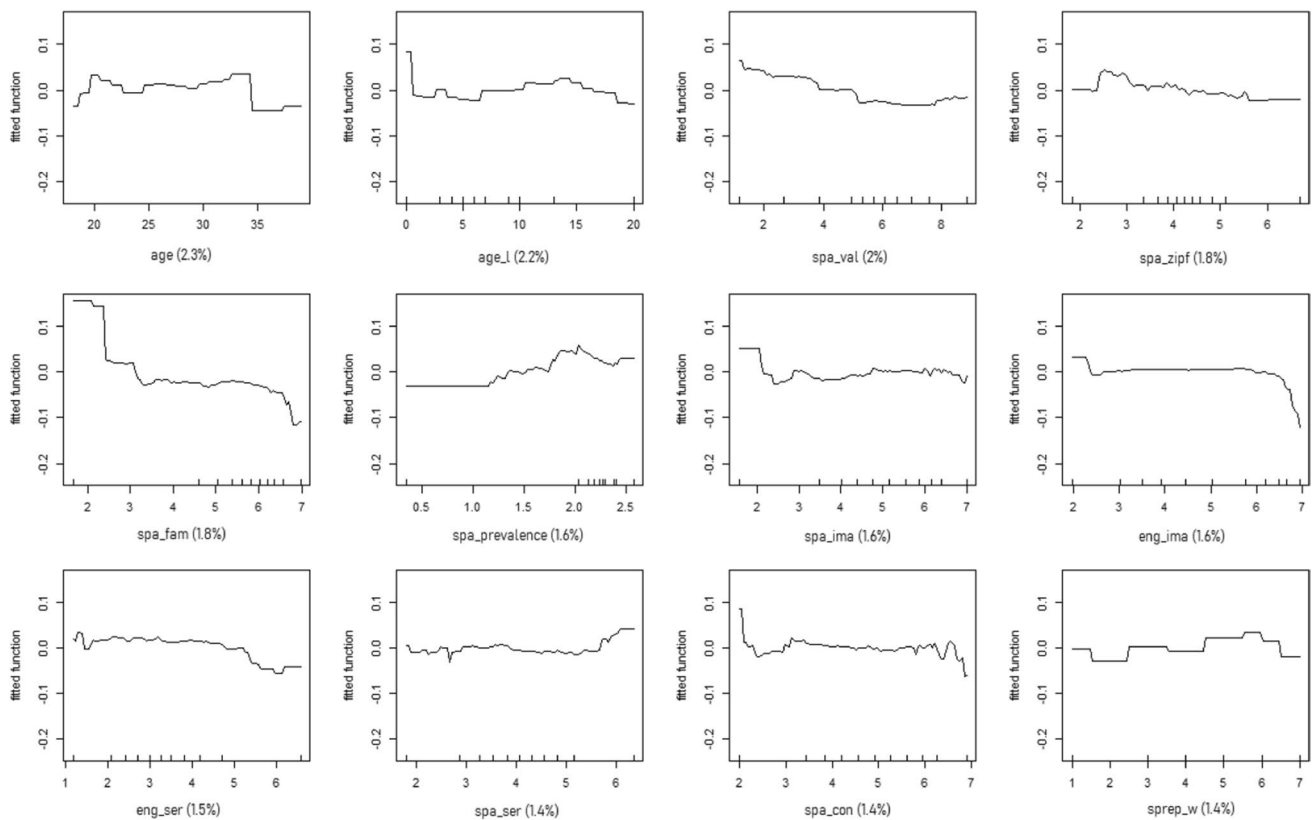


Fig. 4 (continued)

were $r = .207$, $p < .001$, which supports a positive influence of L2 AoA on L2 RTs (i.e., the earlier an L2 word is learned, the faster speakers identify it, and vice versa). In addition, we correlated L2 AoA values and L2 RTs using partial correlations that controlled for the potential influence of AoA in Spanish (values were extracted from Alonso et al., 2015 and were available for 1049 words) and showed that L2 AoA continued to be a reliable predictor, $r = .189$, $p < .001$. We also correlated the Spanish AoA data with the L2 RT data for Spanish-English speakers, controlling for L2 AoA, finding $r = -.027$, $p = .391$, showing that L1 AoA did not exert any influence on L2 RTs.

