

Rated age-of-acquisition norms for over 3,200 German words

Julia M. H. Birchenough¹ · Robert Davies² · Vincent Connelly¹

© Psychonomic Society, Inc. 2016

Abstract Words that have been learned early in life are responded to faster than words that have been acquired later. Subjective ratings of acquisition ages have been successfully employed to study the effect of age of acquisition (AoA). Although a large number of norms exist in many languages, fewer are available for German. Therefore, subjective AoA ratings for 3,259 German words were collected online, including 2,363 nouns and 473 verbs. These words were presented in lists of 140 words, and participants rated the age in years at which they had first learned each word. A split-half correlation testified to a high internal reliability. There were also high correlations with rated AoA values for subsets of the items that had been collected in previous studies, in both German and English. Age and gender were found to influence the ratings very weakly, in that older and male participants tended to give slightly higher age ratings. Education, multilingualism, and frequent usage of languages other than German did not exert an influence on the rating values. These new ratings will extend the currently existing norms available for language and reading research across languages and will provide researchers with a wider choice of word stimuli. The ratings

are available expressed in two measurements: age in years, and AoA rated on a 7-point Likert scale.

Keywords Age-of-acquisition · AoA · Ratings · Norms · Estimates

Age of acquisition (AoA) denotes the age at which a word and its meaning is first learned (Carroll & White, 1973). Evidence is accumulating that AoA exerts an influence on cognitive processes. Early-acquired words are processed faster than words that are acquired later in life in pictured object naming (Holmes & Ellis, 2006; Morrison, Ellis, & Quinlan, 1992), in reading aloud (Cortese & Khanna, 2007; Cortese & Schrock, 2013), in lexical decision (Cortese & Khanna, 2007), in natural reading paradigms employing eyetracking (Juhasz & Rayner, 2006), and in semantic categorization tasks (Brysbaert, Van Wijnendaele, & De Deyne, 2000). Moreover, in aphasic patients, earlier-acquired words seem to be more accessible for word production than are late-acquired words (De Bleser & Kauschke, 2003; Weekes, Davies, Wadey, & Bradley, 2004; for a recent review, see also Brysbaert & Ellis, 2015). In Alzheimer patients, words that are learned earlier in life have also been found to remain intact for a longer time for word production (Cuetos, González-Nosti, & Martínez, 2005) and word recognition (Cuetos, Herrera, & Ellis, 2010). AoA is now a widely accepted effect evident in lexical processing (Juhasz, 2005) and in learning mechanisms more generally (Catling, Dent, Preece, & Johnston, 2013; Stewart & Ellis, 2008). AoA is thus critical to the investigation of language behavior and needs to be taken into account in future language research.

AoA has been determined by averaging participants' subjective estimates of when they first learned a word. Such subjective AoA ratings have been collected in many languages,

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0718-0) contains supplementary material, which is available to authorized users.

✉ Julia M. H. Birchenough
12085385@brookes.ac.uk

¹ Department of Psychology, Social Work and Public Health, Oxford Brookes University, Headington Campus, Gypsy Lane, Oxford OX3 0BP, UK

² Department of Psychology, Lancaster University, Lancaster, UK

including very extensive ratings collections for 30,000 words in English (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) and in Dutch (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014). Other studies have presented smaller sets of ratings, most of these for English (Bird, Franklin, & Howard, 2001; Carroll & White, 1973; Clark & Paivio, 2004; Cortese & Khanna, 2008; Gilhooly & Logie, 1980; Stadthagen-Gonzalez & Davis, 2006). However, AoA rating estimates have also been collected in other languages, including French (Ferrand et al., 2008), Portuguese (Cameirão & Vicente, 2010; Marques, Fonseca, Morais, & Pinto, 2007), Spanish (Cuetos, Samartino, & Ellis, 2012; Moreno-Martínez, Montoro, & Rodríguez-Rojo, 2014), and Italian (Barca, Burani, & Arduino, 2002; Della Rosa, Catricalà, Vigliocco, & Cappa, 2010). In addition, the Snodgrass and Vanderwart (1980) pictures of 260 objects have been rated for AoA, and at times adjusted and extended with further drawings, in French (Alario & Ferrand, 1999), Icelandic (Pind, Jónsdóttir, Gissurardóttir, & Jónsson, 2000), Spanish (Cuetos, Ellis, & Álvarez, 1999), and German (Schröder, Kauschke, & De Bleser, 2003). Finally, a selection of 824 rated AoA estimates is available in German (Schröder, Gemballa, Ruppert, & Wartenburger, 2012). The present AoA ratings extend the ratings available for German considerably, and will thereby aid word stimulus selection for future language research.

Early-acquired words tend to be shorter in length, more frequent, more imageable, more concrete and have more spelling neighbors than words that are learned later in life (Morrison, Chappell, & Ellis, 1997; Stadthagen-Gonzalez & Davis, 2006). The high intercorrelation with frequency has made it particularly difficult to distinguish frequency from AoA effects (Brysbaert & Ghyselinck, 2006; Zevin & Seidenberg, 2002).

The speeded naming task (reading aloud) requires participants to read words aloud as quickly and accurately as possible, thus tapping into processes underlying the transformation of the written word into its spoken form. Morrison and Ellis (1995) reported faster word naming for early-acquired than for late-acquired words (AoA effect) when frequency was taken into account, but they did not find a processing advantage for more frequent words (frequency effect) when AoA was controlled for. They suggested that previously reported frequency effects were actually due to AoA. However, other studies have reported effects of both frequency and AoA effects in word production (Brown & Watson, 1987; Cortese & Khanna, 2007) and thus support the existence of independent AoA and frequency effects in reading aloud.

Studies examining the AoA effect in several tasks have indicated a stronger AoA effect in lexical decision than in word naming (Cortese & Khanna, 2007; Cortese & Schrock, 2013). In the lexical decision task, participants are asked to discriminate between words and non-words and it is assumed

that word meaning is thereby accessed (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Chumbley & Balota, 1984). Although early studies failed to find an AoA effect in the lexical decision task (see Gilhooly & Watson, 1981), subsequent investigations have reported independent effects of AoA and frequency (Cortese & Khanna, 2007; Ghyselinck, Lewis, & Brysbaert, 2004), as well as an interaction such that the AoA effect was stronger for low-frequency than for high-frequency words (Bonin, Chalard, Méot, & Fayol, 2001; Gerhand & Barry, 1999). Importantly, Brysbaert and Cortese (2011) showed that the AoA effect remained present in both word naming and lexical decision when employing frequency norms based on larger text corpora.

In comparison, the AoA effect seems to be stronger in picture naming than in word naming (Lambon Ralph & Ehsan, 2006), and the AoA effect has been reliably shown in picture naming such that pictures of early-acquired items are named faster than pictures of later-acquired items (Holmes & Ellis, 2006). Conflicting evidence has been reported with regard to the role of frequency when AoA has been taken into account in picture naming. Barry, Morrison, and Ellis (1997) found both AoA and frequency effects in picture naming that interacted in such a way that late-acquired words produced a larger frequency effect than early-acquired words. In contrast, other studies concluded that frequency produced only a small effect (Carroll & White, 1973), or no effect (Chalard, Bonin, Méot, Boyer, & Fayol, 2003; Morrison et al., 1992).

The effect of AoA has been found to be larger in size and more clearly distinct from frequency effects in tasks that require a mapping between language and semantics (but see, e.g., Chalard & Bonin, 2006). The semantic locus hypothesis (Brysbaert et al., 2000; van Loon-Vervoor, 1989) proposes that AoA effects in lexical processing occur because the semantic system is organized according to when meanings are acquired. Early-acquired words thus have an advantage and produce faster meaning associations and faster categorization than later-acquired words. This is consistent with the model of semantic growth presented by Steyvers and Tenenbaum (2005) in which AoA functions as an organizing factor of the semantic network. Earlier-acquired word meanings have more connections to other word meanings than later learned ones, and are consequently better connected. AoA effects can then be understood to occur due to enhanced semantic connectivity and reflect the underlying semantic network.

The network plasticity hypothesis (Ellis & Lambon Ralph, 2000) attributes AoA effects to the diminution of network plasticity during the course of development. Ellis and Lambon Ralph (2000) presented a series of simulation studies, in which a connectionist system was presented with patterns for learning in a cumulative and interleaved fashion. The simulations reported by Ellis and Lambon Ralph, and later by P. Monaghan and Ellis (2010), demonstrated that as experience

of input patterns accumulated and connection weights were adapted, earlier introduced items tended to shape the network in their favor. The adaptation of connections was found to be driven by both the frequency of occurrence of input patterns but also, critically, by the order of entry of such patterns such that later-acquired items were learned less securely than earlier-acquired items.

However, in a review of the computational evidence for AoA effects in reading development, Zevin and Seidenberg (2002) identified the arbitrariness of the input–output patterns deployed in the Ellis and Lambon Ralph (2000) simulations as critical to the observation of AoA effects over and above the impact of pattern frequency. In the Zevin and Seidenberg account (arbitrary mapping hypothesis), an enduring independent AoA effect is less likely to be observed if input or output patterns have the rich structure possessed by the orthography or phonology of English words, and if input–output mappings are predictable as in the pronunciation of regular or consistent English word spellings. In this analysis, strong AoA effects were observed by Ellis and Lambon Ralph because input–output patterns were random and input–output mappings were arbitrary. Zevin and Seidenberg argued that their simulations demonstrated that an AoA effect independent of frequency is not observed even if spelling–sound mappings are inconsistent. However, in a series of behavioral studies, J. Monaghan and Ellis (2002) showed that the AoA effect in naming was greater for words with inconsistent than with consistent spelling–sound mappings. P. Monaghan and Ellis demonstrated in further simulations that AoA effects emerge as an order effect when reading development is simulated using input–output patterns and training regimes that more closely approximate those experienced in human development. Importantly, the items with less consistent spelling–sound mappings incurred greater AoA effects. Thus, AoA effects appear to be observable independent of frequency but are conditioned by spelling–sound consistency.

The finding of larger AoA effects for words in English with inconsistent spelling–sound mappings predicts smaller AoA effects in languages in which mappings are more consistent (Davies, Wilson, Cuetos, & Burani, 2014; Ellis & Lambon Ralph, 2000). However, word naming studies in transparent languages have reported mixed results. AoA effects have been reported in transparent orthographies such as Turkish (Raman, 2006) and Spanish (Cuetos & Barbón, 2006). Interestingly, Wilson, Ellis, and Burani (2012) showed that AoA affects reading aloud of Italian words with irregular stress (arbitrary), but not words with regular stress (consistent mapping) in an otherwise very consistent orthography. However, in a naming study in Spanish, AoA showed an effect when words were highly imageable (Davies et al., 2014). These findings suggest some form of semantic involvement is required in reading for an AoA effect to occur (Davies, Barbón, & Cuetos, 2013; Davies et al., 2014; Wilson, Cuetos, Davies, & Burani,

2013). AoA effects have been observed in spelling systems varying in consistency. It is evident that AoA is a live area of enquiry and that more AoA norms are needed to investigate the role of AoA in languages that differ in the consistency of spelling–sound correspondences. This is a further motivation for collecting AoA ratings for German words. German is considered a transparent language, and exception words tend to be loan words from other languages or proper names of people and locations (Ziegler, Perry, & Coltheart, 2000). Within a comparison of several European languages, German was considered to be a more consistent language in terms of grapheme–phoneme correspondences (Seymour, Aro, & Erskine, 2003). One notable inconsistency in German is found in vowel duration (e.g., Landerl & Reitsma, 2005).

As we discussed above, studies have successfully employed rated AoA norms in accounting for variance in performance in various tasks. Alternatively, AoA has also been determined by children’s picture naming ages, which denote the age at which a child reliably names a pictured object, in English (Morrison et al., 1997), French (Chalard et al., 2003), Spanish (Álvarez & Cuetos, 2007), and Italian (Lotto, Surian, & Job, 2010). Whereas picture-naming ages are considered to be more objective estimates, they are also more difficult to obtain for nonpicturable words. AoA has therefore been collected as subjective ratings, which has facilitated the determination of AoA values for much larger numbers of words. Morrison et al. (1997) compared rated AoA with children’s picture naming ages and concluded that rated AoA was a good estimate of objective AoA, but that rated AoA differed from objective measures in reduced reliance on imageability and was more influenced by word length, word familiarity and word frequency when making a judgment. These findings are consistent with comparisons of subjective and objective AoA norms in Icelandic (Pind et al., 2000), French (Chalard et al., 2003), and Spanish (Álvarez & Cuetos, 2007). In a further study in German, Schröder et al. (2003) compared rated AoA norms with children’s picture-naming ages and spontaneous first speech production ages as reported by parents. The results showed that children included words in their spontaneous production vocabulary at an earlier age than they were able to use the same words to name pictures correctly. Schröder and colleagues found that the rated acquisition ages were more similar to children’s picture-naming ages than to their spontaneous speech production. Comparing rated and objective AoA as predictors of picture-naming latencies, Chalard et al. found that although both rated and objective AoA were predictive of latencies, objective AoA accounted for more variance than did rated AoA. Morrison and Ellis (2000) also replicated both AoA and frequency effects in naming and lexical-decision tasks using objective AoA measures. These results suggest that rated AoA

is an adequate measure to investigate age-of-acquisition effects in lexical processing, if objective measures are not available.

Rather than regarding AoA ratings as capturing the veridical age at which words are learned, researchers have tended to conceptualize rated AoA as reflecting the order of acquisition. Kuperman et al. (2012) compared their AoA ratings with an objective measure that targeted the knowledge of word meanings in children (Biemiller & Slonim, 2001). Kuperman and colleagues found that raters tended to overestimate at what age they had learned very early-acquired words, specifically words learned in the first 3 years of life, and highlighted that this is congruent with infantile amnesia. Kuperman et al. also observed a tendency for raters to consider very few words to be learned after the age of 14/15. These observations mirror previous reports of ratings underestimating the AoA of later-learned words (Morrison et al., 1997) and subsequent suggestions that AoA ratings should be considered to reflect rank order rather than the actual ages when words are learned. In this light, recent research on AoA has tended to refer to the order of acquisition rather than the age of acquisition (Joseph, Wonnacott, Forbes, & Nation, 2014; Kuperman et al., 2012; P. Monaghan & Ellis, 2010).

Whether AoA ratings can be taken to be age estimates or order estimates, how might AoA values be influenced by participant characteristics? Previous rating studies have examined the impacts of age, gender, and education on ratings values. Kuperman et al. (2012) reported that older participants tended to rate words as slightly later-learned than younger participants. Ratings by older participants were more predictive of Alzheimer patients' lexical selection times than were ratings from young participants (Cuetos et al., 2012). In contrast, Schröder et al. (2012) compared their German ratings from their older participants with those from their younger group and did not find an age-related difference. Mixed results have also been found with regard to gender: Whereas in Kuperman's sample, female respondents tended to rate words as being later-learned than did male respondents, the reverse was found in other studies conducted in German (Schröder et al., 2003) and English (Winters, Winter, & Burger, 1978). With regard to the influence of education level, Kuperman and colleagues did not find a main effect of education, but they noted that respondents with higher education tended to give earlier word-learning ages than did participants with lower education, although the effect size was very small. Multilingualism was also assessed by Kuperman et al., but few participants had been raised with more than one language, and their ratings did not differ from the rest. It thus appears that AoA ratings weakly correlate with age and may be influenced—with varying directions—by gender, but seemingly not by education or multilingualism. With the aim to ensure the utmost comparability to other ratings, in this study we collected rater characteristics, and their relative influences on the rating values were determined.

In the following sections we will present our AoA ratings for 3,259 German words. To pave the way for more extensive research across languages, the ratings collection was conducted using a methodological approach that resembled as closely as possible the one that Kuperman et al. (2012) had used to collect the currently largest AoA dataset in English: 30,000 ratings using the Amazon Mechanical Turk (<https://www.mturk.com/mturk/welcome>). In the present study, ratings were collected in years, and AoA values are presented in the resulting dataset both as AoA-in-years and as converted values on a 7-point-Likert scale. The need for comparable psycholinguistic norms has previously been highlighted by other researchers (Schmidtke, Schröder, Jacobs, & Conrad, 2014), and through collection of the German AoA ratings reported in the present article, we endeavored to broaden researchers' stimulus choices.

Method

Ratings were collected using online questionnaires (Qualtrics LLC, 2014) advertised via a large number of Internet forums, mailing lists, flyers, and personal communication between January and September 2014. Respondents were encouraged to disseminate knowledge about the study in order to attract further participants.