Discussion

In the current article, we present a database of subjective L2 AoA ratings for 1604 English words provided by Spanish L2 speakers of English. It provides more ecological data for psycholinguistic studies with Spanish L2 speakers of English. In addition, we wanted to examine whether, and how, several variables related, on the one hand, to the words (both lexico-semantic and affective variables), and, on the other hand, to the participants (sociolinguistic and proficiency variables), modulated the AoA ratings. For the database, we gathered the ratings from 292 participants, and we make it available to the research community. Our L2 AoA

ratings follow the pattern of previous studies, such as those of Dirix and Duyck (2017) or Wang and Chen (2020). Our results were highly correlated with these other databases, which supports the validity of our data. Specifically, high correlations were found between our ratings and those reported in other studies looking at subjective L2 AoA (Dirix & Duyck, 2017; Wang & Chen, 2020). Our ratings were also correlated, to a slightly lesser extent, with L1 AoA ratings in English (Scott et al., 2019) and Spanish (Alonso et al., 2015). Therefore, we are confident we are offering a solid resource for L2 psycholinguistic research.

Regarding the exploration of how word- and participant-related factors modulated AoA ratings, 46 predictors were included in the BRTs analyses. Our results showed that AoA in English was the factor that contributed the most to the simplified model; in other words, the age at which native English speakers acquire certain words could predict the age at which non-natives acquire those words as well. Below, we discuss some further results that may be of interest.

Following previous literature, we stated a few initial hypotheses. First, we expected to find a positive relationship between L1 (Spanish) AoA and L2 AoA scores. This was motivated by previous literature gathering L2 AoA ratings, where both Dirix and Duyck (2017) and Wang and Chen (2020) found a moderate positive correlation between the AoA of L1 and L2 words in their studies. We also found a moderate correlation with the

L1 AoA data gathered by Alonso et al. (2015), and additionally found L1 AoA to be a relevant predictor in the BRTs simplified model. Therefore, it seems that the role of L1 AoA is undeniable for L2 AoA (at least, when English is the L2), regardless of the L1 of the speakers. Also, it is remarkable that our L2 AoA data correlated more highly with other L2 AoA data (i.e., the Dutch-English and Chinese-English data by Dirix & Duyck, 2017, and Wang & Chen, 2020, respectively) than with L1 AoA data in Spanish or English. Although this would require further examination, it seems that, regarding AoA, the acquisition of words in an L2 follows a similar pattern for L2 speakers from different backgrounds, with more similarity between L2 AoAs in different languages than between the AoA in L2 and the AoA of the two languages of an L2 speaker, which in our case is Spanish and English. However, differences in the pattern of L2 AoA are also noticeable between our data and the previous work by Dirix and Duyck (2017) and Wang and Chen (2020). The data shown in the boxplots (Fig. 3) and the results from the ANOVAs illustrate two main findings: (1) Spanish natives acquire L2 words earlier, and (2) AoA scores in L2 from Spanish natives show more variability than those from Dutch and Chinese natives. To account for these differences, we would need to consider educational and social factors. In Spain, a foreign language is learned at school from the first year of compulsory education, that is, from 6 years of age (LOMLOE, 2020). English formal education for the Dutch sample is reported to start at age 13 in the case of Dirix and Duyck (2017), and Wang and Chen (2020) declared that most of their participants started learning English at 7 to 9 years of age. Therefore, at least in terms of formal education, participants have different starting points.³

³ As one reviewer suggested, we explored whether our lower AoA average is due to population differences, as we propose, or whether it could respond to the fact that we provided a top threshold (i.e., +15) in our study, whereas Dirix and Duyck (2017), and Wang and Chen (2020) did not. To examine this issue, we selected those words in our database for which participants never chose the 15+ option ($n=412$) and compared them with the corresponding words available in the Dutch database ($n=188$) and the Chinese database ($n=263$). The AoA values showed a correlation of .657 ($p<.001$) with those from Wang and Chen (2020) and a correlation of .663 ($p<.001$) with the values of Dirix and Duyck (2017). For shared words in both databases, our average L2 AoA is lower ($M=7.52$) and the dispersion in our sample is smaller ($SD=1.14$; contrary to when using the full dataset) than for the data obtained with native speakers in Dutch ($M=11.45$, $SD=1.17$). Something similar happens in comparison to the data of native speakers of Chinese ($M=7.52$, $SD=1.15$ in our sample; $M=12.29$, $SD=1.18$ for the Chinese sample). The results of two one-way ANOVAs (one comparing our data with those of Dirix & Duyck, 2017; and other comparing our data with those of Wang & Chen, 2020) showed that those differences are significant. That is, the average L2 AoA estimate made by our participants (considering only the shared words between databases) is significantly lower than that estimated by Dutch native speakers, $F(1, 374)=1087.68$, $MSE=1.334$, $p<.001$, $\eta_p^2=.74$, and by Chinese native speakers, $F(1, 524)=2193.32$, $MSE=1.362$, $p<.001$, $\eta_p^2=.81$. In sum, the pattern of results does not change substantially when eliminating the +15 words, which speaks in favour of the population differences.