Participants

Eligibility for participation was confined to native German speakers who at the time of participation were 18 years of age or over and lived in a German-speaking country. At the start of the questionnaire, all participants were invited to answer some questions on demography, but it was not obligatory to provide any of this information, and participants were able to complete only the ratings section if they so wished. First of all, participants were asked to indicate their age and gender. Education level was probed in two ways, both as the number of years of education received, including primary schooling, and as a formalized level of education, with eight choices given: *weiterführende Schule* (secondary school), *Mittlere Reife oder Äquivalent* (school qualification at age 16 or equivalent), *Abitur oder Äquivalent* (school qualification at age 18 or equivalent), *nicht abgeschlossene Ausbildung nach Schule* (not completed training after school), *abgeschlossene Ausbildung nach Schule* (completed training after school), *abgeschlossenes Erststudium* (first degree), *Zweitstudium oder weiterführendes Studium* (second degree or postgraduate studies), *nicht hier aufgeführt* (not listed). Respondents' language environment was assessed by asking participants to indicate whether they had grown up speaking more than one language, and whether German was the language they spoke most at the time of completing the questionnaire. The options

given were *yes*, *no*, and *I don't know*. If respondents spoke other languages frequently, they were asked to specify which ones. At the end of the questionnaire, the respondents were invited to take part in a prize draw to win one of five €25 Amazon vouchers. Winners were determined and notified in December 2014.

Stimuli

We aimed to collect AoA ratings for a varied set of German word stimuli, to afford experimenters as great a breadth as possible in future experimental stimulus selection. Originally 3,260 words were selected for ratings data collection, but due to an input error in constructing the Qualtrics questionnaire, one word was inputted twice, so that data were obtained for 3,259 words, including 2,363 nouns, 473 verbs (in the infinitive form), 371 adjectives, and 17 adverbs. The remainder of the word forms fulfilled multiple grammatical functions. For example, the word *allein* (alone) could be categorized as an adjective, adverb, or conjunction (from Duden Online: www.duden.de/suchen/dudenonline/allein). The words were 2–17 letters long ($M = 6.73$), and had one to seven phonological syllables ($M = 2.25$). Their frequencies ranged between 0 and 2,839.84 ($M = 33.31$) occurrences per million, according to SUBTLEX-DE (Brysbaert et al., 2011).

The set of words for which we collected AoA values shares 2,244 words with the BAWL-R database for ratings of emotional arousal, valence, and imageability (Vö et al., 2009), and 425 words with the more recent German ANEW database for affective word norms (Schmidtke et al., 2014). All words apart from *jauchzen* (“shout for joy”) and *All* (“universe”) can be found in the cleaned version of the SUBTLEX-DE database (Brysbaert et al., 2011). This considerable overlap in word choice with other psycholinguistic databases will be a useful asset for researchers when choosing their stimuli. A screenshot from the final database may be found in Appendix C.

Data collection

The original set of 3,260 words was divided into 3,240 rating items and 20 control items. The 3,240 rating items were separated into 27 questionnaires, each containing 120 words. Each participant completed one questionnaire only. A one-way independent analysis of variance (ANOVA) showed that the questionnaires did not differ in terms of log10 frequency, $F(1, 3238) = 0.07$, $p = .79$. In each online questionnaire, the 120 items were distributed over 12 consecutive question pages. To vary the items that were presented together on one question page, we created three pseudorandomized versions of each 120-item questionnaire. [Note that AoA values on the full dataset collected by the three different versions did not differ significantly: Kruskal–Wallis test comparison, $H(2) =$

5.66 , $p = .059$.] Further randomization occurred at two levels. The order of the question pages within each questionnaire and the order of words within each question page were randomized for each participant.

The 20 control words (see Appendix A) were chosen from the original word set for all participants to rate, and thus appeared in all questionnaires. Kuperman et al. (2012) had used a set of 52 control words to be rated by all participants. In the present study, ratings for the control words were collected so as to have ratings given by all participants on the same 20 items, which could then be used to (1) identify outliers and (2) investigate the influence of rater differences on AoA ratings. These items were selected to mirror a range of acquisition ages, while having similar frequencies per million ($M = 33$, range = 29 to 35). The 20 control words were randomly added to each questionnaire, one or two per question page. The addition of control words meant that each participant rated 140 words.

Once participants had agreed to take part and answered the questions on demography, they were asked to start with the rating task. At the top of each question page, participants were shown instructions in German, a translation of those used by Stadthagen-Gonzalez and Davis (2006) and Kuperman et al. (2012); the instructions are reproduced in Appendix B. Participants were asked to estimate at what age they had learned each word with its meaning, by inserting the age estimates in years in a text box next to each presented item, or to indicate with an “x” if they did not know an item.

Results

In the following sections, we report how the raw data were prepared for inclusion in the ratings database, how we derived our rated AoA estimates from the raw data, and how we checked their internal and external reliabilities. We further examined the validity of our estimates by comparing them with German children’s production ages (Suchodoletz, 2010). In addition, we assessed the relationship of the present AoA estimates with other psycholinguistic variables. We then outline the characteristics of the participants who contributed to the present ratings. Finally, we examine whether the raters’ characteristics influenced the AoA values and explore the relative effects of lexical variables and participant characteristics on AoA ratings. For all analyses, we used the R language and statistical computing environment (R Development Core Team, 2014).

For the sake of clarity, we will make a distinction between the different types of rating values that we used within the data analysis. We distinguish between the raw ratings, which denote a single observation from one participant (one rating from one participant for one word), and the AoA ratings/estimates, which are all averaged raw ratings from all participants

for that word (our actual AoA norms). Please note that for the purpose of the below analysis, we used the unrounded values for AoA estimates in years and in points on the Likert scale. The values in the database as supplement to this article have been rounded to two decimal places. We will also refer to the control word means (CWMs), which are averages of the raw AoA ratings given by each participant to the words in the control word list.

Database cleaning

The questionnaires were accessed 1,056 times, but only 544 (51 %) people submitted their responses. Three completed questionnaires were excluded outright because they contained only invalid entries or had marked all text boxes with crosses. Within the remaining 541 completed questionnaires, each single raw rating was examined to detect any nonvalid entries. Nine of the raw ratings were deleted because they did not clearly indicate an age. Where the age of learning was given but additional text was also added (e.g., “years” or a question mark), the additional text was deleted. This was the case for four raw ratings. Where participants had indicated with an “x” that the item to be rated was unknown, this was recorded and coded as a missing value. In total, only 179 “x” (unknown) responses were recorded, indicating that overall the words in the stimulus set are likely to be known by adult speakers of German. Information about which words were unknown was added to the final ratings database. Four participants entered “x” (unknown) responses for more than 10 % of the words in the questionnaires that they completed. For these responses, the “x” seemed to function as an indicator that the AoA was unknown, and not that the words themselves were. This assumption was supported by the fact that these participants did not leave any cells blank, to differentiate between not knowing the acquisition age and not knowing the meaning of a word. The words marked with “x” for these four participants were not coded as unknown but as missing values in the database. Where participants entered two different acquisition ages for a word, we entered the mean of the ages in our database, provided that the two values were neighboring integers—that is, for the entry “5 and 6 years,” 5.5 years was used. These entries were interpreted to signify that the participant indicated an approximate age, and such averaging was applied for 57 raw ratings. For six further raw ratings, in which two entries were made that were separated by more than a year, the lower age was used. For example, if “4/11 years” had been entered, “4 years” was used. The different treatment seemed appropriate, because greater differences were interpreted to indicate the learning of the different meanings of polysemous words. For example, the word *Linse* could have been interpreted to mean “lentil” or to refer to an optical lens, and we assumed that these double entries were made to capture the acquisition of different word meanings.

Due to an input error, one word was lost, which resulted in the reduction of the final word count from 3,260 to 3,259

items. Three further items were erroneously presented for rating twice within one of the three different-order questionnaire versions. In one version, the word *Parfum* was entered twice and the word *Nachteil* was left out. The word *mehr* was also presented twice in one questionnaire version. On all occasions, we deleted at random one of the two responses for the word accidentally presented twice.

In line with Kuperman et al. (2012), entries indicating acquisition ages above 25 years were removed; this was done for 134 raw ratings. Checks showed that these had been contributed by 73 different participants.

Following Kuperman et al. (2012), we removed all of some participants’ responses from the dataset because the AoA values they supplied appeared to be outliers in relation to the responses supplied by other participants. The detection of outliers was based on data collected using the control words. All participants had been asked to rate 20 control words, and for all but three participants (who did not provide enough control word ratings) we could average across the ratings given for the 20 control words to obtain the mean of this person’s control word ratings (CWMs). We then compared these to the sample’s overall CWM. We identified four participants whose CWMs correlated below .4 with the mean. By removing these four participants’ contributions, 459 raw ratings were excluded. Although the cutoff of .4 was arbitrary, it emulates that used by Kuperman and colleagues, who also removed lists that had a correlation below .4 with the Bristol norms. After these exclusions, we used the remaining 73,948 raw ratings from 537 participants to determine the average AoA ratings per word. The average correlation between the overall sample’s CWM and each participant’s CWM was $r = .83$.

Data analysis

AoA ratings Words (excluding control words) were rated on average 19.57 times (range = 7 to 27), a raw ratings per word ratio similar to those recorded for the Kuperman et al. (2012) and Schröder et al. (2012) datasets. (The 20 control words received an average of 528.7 raw ratings, with a range of 518 to 534.) AoA estimates were computed by averaging across all raw ratings received per word, following the exclusions discussed previously. The overall mean AoA estimate was 7.01 years ($SD = 2.62$, range = 1.67 to 15.83 years). The ten words rated as the latest acquired (including control words) were *Blog* (blog), *Quiche* (quiche), *Webcam* (webcam), *Fakultät* (faculty), *Kurator* (curator), *Mensa* (refectory), *Lobbyist* (lobbyist), *Veganer* (vegan), *Vektor* (vector), and *Webseite* (website). The ten words that were rated as the earliest acquired were *Papa* (dad), *Mama* (mum), *Bett* (bed), *Nase* (nose), *Bauch* (belly), *Mund* (mouth), *Hose* (trousers), *Kind* (child), *Hand* (hand), and *Puppe* (doll). The words that were the most often marked as unknown were *Quorum* (quorum), *Pomp* (pomp), *Alkoven* (alcove), *Charta* (charter), *Jade* (jade), *Noblesse* (nobility), and *Viola* (viola).

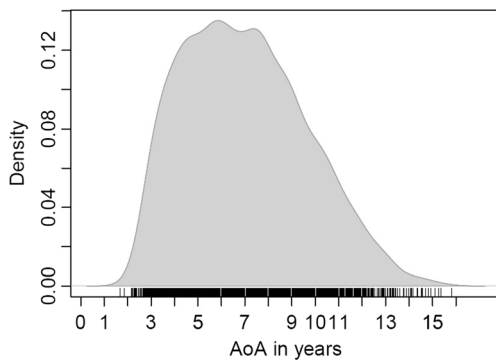


Fig. 1 Distribution of AoA estimates in years for 3,259 German words

The frequency distribution in Fig. 1 shows that averaged AoA estimates were slightly positively skewed ($\text{skew} = 0.41$, $\text{kurtosis} = -0.45$), peaking at around 6–7 years. Few items were rated as having been acquired before the age of 2 years. Acquisition was considered to occur mostly from the age of 2.5 years onward, with a slowly diminishing acquisition rate in later childhood and the teenage years.

Internal reliability A split-half correlation was performed to further examine internal reliability. Participants were divided into two groups of even ($N = 268$) and odd ($N = 269$) numbered participants, and the averaged AoA estimates were calculated for each group for all 3,259 words. The ratings from the two groups were found to correlate very highly ($r_s = .91$, $p < .001$).

External reliability The present ratings were compared to other existing ratings to gauge external reliability. Schröder et al. (2012) presented the current largest set of German AoA ratings in Likert scale format (1 = 0–2 years, 2 = 3–4 years until 7 = 13+ years). For comparison purposes, we converted our raw ratings accordingly, as if respondents had used the Likert scale to issue a response. The by-items average AoA values were then computed to generate new AoA estimates on a 7-point Likert scale. These ratings in Likert scale can also be found in the downloadable supplement. Our set included 203 of the words in Schröder et al.’s 824 items, and the AoA values for these words in the Schröder et al. and in our

data were highly correlated, $r_s = .91$, $p < .001$. Of the words the two studies have in common, more words had been rated to be early-acquired (between the ages of 3–8) and fewer had been rated as late-acquired (after the age of 8) in the present study compared to Schröder’s ratings.

The present ratings were also compared to AoA estimates that had been collected in English. Table 1 shows that the cross-language comparisons yielded high correlations. The present stimulus set shared 2,683 words with the Kuperman et al. (2012) set of American English, and the AoA ratings were strongly positively correlated ($r_s = .74$, $p < .001$). Our stimulus set also had 782 words in common with the Cortese and Khanna (2008) norms, and again correlated highly ($r_s = .70$, $p < .001$). Likewise, the AoA estimates for the 1,031 words appearing in both the Bristol norms (Stadthagen-Gonzalez & Davis, 2006) and our norms correlated highly ($r_s = .77$, $p < .001$). (Note that we compared the Likert-scale AoA estimates from the Bristol and the Cortese & Khanna norms with the AoA estimates converted to Likert-scale values in our norms.)

External validity We examined the external validity of our AoA estimates by comparing them to German data on early speech production. Suchodoletz (2010) had asked parents of 20- to 25-month-old children which words their children produced. This resulted in a list of 406 words, and for each word the percentage of children who produced it was reported. The present dataset had 229 words in common with Suchodoletz’s production word list, and we found a moderate relationship of $r = -.53$ ($p < .001$) between the percentage of 20- to 25-month-olds who produced a word and our AoA estimates for each word (see Fig. 2). The words estimated by adults as having been acquired earlier were more likely to have been produced by more children in spontaneous speech.

Relationship of AoA with other word characteristics As we mentioned above, previous research has shown that early-acquired words tend to be shorter and more frequent and to have more orthographic neighbors. To inspect these relationships for our AoA estimates, a number of correlations with known word properties were performed. Frequency estimates

Table 1 Correlation coefficients of present German age-of-acquisition (AoA) norms to extant rating collections

Ratings Study/Collection	Language	Measurement	Number of Words in Common	Correlation Coefficient Spearman’s Rho
Schroder et al. (2012)	German	Likert (1–7 scale)	203	.91***
Kuperman et al. (2012)	American English	years	2,683	.74***
Cortese and Khanna (2008)	American English	Likert (1–7 scale)	782	.70***
Stadthagen-Gonzalez and Davies (2006; Bristol norms)	British English	Likert (1–7 scale)	1,031	.77***

*** $p < .001$

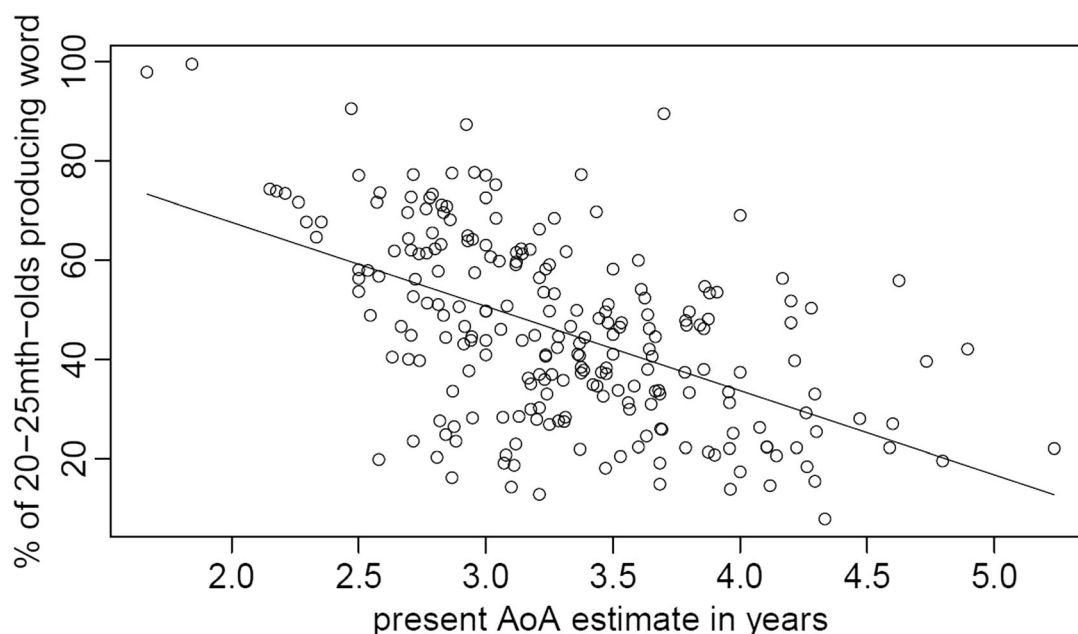


Fig. 2 Scatterplot showing the relationship between the percentage of children 20–25 months of age producing a word and our present AoA ratings