Secondly, a negative relationship between word familiarity in English and L2 AoA was expected, meaning that more familiar words would be acquired earlier, in line with Wang and Chen (2020). Our data showed that familiarity of the English words was the second most relevant predictor in the simplified model, only after the AoA of the English word. Also, as can be observed in Fig. 4, this relationship is negative: the higher the familiarity, the earlier the word was acquired in the L2. Therefore, our results are in line with those from Wang and Chen (2020), even though they collected familiarity scores provided by their Chinese L2 English speakers, while we collected familiarity ratings in English provided by native speakers from the Glasgow norms (Scott et al., 2019), and not from our participants.

Our third objective was to explore whether, and how, some variables linked to the speaker (i.e., sociolinguistic variables and proficiency) could affect L2 AoA ratings. In a study where valence and arousal ratings for L2 words were collected, Imbault et al. (2021) found that participants with certain characteristics provided more native-like ratings. These features included (a) learning English at a young age, (b) for a long number of years, and (c) in an immersive context; (d) having high self-reported proficiency and (e) using English frequently in their day-to-day lives (Imbault et al., 2021). These five variables were included in our model. “Learned English at a young age” was assessed through four variables, namely age at which English was acquired for speaking, reading, writing and listening. We found that all these variables contributed to the model, with the age at which English was acquired for writing having the largest RC, followed by speaking and reading, and finally listening. “Learning English for a long number of years” would be equivalent to “number of years that they have been studying English”, which was the fourth most relevant variable in the model. It is important to note that in this case, our data show a non-linear relationship where learning for more years does not always (i.e., around 6 and 18–19 years; see Fig. 4) involve earlier AoA ratings. “Learning English in an immersive context” does not have an equivalent in our study, but it could be inferred from variables such as “learning English at home” or at an “English-speaking country”; however, these two variables were not selected for the simplified model given their low RC. “High self-reported proficiency” was assessed in this case with self-rated proficiency for speaking, reading, writing and listening, which entered the model, with self-reported proficiency for reading having the largest RC, followed by writing, and speaking and listening (although the RC for all these variables was rather small). Finally, the percentage of use of English also contributed to the model, and, in addition, it correlated positively with the speakers’ proficiency. In sum, our results support the notion that sociolinguistic variables influence L2 ratings, extending to L2 AoA what Imbault et al. (2021) found for L2 valence and arousal.

Our study also identified other variables that contribute to AoA L2 ratings. Considering that 32 predictors entered the simplified model, we would like to comment on those that had larger RCs, presenting them separately in variables related to the words (i.e., lexico-semantic and affective) and to the participants (i.e., sociolinguistic and proficiency). Regarding those linked to the words, after AoA in English and English familiarity, the variables that had a larger RC were cognate status and frequency in English (please note that a negative relationship between AoA and frequency—that is, the higher the frequency, the lower the AoA—seems to arise mostly for medium- to high-frequency scores). Our results follow the same trends as those of studies such as Tonzar et al. (2009) and Comesaña et al. (2012), who showed that cognates facilitate L2 acquisition, and of others like Citron et al. (2014), who reported that high-frequency words are also acquired earlier (frequency in Spanish also contributed to the model but to a lesser extent than frequency in English). With a smaller contribution, we found other results agreeing with previous literature (although these data should be taken with caution, as the RCs for these predictors were rather small). For instance, the contribution of valence in Spanish and valence in English (with a smaller RC value) agrees with Citron et al. (2014), Hinojosa et al. (2016a) and Warriner et al. (2013), and indicates that at young ages, people tend to prefer positive stimuli or overestimate the positive valence of stimuli, as observed by previous studies with children (see, for instance, Ponari et al., 2018; Sabater et al., 2020; Sylvester et al., 2016). We have also found an influence on our L2 AoA ratings of prevalence of Spanish words, and to a lesser extent, of English words (supporting Brysbaert et al.'s findings, 2009, 2016), imageability in Spanish and English (in line with Citron et al., 2014) and of SERs both in English and Spanish (agreeing with Hinojosa et al., 2016b). However, it cannot be said that our results are conclusive, because the relationships that we have found for these variables are not linear. Also, results are inconclusive for the rest of the lexico-semantic and affective variables entering the simplified model, either because they depict a non-linear relationship or because a particular variable enters the model for one language but not for the other (i.e., concreteness in Spanish, iconicity in English, arousal in English).