(log10) were taken from the SUBTLEX-DE database (Brysbaert et al., 2011). Word length was measured in letters and in number of phonological syllables. Two different measures were employed so as to capture the orthographic similarity neighborhood characteristics of the words. For each word, we calculated Coltheart's *N* (Coltheart, Davelaar, Jonasson, & Besner, 1977), an estimate of how many words of the same length can be created by changing one letter in a target word. Since Coltheart's *N* has been argued to be insensitive to orthographic similarity neighborhoods for longer words, for each word we also calculated the orthographic Levenshtein distance (Yarkoni, Balota, & Yap, 2008), a count of how many changes (insertions, deletions, or substitutions) are needed to transform a word into its 20 nearest orthographic neighbors. Both measures were computed using the R package *vwr* (Keuleers, 2013), on the basis of the Clearpond word list of 27,751 German words (Marian, Bartolotti, Chabal, & Shook, 2012). As expected, words that were acquired earlier tended to be shorter and more frequent, and to have more orthographic neighbors (see Table 2). Both length measures were only weakly correlated with AoA.

A regression analysis showed that the variance in German AoA estimates was explained, in part, by variation in word frequency, length in syllables, and old20 as a measure of orthographic similarity, $R^2 = .34$, $F(3, 3255) = 567.9$, $p < .001$ (see the regression summary of Model 1 in Table 3). Given that old20 better captures orthographic similarity in long words (Yarkoni et al., 2008), we chose old20 rather than ColtN as a measure for orthographic similarity. However, substituting ColtN for old20 yielded similar results. We completed further analyses to check whether it would be useful to include the word length in letters variable as a predictor in the analysis of AoA values, either instead of or in addition to the inclusion of word length in syllables,

and found that the inclusion of syllable length alone accounted for most variance in AoA. The regression analysis with linear variables showed that words were rated as being earlier-acquired when they were more frequent, shorter, and more orthographically similar to other words.

When inspecting the scatterplots in Fig. 3, curvilinear relationships appeared likely to explain added variance. Therefore, quadratic terms for old20 and syllable numbers were gradually added to the model. Adding the term to describe nonlinearity for old20 improved the model fit. With the inclusion of the quadratic term for syllable numbers, the linear term for number of syllables was no longer predictive, but it was retained in the model because higher-order terms (like the quadratic term) cannot be estimated accurately without including lower-order terms (Cohen, Cohen, West, & Aiken, 2003). This final model [$R^2 = .36$, $F(5, 3253) = 364.2$, $p < .001$; see Model 2 in Table 3] included all predictors (log frequency, old20, and number of syllables) and the quadratic terms for old20 and number of syllables. A comparison between Model 1, which only estimated linear relationships, and Model 2, which also included the predictors' quadratic terms, demonstrated a significance difference [$F(2, 3253) = 38.87$, $p < .001$]. The inclusion of the quadratic term for number of syllables showed that words with large numbers of syllables tended to be rated as later acquired, whilst other words tended to be rated independent of their acquisition age. The fact that the inclusion of the quadratic term for orthographic similarity was beneficial for the model demonstrated that words with more neighbors tended to be rated as being earlier acquired, but that this relationship is attenuated as neighborhood becomes sparser.¹

¹ We thank an anonymous reviewer for suggesting that we investigate nonlinearity.

Table 2 Correlation coefficients for multiple comparisons of the present AoA norms for German words to other word characteristics

Ratings Study/Collection	AoA Estimate	Log Frequency	Length in Letters	Syllable Number	Old20
Log frequency	-.53***				
Length in letters	.29***	-.26***			
Syllable number	.36***	-.26***	.75***		
Old20	.41***	-.40***	.81***	.66***	
ColtN	-.35***	.33***	-.49***	-.42***	-.58***

Log frequency (SUBTLEX-DE frequency estimates; Brysbaert et al., 2011), old20 (orthographic Levenshtein distance; Yarkoni et al., 2008), ColtN (Coltheart's *N*; Coltheart et al., 1977). *** $p < .001$

Rater characteristics Ratings from 537 participants (male = 170, female = 367) contributed to the final AoA estimates. Of the total sample, 531 participants volunteered age information ($M = 35.19$, $SD = 14.04$, range = 18 to 83 years). The age distribution is shown in Fig. 4. The age group between 18 and 20 years old comprised 6.21 % of all participants. The largest participant group consisted of 21- to 30-year-olds, with 44.63 %, followed by 19.4 % who were between 31 and 40 years old. Another 14.5 % were between 41 and 50 years of age, and 7.91 % were between 51 and 60 years old. Only 4.52 % fell into the 61–70 age group, and the remaining 2.82 % were 71 or older.

Education was assessed in two ways, by asking for both years of education and formal level of education. All participants gave information about their formal level of education. Figure 5 shows the exact percentages for each possible choice given in the questionnaire. Overall, about a quarter of all participants (25.13 %) were either still in or had completed secondary education at the time of responding. The majority (70.19 %) reported that they had completed job training or a first degree. This group included 20.48 % who were educated at a postgraduate level. Finally, 4.66 % felt that their education level was not represented by the choices given.

Participants were also asked to indicate the number of years in education including primary education, and 506

respondents contributed this information (in years, $M = 17.5$, $SD = 3.33$, min = 6, max = 33). Not including the participants whose level of education was not represented, the two measures correlated moderately but significantly ($N = 481$, $r_s = .55$, $p < .001$). Because more respondents had specified their formal level of education than their number of education years, the former measure was used for all further analysis.

All but one participant gave information on whether they had grown up with more than one language: 9.51 % (51 participants) reported that they had been raised with languages in addition to their mother tongue (German). Of the 536 responses given, fewer than 1 % stated that they did not use German as their main everyday language, but 68 % reported using other languages regularly, with English as the most common one. Dialects were not included in the count.

Did ratings vary with rater characteristics? To examine whether rating values varied with participants' age, gender, education, multilingualism, or language usage, the control words' mean AoA (CWM) for each participant was computed. The CWM was the average of the AoA values assigned by the participant to the words in the control word list ($M = 5.63$ years, $SD = 1.49$). Only the responses of participants who had rated all 20 control words and also given information

Table 3 Simultaneous regression analysis with the present German AoA norms as the outcome variable

	Model 1 Standardized Coefficient	Coefficient	SE	t Value	Model 2 Standardized Coefficient	Coefficient	SE	t Value
Log frequency	−0.44	−1.48	0.05	−28.47***	−0.43	−1.43	0.05	−27.56***
Syllable number	0.15	0.48	0.06	7.92***	0.01	0.02	0.19	0.13
Old20	0.13	0.37	0.05	6.72***	0.68	1.85	0.17	10.45***
Syllable number ²					0.13	0.08	0.04	2.23*
Old20 ²					−0.55	−0.23	0.03	−8.76***
R ²	.34				.36			

Model 1 includes the word characteristics log frequency, number of phonological syllables, and the orthographic similarity measure orthographic Levenshtein distance (old20) as predictor variables. Model 2 additionally adds quadratic terms for old20 and the number of syllables. *SE* is the standard error for the coefficient. *** $p < .001$, * $p < .05$

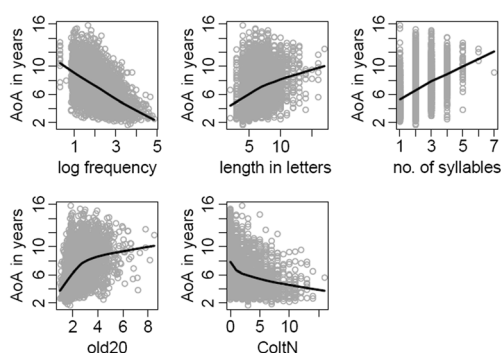


Fig. 3 Scatterplots of AoA with the lexical variables log frequency (SUBTLEX-DE frequency estimates; Brysbaert et al., 2011), length in letters, number of syllables, old20 (orthographic Levenshtein distance; Yarkoni et al., 2008), and ColtN (Coltheart's N ; Coltheart et al., 1977). Lowess smoothers are shown

on all participant characteristics were included in this analysis. This reduced the number of participants to 453.