As for those factors related to the participants, the most relevant variable seems to be proficiency, as more proficient speakers assigned lower AoA scores (where this is accentuated for very proficient speakers).

Finally, we aimed at exploring the relationship between L2 AoA and L2 processing by retrieving L2 RT data from a previous independent study (i.e., Berger et al., 2019). This was motivated by Izura and Ellis's (2002) work. The authors looked at the relationship between the AoA of the speaker's first and second language, and the RTs in language

processing tasks, adding to the theoretical discussion regarding the locus of the AoA effect. They found that L2 AoA does not correspond with the order in which the equivalent word meanings were acquired in the L1. Following Izura and Ellis (2002), if L2 AoA depended on L1 AoA (i.e., in the process of acquiring a word in the L2, this new word form would be associated with an old semantic representation), the source of the AoA effects would lie in the semantic system, so the age at which the concept was firstly acquired (that is, in the L1) would determine RTs in the L2 too. In practical terms, the AoA of the speakers' L1 (in this case, Spanish) should positively correlate with L2 RTs. However, if the locus is not in the semantic system and were to lie in the mapping between lexical and semantic representations (as predicted by the *mapping hypothesis*), L2 AoA, but not L1 AoA, should positively correlate with L2 RTs. The latter was found by Izura and Ellis (2002), concluding that the L2 AoA effect mirrors the order in which the words were acquired in the L2, rather than the order in which the translation equivalents were acquired in the native language. We have been able to replicate Izura and Ellis' (2002) findings with a larger dataset, adding to the theoretical discussion of the locus of the AoA effect by advocating in favour of the *mapping hypothesis* and against a semantic locus for this effect.

At this point, some remarks are needed. First of all, this is the first time, to our knowledge, in which BRTs are used to explore word-related and participant-related data on AoA. As pointed out above, BRTs enable one to explore the relationship between variables without prior assumptions, allowing researchers to establish patterns in their data that might not have been explored if following other types of analysis. Also, previous research tends to consider only a few of these variables as predictors, but we included 32 predictors in our simplified model (considering data in English and Spanish). On top of that, what we know about the influence of these lexico-semantic and affective variables for AoA is limited to AoA ratings provided by native speakers, so the hypotheses that can be derived using these norms are limited as well. Lastly, it must be kept in mind that the lexico-semantic and affective values that we have employed in the analyses carried out here have been normed by L1 speakers. Therefore, in order to clearly elucidate which lexico-semantic and affective variables influence L2 AoA, it would be beneficial to use ratings normed by L2 speakers, who ideally share a native language.

Finally, we would like to highlight the relevance of including variables related to the words (lexico-semantic and affective) and to the speakers (sociolinguistic and proficiency) in this sort of research. The results of our model show that a combination of these factors is key to explaining L2 AoA, where if we were to overlook, for instance, the role of the variables related to the speakers, we would have lost valuable information. Therefore, in our quest to make

L2 research more ecologically valid, we need to include not only ratings performed by the type of speaker that we are going to investigate (that is, ratings by and for L2 speakers), but also information about sociolinguistic factors that indeed shape L2 acquisition.

Conclusions

In the current article, we provide the research community with an L2 AoA database by and for Spanish-English speakers. We also carefully examined the relationship between lexico-semantic, affective, sociolinguistic and proficiency variables and L2 AoA, finding that a combination of these variables is needed to obtain comprehensive models of L2 AoA. Our results also support the *mapping hypothesis* regarding the locus of the AoA effects. We hope that our contributions prove useful for researchers involved in psycholinguistic research with Spanish L2 speakers of English and can ultimately be informative for second language teaching or bilingual education programmes.

Acknowledgments We would like to thank all the participants who took part in our study. We thank Lynn K. Perry and Bodo Winter for their assistance in accessing iconicity ratings in English. We would also like to thank Mike Cortese and Nicolas Dirix for their insightful reviews. We are very grateful to Cynthia Berger, Scott Crossley and Stephen Skalicky for sharing their L2 RT data with us.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by the “Proyectos de I+D de Jóvenes Doctores” funded by the Autonomous University of Madrid and the Community of Madrid (Title: “The acquisition and development of affective vocabulary in a second language from childhood to adolescence”; reference SI3/PJI/2021-00249) awarded to Sara Rodríguez-Cuadrado; by the Ministerio de Ciencia e Innovación under Grants PGC2018-098558-B-I00 to José A. Hinojosa, and PID2019-107206GB-I00 and RED2018-102615-T to Pilar Ferré, PID2019-108092GA-I00/AEI/10.13039/501100011033 to Carlos Romero-Rivas, and PID2019-106868GB-I00 to Paz Suárez-Coalla; by Comunidad de Madrid under Grant H2019/HUM-5705 to José A. Hinojosa, and Universitat Rovira i Virgili under Grant 2019PFR-URV-B2-32 to Pilar Ferré. Lucía Sabater was hired by the aforementioned grant PGC2018-098558-B-I00 awarded to José A. Hinojosa.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval The study was approved by the Ethics Committee of the Principality of Asturias (reference 153/19) and was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki.

Consent to participate Consent to participate was requested and granted for all participants. Participants were informed about the purpose of the study, its procedure and length, the voluntary nature of their participation and the possibility of deleting their data after participation

upon request, the benefits of their participation, the lack of physical or moral risks associated to the task, and finally, confidentiality details.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguasvivas, J. A., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E., & Duñabeitia, J. A. (2018). SPALEX: A Spanish lexical decision database from a massive online data collection. *Frontiers in Psychology, 9*, 2156. <https://doi.org/10.3389/fpsyg.2018.02156>
- Alonso, M. A., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods, 47*(1), 268–274. <https://doi.org/10.3758/s13428-014-0454-2>
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445–459.
- Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers, 34*(3), 424–434.
- Berger, C., Crossley, S., & Skalicky, S. (2019). Using lexical features to investigate second language lexical decision performance. *Studies in Second Language Acquisition, 41*(5), 911–935. <https://doi.org/10.1017/S0272263119000019>
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 441. <https://doi.org/10.1037/xhp0000159>
- Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods, 51*(2), 467–479. <https://doi.org/10.3758/s13428-018-1077-9>
- Cameirao, M. L., & Vicente, S. G. (2010). Age-of-acquisition norms for a set of 1,749 Portuguese words. *Behavior Research Methods, 42*(2), 474–480. <https://doi.org/10.3758/BRM.42.2.474>
- Caselli, N. K., & Pyers, J. E. (2017). The road to language learning is not entirely iconic: Iconicity, neighborhood density, and frequency facilitate acquisition of sign language. *Psychological Science, 28*(7), 979–987. <https://doi.org/10.1177/0956797617700498>
- Chalard, M., Bonin, P., Méot, A., Boyer, B., & Fayol, M. (2003). Objective age-of-acquisition (AoA) norms for a set of 230 object names in French: Relationships with psycholinguistic variables, the English data from Morrison et al. (1997), and