Participant age was very weakly positively correlated with the CWM ($r = .16, p < .001$), suggesting that older participants rated words as being slightly later learned than did younger participants. This concurs with the findings by Kuperman and colleagues (2012) in their American English sample. We investigated this further by examining the correlation between age and the raw control word ratings (the original AoA values for each control word given by each participant). Correlations varied between $r = -.15$ and $.38$. The control words that showed the strongest positive association with age were *Karriere* ($r = .38$, “career”), *Hubschrauber* ($.33$, “helicopter”), and *Urlaub* ($.24$, “holidays”), and at the other end of the spectrum, the words *Milch* ($-.15$, “milk”), *reparieren* ($-.07$, “repair”), and *Tier* ($-.05$, “animal”) showed extremely low and negative age correlations. This demonstrates that the positive direction of the association was not

valid for all control words, but may arguably have been influenced by more modern words. To examine whether the age correlation would disappear if the raw ratings from older participants were excluded, we computed CWMs again for all participants up to the age of 59. The correlation between CWMs and age was then no longer significant ($N = 426$, $r = .05, p = .285$).

We investigated whether the CWMs differed between male and female participants. In the present sample, female respondents ($N = 314$, $M = 5.51$, $SD = 1.44$) tended to give slightly lower ratings than did male respondents ($N = 139$, $M = 5.90$, $SD = 1.56$). This result was significant, $t(245.664) = 2.53, p = .012$, but only had a small effect size, $r = .12$.

CWMs did not differ with regard to participants' formal education levels [one-way ANOVA: $F(7, 445) = 1.75, p = .096$], although there was a tendency for participants with the lowest formal education levels and those who had chosen the “not listed” option to rate words as being later acquired (see Table 4). However, very small participant numbers were included in our sample for these education levels. A post-hoc Tukey multicomparison between groups was conducted, in which groups were defined by education level and the observations were the CWMs (average AoA ratings of control words per participant). The participants in different education groups did not significantly differ in terms of CWMs.

Participants with multilingual ($N = 40$, $M = 5.75$, $SD = 1.72$) and monolingual ($N = 413$, $M = 5.62$, $SD = 1.46$) upbringing did not differ in CWMs [$t(44.667) = 0.46, p = .644$]. Equally, whether respondents used other languages frequently ($N = 313$, $M = 5.57$, $SD = 1.43$) or not ($N = 140$, $M = 5.76$, $SD = 1.61$) did not impact the CWMs [$t(239.928) = -1.21, p = .227$].

To verify these results based on the means of the 20 control words, we entered the full raw data set into a

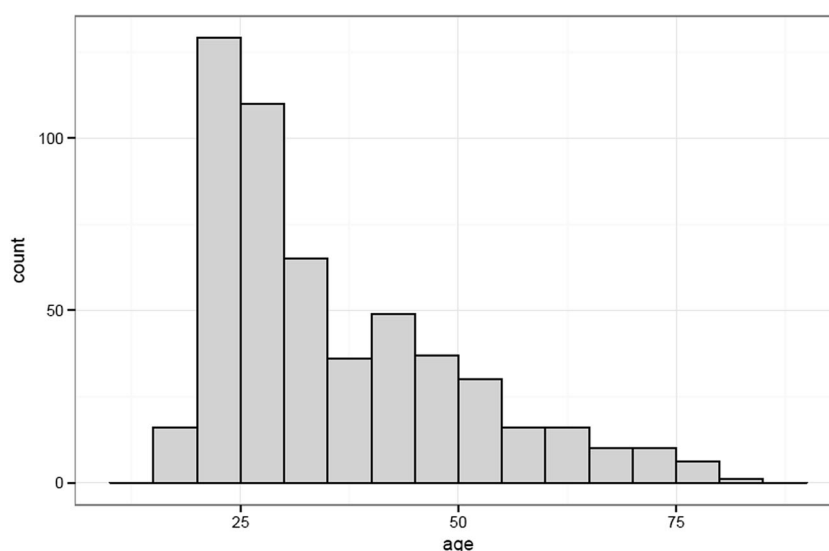


Fig. 4 Age distribution of the 531 participants who gave information about their age

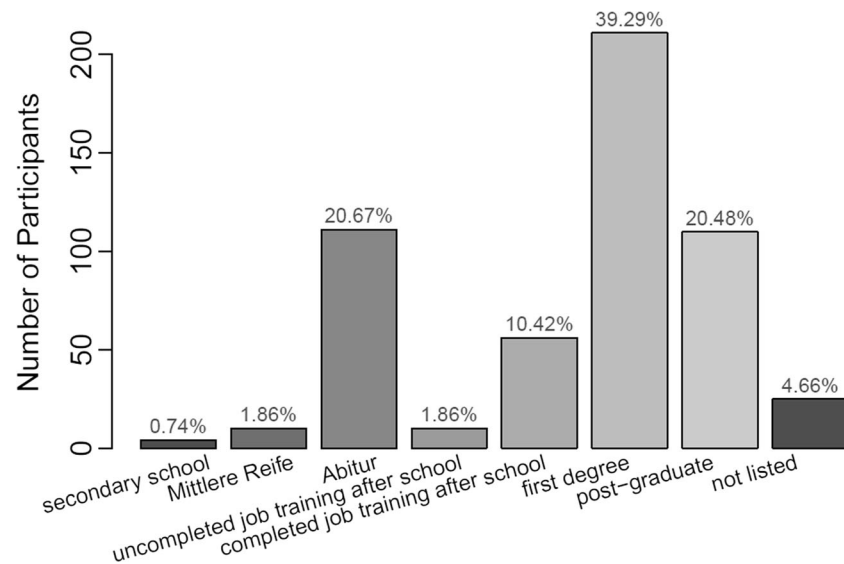


Fig. 5 Percentages of 537 participants within formal levels of education. Mittlere Reife, *Mittlere Reife oder Äquivalent* (school qualification at age 16 or equivalent); Abitur, *Abitur oder Äquivalent* (school qualification at age 18 or equivalent)

mixed-effects model using lme4 (Bates, Maechler, Bolker, & Walker, 2014). Eliminating empty fields resulted in reducing the participant numbers to 528. All of these participants had given all demographic information. The first model (Model 1) included only random effects for participants and words, but no fixed effects. Groups of predictor variables were then added progressively to each new model, and models were compared by using the likelihood ratio test (LRT). The following list shows the order in which the predictor groups were added, with results comparing each model to its predecessor given in parentheses:

Model 2: log frequency, syllable numbers, and old20 (LRT, $\chi^2 = 1400.9$, 3 *df*, $p < .001$);

Model 3: syllable numbers² and old20² (LRT, $\chi^2 = 73.411$, 2 *df*, $p < .001$);

Model 4: Questionnaire version, age, gender, formal education level, multilingual upbringing, and frequent use of other languages (LRT, $\chi^2 = 29.847$, 13 *df*, $p = .005$);

Model 5: age² (LRT, $\chi^2 = 9.5151$, 1 *df*, $p = .002$);

Model 6: random slope for frequency (LRT, $\chi^2 = 1,782.2$, 2 *df*, $p < .001$)

In the final model (see Table 5, showing Model 6), the lexical variables log frequency, old20, the quadratic terms syllable numbers² and old20², and the participant characteristics gender, age, and age² emerged as significant predictors. Adding a term to capture the effect of random participant variation in frequency also made a significant difference to model fit. The variable frequent-other-language-usage remained predictive until the inclusion of the random slopes for frequency in Model 6, suggesting that participants who frequently spoke other languages rated words as earlier-learned.

The fact that the model fit was significantly improved by the inclusion of the random frequency effect of participants indicates unexplained or unmeasured variation among the participants in the relationship between the frequency of words and participants' AoA estimates. On average, higher-frequency words are associated with earlier AoA ratings. For some participants, however, the slope is steeper, and for some the slope is shallower, than the average. We can conjecture that this effect may overlap with the impact of frequent-other-language-usage. However, this is a subject for future research to investigate.