- naming latencies. *European Journal of Cognitive Psychology*, 15(2), 209–245. <https://doi.org/10.1080/09541440244000076>
- Chen, X., & Dong, Y. (2019). Evaluating objective and subjective frequency measures in L2 lexical processing. *Lingua*, 230, 102738. <https://doi.org/10.1016/j.lingua.2019.102738>
- Citron, F. M., Weekes, B. S., & Ferstl, E. C. (2014). How are affective word ratings related to lexicosemantic properties? Evidence from the Sussex Affective Word List. *Applied PsychoLinguistics*, 35(2), 313–331. <https://doi.org/10.1017/S0142716412000409>
- Comesaña, M., Soares, A. P., Sánchez-Casas, R., & Lima, C. (2012). Lexical and semantic representations in the acquisition of L2 cognate and non-cognate words: Evidence from two learning methods in children. *British Journal of Psychology*, 103(3), 378–392. <https://doi.org/10.1111/j.2044-8295.2011.02080.x>
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, 60(8), 1072–1082. <https://doi.org/10.1080/17470210701315467>
- Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, 66(5), 946–972. <https://doi.org/10.1080/17470218.2012.722660>
- Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283. <https://doi.org/10.1037/T0278-7393.26.5.1283>
- Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, 31(4), 650–658.
- Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & Cognition*, 41(2), 297–311. <https://doi.org/10.3758/s13421-012-0263-8>
- Degner, J., Doycheva, C., & Wentura, D. (2011). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition*, 15, 181–189. <https://doi.org/10.1017/S1366728911000095>
- Díez-Álamo, A. M., Díez, E., Wojcik, D. Z., Alonso, M. A., & Fernandez, A. (2019). Sensory experience ratings for 5,500 Spanish words. *Behavior Research Methods*, 51(3), 1205–1215. <https://doi.org/10.3758/s13428-018-1057-0>
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science*, 14(1), 14–18. <https://doi.org/10.1111/1467-9280.t01-1-01412>
- Dirix, N., & Duyck, W. (2017). The first- and second-language age of acquisition effect in first- and second-language book reading. *Journal of Memory and Language*, 97, 103–120. <https://doi.org/10.1016/j.jml.2017.07.012>
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45(4), 1246–1258. <https://doi.org/10.3758/s13428-013-0326-1>
- Eliith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, 40(4), 1049–1054. <https://doi.org/10.3758/BRM.40.4.1049>
- Ferré, P., Guasch, M., Moldovan, C., & Sánchez-Casas, R. (2012). Affective norms for 380 Spanish words belonging to three different semantic categories. *Behavior Research Methods*, 44(2), 395–403. <https://doi.org/10.3758/s13428-011-0165-x>
- Ferré, P., Sánchez-Casas, R., Comesaña, M., & Demestre, J. (2017). Masked translation priming with cognates and noncognates: Is there an effect of words? concreteness? *Bilingualism: Language and Cognition*, 20, 770–782. <https://doi.org/10.1017/S1366728916000262>
- Ferré, P., Anglada-Tort, M., & Guasch, M. (2018). Processing of emotional words in bilinguals: Testing the effects of word concreteness, task type and language status. *Second Language Research*, 34(3), 371–394. <https://doi.org/10.1177/0267658317744008>
- Ferré, P., Guasch, M., Hinojosa, J. A., Stadthagen-González, H., Fraga, I., Pérez-Sánchez, M. A. (2022a). The contribution of the components of emotional experience to the emotional prototypicality of words (in preparation).
- Ferré, P., Guasch, M., Stadthagen-Gonzalez, H., & Comesaña, M. (2022b). Love me in L1, but hate me in L2: How native speakers and bilinguals rate the affectivity of words when feeling or thinking about them. *Bilingualism: Language and Cognition*, 1–15. <https://doi.org/10.1017/S1366728922000189>
- Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, 45(3), 468–492. <https://doi.org/10.1006/jmla.2000.2784>
- Garrido, M. V., & Prada, M. (2021). Comparing the valence, emotionality and subjective familiarity of words in a first and a second language. *International Journal of Bilingual Education and Bilingualism*, 24(2), 275–291. <https://doi.org/10.1080/13670050.2018.1456514>
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2), 228–244. [https://doi.org/10.1016/0749-596X\(90\)90074-A](https://doi.org/10.1016/0749-596X(90)90074-A)
- Guasch, M., Sánchez-Casas, R., Ferré, P., & García-Albea, J. E. (2008). Translation performance of Beginning, Intermediate and Proficient Spanish-Catalan Bilinguals: Effects of Form and Semantic Relations. *The Mental Lexicon*, 3, 289–308. <https://doi.org/10.1075/ml.3.3.03gua>
- Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, 45(3), 765–771. <https://doi.org/10.3758/s13428-012-0296-8>
- Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48(4), 1358–1369. <https://doi.org/10.3758/s13428-015-0684-y>
- Hasegawa, Y. (2010). Context effects on EFL vocabulary learning: Focusing on word imageability and test formats. *JLTA Journal Kiyō*, 13, 145–161. https://doi.org/10.20622/jltaj.13.0_145
- Hinojosa, J. A., Méndez-Bértolo, C., & Pozo, M. A. (2010). Looking at emotional words is not the same as reading emotional words: Behavioral and neural correlates. *Psychophysiology*, 47(4), 748–757. <https://doi.org/10.1111/j.1469-8986.2010.00982.x>
- Hinojosa, J. A., Martínez-García, N., Villalba-García, C., Fernández-Folgueiras, U., Sánchez-Carmona, A., Pozo, M. A., & Montoro, P. R. (2016a). Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behavior Research Methods*, 48(1), 272–284. <https://doi.org/10.3758/s13428-015-0572-5>
- Hinojosa, J. A., Rincón-Pérez, I., Romero-Ferreiro, M. V., Martínez-García, N., Villalba-García, C., Montoro, P. R., & Pozo, M. A. (2016b). The Madrid Affective Database for Spanish (MADS): Ratings of dominance, familiarity, subjective age of acquisition and sensory experience. *PLoS One*, 11(5), e0155866. <https://doi.org/10.1371/journal.pone.0155866>

- Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2020). Affective neuro-linguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience*, 35(7), 813–839. <https://doi.org/10.1080/23273798.2019.1620957>
- Hinojosa, J. A., Haro, J., Magallares, S., Duñabeitia, J. A., & Ferré, P. (2021). Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods*, 1–14. <https://doi.org/10.3758/s13428-020-01496-z>
- Hubers, F., Cucchiari, C., & Strik, H. (2020). Second language learner intuitions of idiom properties: What do they tell us about L2 idiom knowledge and acquisition? *Lingua*, 246, 102940. <https://doi.org/10.1016/j.lingua.2020.102940>
- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 369(1651), 20130298. <https://doi.org/10.1098/rstb.2013.0298>
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54–65. <https://doi.org/10.1016/j.cognition.2008.07.015>
- 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>
- Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica*, 23(2).
- Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, 13(7–8), 789–845. <https://doi.org/10.1080/13506280544000066>
- Juhász, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, 131(5), 684. <https://doi.org/10.1037/0033-2909.131.5.684>
- Juhász, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1312. <https://doi.org/10.1037/0278-7393.29.6.1312>
- Juhász, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, 13(7–8), 846–863.
- Juhász, B. J., & Yap, M. J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods*, 45(1), 160–168. <https://doi.org/10.3758/s13428-012-0242-9>
- Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, 35(3), 575–586. <https://doi.org/10.1111/j.1551-6709.2010.01169.x>
- 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>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990. <https://doi.org/10.3758/s13428-012-0210-4>
- 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>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Ley Orgánica 2/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de educación. *Boletín Oficial del Estado*, núm. 340, de 30 de diciembre de 2020, pp. 122868 a 122953. <https://www.boe.es/eli/es/lo/2020/12/29/3>. Accessed 27 Apr 2011.
- Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, 39(2), 192–198.
- Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. *PLoS One*, 6(1), e16505. <https://doi.org/10.1371/journal.pone.0016505>
- Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Anđelković, D., ... & Ünal-Logacev, Ö. (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words?. *Behavior Research Methods*, 48(3), 1154–1177. <https://doi.org/10.3758/s13428-015-0636-6>
- Łuniewska, M., Wodniecka, Z., Miller, C. A., Smolík, F., Butcher, M., Chondrogianni, V., ... & Haman, E. (2019). Age of acquisition of 299 words in seven languages: American English, Czech, Gaelic, Lebanese Arabic, Malay, Persian and Western Armenian. *PLoS one*, 14(8), e0220611. <https://doi.org/10.1371/journal.pone.0220611>
- Mañas Antón, O. (2019). Datos y cifras de la educación 2019- 2020. Consejería de Educación y Juventud. <http://www.madrid.org/bvirtual/BVCM050013.pdf>. Accessed 27 Apr 2011.
- Massaro, D. W., & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers in Communication*, 2, 4. <https://doi.org/10.3389/fcomm.2017.00004>
- Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A. L., ... & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, 45(1), 169–177.
- Morris, P. E. (1981). Age of acquisition, imagery, recall, and the limitations of multiple-regression analysis. *Memory & Cognition*, 9, 277–282.
- Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology Section A*, 50(3), 528–559. <https://doi.org/10.1080/027249897392017>
- Peeters, D. (2016). Processing consequences of onomatopoeic iconicity in spoken language comprehension. In: *38th Annual Meeting of the Cognitive Science Society (CogSci 2016)* (pp. 1632-1647). Cognitive Science Society.
- Perani, D., & Abutaleb, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, 15(2), 202–206. <https://doi.org/10.1016/j.conb.2005.03.007>
- Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 369(1651), 20130300. <https://doi.org/10.1098/rstb.2013.0300>
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS One*, 10(9), e0137147. <https://doi.org/10.1371/journal.pone.0137147>
- Piñero, A., & Manzano, M. (2000). A lexical database for Spanish-speaking children. *Behavior Research Methods, Instruments, & Computers*, 32(4), 616–628.
- Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2), e12549. <https://doi.org/10.1111/desc.12549>
- Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.9. <https://CRAN.R-project.org/package=psych>. Accessed 28 May 2011.