Until the inclusion of the quadratic age term, the linear term for the age effect was associated with a positive coefficient, suggesting that older participants gave higher ratings. With the inclusion of the quadratic age term, the linear age effect was

Table 4 Descriptives of control word mean AoA ratings by formal education level

Education Level	N	%	CWM	SD
Secondary school	3	0.66	6.70	0.87
<i>Mittlere Reife</i> or equivalent	7	1.55	6.64	1.86
<i>Abitur</i> or equivalent	96	21.19	5.46	1.37
Uncompleted job training after school	9	1.99	4.84	0.92
Completed job training after school	46	10.15	5.69	1.48
First degree	186	40.40	5.61	1.49
Postgraduate	88	19.43	5.64	1.61
Not listed	21	4.64	6.23	1.34

N = number of participants, total N = 453; % = percentage of 453 participants; CWM = control word mean (averaged AoA ratings, in years, of 20 control words)

Table 5 Final mixed-model summary table of word and participant characteristics, including quadratic terms

Fixed Effects	Estimate	SE	<i>t</i> Value	Wald Confidence Intervals	
				2.50 %	97.50 %
(Intercept)	8.3941	0.8596	9.77	6.7093	10.0788
Log frequency	−1.4341	0.0555	−25.84	−1.5429	−1.3253
Syllable number	0.0555	0.1902	0.29	−0.3173	0.4283
Old20	1.7928	0.1747	10.26	1.4504	2.1353
Syllable number ²	0.0753	0.0359	2.10	0.0049	0.1456
Old20 ²	−0.2163	0.0254	−8.51	−0.2661	−0.1665
Version2	0.0553	0.1392	0.40	−0.2175	0.3281
Version3	0.1051	0.1390	0.76	−0.1673	0.3775
gender_female	−0.4948	0.1217	−4.07	−0.7333	−0.2564
Age	−0.0494	0.0239	−2.07	−0.0962	−0.0025
Formal.education.level_MittlereReife	−0.4104	0.7668	−0.54	−1.9134	1.0925
Formal.education.level_Abitur	−0.9533	0.6634	−1.44	−2.2536	0.3469
Formal.education.level_training_uncompleted	−1.1883	0.7662	−1.55	−2.6901	0.3134
Formal.education.level_training_completed	−0.6169	0.6721	−0.92	−1.9342	0.7004
Formal.education.level_first_degree	−0.8092	0.6568	−1.23	−2.0965	0.4782
Formal.education.level_postgrad	−0.5175	0.6654	−0.78	−1.8217	0.7866
Formal.education.level_not_listed	−0.3669	0.7021	−0.52	−1.7431	1.0093
Multilingual.upbringing_no	−0.0804	0.1942	−0.41	−0.4609	0.3002
Freq.other.language.usage_no	0.0878	0.1235	0.71	−0.1542	0.3298
Age ²	0.0007	0.0003	2.58	0.0002	0.0012
Random Effects			Variance	Std. Dev.	Corr.
Word (Intercept)			4.059	2.015	
Participant (Intercept)			5.995	2.448	
lgSUBTLEX			0.235	0.485	−.87
Residual			3.796	1.948	

The model specified random intercepts for both words and participants, as well as random slopes for frequency over participants. For categorical fixed effects variables, estimates indicate effect with regard to the reference variable, which were "Version 1" (for Questionnaire Version), "male" (for gender), "secondary school" (for Formal.education.level), "yes to multilingual upbringing" (for Multilingual.upbringing), and "yes to frequent-other-language-usage" (for Freq.other.language.usage).

estimated to have a negative-signed coefficient, whereas the quadratic age effect was positive. Since the inclusion of age² significantly improved the model, this may indicate that participant age only affected ratings in the older participants, rather than in all age groups. Whereas older participants showed a slight tendency to rate words as being later-learned, there did not seem to be systematic variation among the younger participants. This finding concurs with our previous results regarding CWMs. However, the coefficient estimates for age were small, especially for the quadratic term, indicating that the overall age effect is small.

Formal education level was not predictive of raw ratings in any of the models. We also investigated whether age interacted with formal education level. We tested the Age × Formal Education level interaction twice. Adding it to Model 4 did not improve the model (LRT, $\chi^2 = 5.7562$, 7 *df*, $p = .569$). Adding the interaction after the inclusion of the age² term into Model 5

also did not improve the model fit (LRT, $\chi^2 = 3.8022$, 7 *df*, $p = .802$), and the interaction term was therefore not retained in the model. It may nevertheless be noteworthy that age became nonsignificant when the Age × Formal Education variable was added to the model. The fact that the Age × Education interaction did not have an effect may be due to two reasons. First, there were very few individuals in the lowest education categories "secondary school" and "Mittlere Reife." Second, the education options given as choices to participants could have been interpreted in a nonordinal way. The choices "uncompleted training after school" and "completed training after school" did not specify whether or not previous school education had to have been completed. Arguably, then, this option could comprise participants both with and without previous completed formal school education, and hence could possibly include participants who had spent less time in education than those who chose the preceding education level, "Abitur" or equivalent (A-levels or

equivalent).” Consequently, our formal education choices may not describe incrementally ordered education categories. Likewise, the last education level of “not listed” cannot be classified as an education category. We examined whether the exclusion of the “not listed” education category would reveal the formal education variable to be a significant predictor of AoA ratings. After this exclusion, a linear mixed-effects model analysis on the reduced number of 503 participants indicated that the effects of the predictor variables remained the same, and that formal education did not emerge as a predictor.

The mixed-effects model analysis identified gender as a predictor for raw AoA ratings, with women showing a tendency to rate words as being learned earlier. Interestingly, gender lost its predictive power when we added an Age \times Gender interaction term to Model 4. However, as there was no Age \times Gender interaction effect, the interaction term was not kept in the model.

Discussion

We have presented 3,259 AoA estimates for German words, in years and in the form of points on a 7-point Likert scale. The present AoA collection extends existing norms and widens the choice of word stimuli for researchers. The present collection has a considerable number of words in common with normative databases for other word properties. Furthermore, the high internal and external reliabilities of the present norms were demonstrated. Specifically, the high correlation of item ratings in common with Schröder et al. (2012) suggests that the present ratings can be used in conjunction with Schröder’s norms. The correlations with AoA estimates collected in American and British English were high and comparable to such cross-language correlations in other studies (Ferrand et al., 2008). In fact, such high cross-language correlations are remarkable, considering that translations are not always unequivocally one to one. These translation difficulties naturally result in some attenuation in the strength of correlations.

Our AoA estimates were positively correlated with the frequency with which words occurred in the spontaneous speech production of 20- to 25-month-old children (Suchodoletz, 2010). Although Schröder et al. (2003) had previously found that rated AoA was more closely associated with picture-naming ages than with production ages, this association nevertheless indicates the validity of the present set of AoA estimates. It may also be pointed out that the size of the correlation is even more striking, considering that the Suchodoletz word production percentages are based on the vocabulary of children up to the age of 2 years only. Furthermore, this lends

support to the concept of rated AoA as reflecting order of acquisition.

We used regression to explore which lexical variables and which participant characteristics accounted for variance in the AoA estimates. In accordance with previous research (Stadthagen-Gonzalez & Davis, 2006), the present rated AoA estimates were influenced by word frequency and orthographic neighborhood, so that early-learned words tended to be more frequent and more similar in spelling with other words. The curvilinear effect of word length in syllables indicates that only very long words showed a tendency to be rated as later acquired. This finding was further confirmed when exploring predictors of the raw AoA ratings in a mixed-effects model.

Previous studies have examined which factors other than word characteristics influence AoA ratings. In the present study, we followed Kuperman et al.’s (2012) lead and looked at the impacts of age, gender, education, and general language usage on our ratings. This was first examined by analyzing the effects of participant characteristics on CWMs.

Consistent with the findings reported by Kuperman et al. (2012), older participants in this sample tended to give slightly higher ratings, although the correlation was very weak. Kuperman et al. proposed that older participants have a greater age range available to choose from, and are therefore more prone to give higher ratings. Other research, however, has suggested that ratings do not vary with age. Schröder et al. (2012) did not find an age difference, but they only employed a small sample size. Walley and Metsala (1992) demonstrated that 5-year-old children give AoA estimates similar to those of adults.

In the present sample, age was confirmed as a predictor of AoA ratings, when added to the mixed-effects model on raw estimates. The fact that the model fit was improved by the inclusion of the quadratic age term, corresponding to the non-linear component in the age effect, confirmed that the relationship between AoA estimates and age is curvilinear in such a way that older participants tend to rate words as being later-learned.