- Sabater, L., Guasch, M., Ferré, P., Fraga, I., & Hinojosa, J. A. (2020). Spanish affective normative data for 1,406 words rated by children and adolescents (SANDchild). *Behavior Research Methods*, 52(5), 1939–1950. <https://doi.org/10.3758/s13428-020-01377-5>
- Schröder, A., Gemballa, T., Ruppig, S., & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods*, 44(2), 380–394. <https://doi.org/10.3758/s13428-011-0164-y>
- 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.biopsycho.2008.03.010>
- Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, 51(3), 1258–1270. <https://doi.org/10.3758/s13428-018-1099-3>
- Sereno, S. C., & O'Donnell, P. J. (2009). Participant and word gender in age of acquisition effects: The role of gender socialization. *Sex Roles*, 61(7), 510–518. <https://doi.org/10.1007/s11199-009-9649-x>
- Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *Quarterly Journal of Experimental Psychology*, 62, 1115–1122. <https://doi.org/10.1080/17470210802618900>
- Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition*, 12(1), 164–181. <https://doi.org/10.1017/langcog.2019.36>
- Sidhu, D., Williamson, J., Slavova, V., & Pexman, P. M. (2021). An investigation of iconic language development in four datasets. *Journal of Child Language*. <https://doi.org/10.1017/S0305000921000040>
- Stadthagen-Gonzalez, H., Imbault, C., Sánchez, M. A. P., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 Spanish words. *Behavior Research Methods*, 49(1), 111–123. <https://doi.org/10.3758/s13428-015-0700-2>
- Sylvester, T., Braun, M., Schmidtke, D., & Jacobs, A. M. (2016). The Berlin affective word list for children (kidBAWL): exploring processing of affective lexical semantics in the visual and auditory modalities. *Frontiers in Psychology*, 7, 969. <https://doi.org/10.3389/fpsyg.2016.00969>
- Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological Science*, 23(12), 1443–1448. <https://doi.org/10.1177/0956797612459763>
- Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, 59(3), 623–646. <https://doi.org/10.1111/j.1467-9922.2009.00519.x>
- Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3), 458–483. <https://doi.org/10.1006/jmla.1998.2584>
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Vélez-Urbe, I., & Rosselli, M. (2019). The auditory and visual appraisal of emotion-related words in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 22(1), 30–46. <https://doi.org/10.1017/S1366728917000517>
- Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, 40(4), 1079–1087. <https://doi.org/10.3758/BRM.40.4.1079>
- Walley, A. C., & Metsala, J. L. (1992). Young children's age-of-acquisition estimates for spoken words. *Memory & Cognition*, 20(2), 171–182.
- Wang, J., & Chen, B. (2020). A database of Chinese-English bilingual speakers: ratings of the age of acquisition and familiarity. *Frontiers in Psychology*, 11, 554785. <https://doi.org/10.3389/fpsyg.2020.554785>
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. <https://doi.org/10.3758/s13428-012-0314-x>
- Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging*, 32(2), 158–177. <https://doi.org/10.1037/pag0000151>
- Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic?: Iconicity in English sensory words. *Interaction Studies*, 18(3), 443–464. <https://doi.org/10.1075/is.18.3.07win>
- Winter, B., Perlman, M., Perry, L. K., Dingemans, M., & Lupyan, G. (2022). Iconicity ratings for 14000+ English words (submitted).
- Yao, B., Vasiljevic, M., Weick, M., Sereno, M. E., O'Donnell, P. J., & Sereno, S. C. (2013). Semantic size of abstract concepts: It gets emotional when you can't see it. *PLoS One*, 8(9), e75000. <https://doi.org/10.1371/journal.pone.0075000>
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5), 971–979. <https://doi.org/10.3758/PBR.15.5.971>

Open practices statement The data and materials for the experiment and the raw data are available at https://osf.io/gr8xd/?view_only=73b01dccb4d7897c8d104d3d68c46. The experiment was not preregistered.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.