In the present sample, the correlation of participant age with AoA CWMs disappeared when the ratings from older participants were no longer taken into account. This is consistent with the findings of Cuetos et al. (2012), who showed that ratings from elderly participants were better predictors of word recognition performance in Alzheimer’s patients. Cuetos and colleagues found that although the ratings from elderly and young adults correlated highly, the ratings also differed greatly for some words (e.g., *robot*, *television*), indicating that the day-to-day vocabulary that people use changes with time. In the present sample, some of the later-acquired words (*career*, *helicopter*) may have driven the association between age and AoA, and it is conceivable that this was due to a

generational change in day-to-day vocabulary. However, some caution may be advisable here, since De Deyne and Storms (2007) found that participants who were in their 50s rated words as being significantly later-learned than did participants in their 20s, and that this was true for both common words and modern words. As was suggested by an anonymous reviewer, the question of whether the small age effect present in our data was due to a genuine tendency of older participants to rate words as being later-acquired or to a generational difference in word acquisition (common vs. modern words) could be addressed more conclusively by taking into account the appearance of the words in the general language vocabulary. Unfortunately, this is beyond the scope of this article.

We found that men tended to rate words as a little later-acquired than did women. This is consistent with Winters et al. (1978), who suggested that female respondents give lower ratings because, in contrast to male participants, female respondents may base their AoA estimates on their current children's usage and knowledge of words. The inclusion of an Age \times Gender interaction did not significantly improve the model. We note that the gender effect that we observed is opposite in direction to that observed by Kuperman et al. (2012), whereas Schröder and colleagues did not find a gender difference, equally indicating inconsistency in the empirical picture. We think that the precise shape of the relationship between age and AoA ratings, and the potential modulation of the age effect by a gender effect, merits further investigation in future research.

In the present analysis, education level did not affect the ratings, in line with Kuperman et al.'s (2012) findings, but we have highlighted that participant numbers were very unevenly distributed over education categories. We also examined whether the respondents' language environment and usage influenced the ratings. Multilingualism has been linked to greater linguistic awareness (Bialystok, 1988). Bilingual children tend to have smaller vocabularies than monolingual children, when comparing word learning in one language only (Bialystok, 2009). In the present study, almost 10 % of participants had been raised with more than one language. However, consistent with Kuperman et al.'s (2012) findings, AoA ratings did not differ as a function of multilingualism. In the present sample, 68 % used other languages frequently, which intuitively could also be argued to lead to greater language awareness. However, a simple group comparison showed that in the present sample, the frequent use of languages other than the mother tongue did not influence the ratings given.

In the mixed-effects model, log frequency emerged as the strongest predictor for the raw AoA estimates. It is therefore interesting to note that the model fit was significantly

improved by taking into account the random variation of the frequency effect between participants. This indicates that although higher-frequency words were, on average, associated with earlier AoA ratings, there was a measureable variation among participants in the slope of that effect. The reasons for such variation, we think, also merit further investigation in future research.

We have presented German AoA ratings with high internal and external reliabilities for 3,259 words comprising a number of word classes. A comparison to early German word production ages provided evidence for the ratings' validity as reflecting the order of word acquisition. Ratings were weakly influenced by gender, but not by education, multilingualism, or frequent use of other languages. Although a weak correlation with age was found, this association may have been driven by modern words being rated as later-acquired by older participants, which may inform sampling for future ratings collections.

Author note This study was supported by Oxford Brookes University through a PhD studentship awarded to the first author. The study is part of the first author's doctoral dissertation. The authors thank all of the respondents for their participation.

Appendix A: List of 20 control words

Control Word	Translation
Milch	milk
Tier	animal
Stuhl	chair
warm	warm
schwach	weak
Hubschrauber	helicopter
Urlaub	holiday
Nachmittag	afternoon
Sommer	summer
kompliziert	complicated
reparieren	repair
Fahrer	driver
Gebiet	territory
Interesse	interest
Tatsache	fact
zulassen	allow
Zustand	condition
Dienst	service
Karriere	career
annehmen	assume

Appendix B: Instructions given to participants

This is a direct translation into German of the instructions used by Stadthagen-Gonzalez and Davis (2006) and Kuperman et al. (2012):

Bitte geben Sie an, in welchem Alter (in Jahren) Sie jedes Wort in der Liste jeweils gelernt haben. Eine ungefähre Altersangabe ist ausreichend für diese Einschätzung. Wenn Sie die Bedeutung eines Wortes nicht kennen, dann schreiben Sie einfach ein X in den Textkasten. Mit „ein Wort lernen“ meinen wir das Alter, in dem Sie das Wort verstanden hätten, wenn es jemand in Ihrer Anwesenheit benutzt hätte, auch wenn Sie es nicht selbst zu diesem Zeitpunkt hätten benutzen, lesen oder schreiben können.

Original in English (Stadthagen-Gonzalez & Davis, 2006):

Please indicate (in years) the age at which you learned each of the words on the list. An approximate age is good enough for this rating. If you do not know the meaning of a word, just write an X on that space. By “learning a word” 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.

Appendix C: Screenshot of downloadable German AoA database

The following information is given in the columns: Word (original word), Upper (Word in capital letters), BE_TWord (British English translation), AM_TWord (American English translation), RatperWord (Number of ratings each word has received), AoAestimate (German AoA, computed as average from all individual ratings received), SD (standard deviation of AoAestimate), min (lowest individual rating received), max (largest individual rating received), unknown (number of times the word has been marked as not known), AoALikert (AoAestimate converted into score on 7-point-Likert scale), SDLikert (standard deviation of AoALikert), minLikert (lowest score received on Likert scale), and maxLikert (largest score received on Likert scale).

Values are shown to two places after the decimal point, where applicable.

Please note for German users: decimal mark used is the point (.).

Word	Upper	BE_TWord	AM_TWord	RatperWord	AoAestimate	SD	min	max	unknown	AoALikert	SDLikert	minLikert	maxLikert
Aal	AAL	eel	eel	21	6.43	2.27	3	11	0	3.48	1.17	2	6
Aas	AAS	carion	carion	17	8.65	2.74	5	14	0	4.53	1.37	3	7
Abbau	ABBAU	dismantling	dismantling	16	9.25	2.79	4	13	0	4.69	1.49	2	7
abbauen	ABBAUEN	dismantle	dismantle	17	7.71	3.74	2	15	0	3.94	1.85	1	7
Abbild	ABBILD	likeness	likeness	27	9.37	2.68	3	14	0	4.93	1.41	2	7
Abbruch	ABBRUCH	demolition	demolition	17	8.12	3.28	5	18	0	4.06	1.25	3	7
Abdruck	ABDRUCK	mark	mark	26	6.85	3.11	3	16	0	3.62	1.50	2	7
Abend	ABEND	evening	evening	19	3.05	1.03	1	5	0	1.84	0.60	1	3
Abendrot	ABENDROT	red sunset	red sunset	19	4.21	1.90	2	8	0	2.37	0.96	1	4
abends	ABENDS	in the evening	in the evening	18	3.69	1.32	1	7	0	2.11	0.68	1	4
Abfahrt	ABFAHRT	departure	departure	14	6.43	2.44	4	10	0	3.36	1.22	2	5
Abfall	ABFALL	rubbish	rubbish	17	4.29	2.71	2	12	0	2.35	1.37	1	6
Abflug	ABFLUG	take off	take off	18	7.00	2.97	4	14	0	3.72	1.53	2	7
Abfuhr	ABFUHR	snub	snub	14	9.36	2.21	6	13	0	5.00	1.18	3	7
Abgang	ABGANG	leaving	leaving	17	9.35	3.22	3	14	0	4.94	1.60	2	7
abgeben	ABGEBEN	hand over	hand over	19	4.24	1.92	1.5	10	0	2.37	0.90	1	5
Abgrund	ABGRUND	abyss	abyss	17	7.65	2.42	4	12	0	3.94	1.25	2	6

References

- Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, 31, 531–552. doi:10.3758/BF03200732
- Álvarez, B., & Cuetos, F. (2007). Objective age of acquisition norms for a set of 328 words in Spanish. *Behavior Research Methods*, 39, 377–383. doi:10.3758/BF03193006
- 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, 283–316. doi:10.1037/0096-3445.133.2.283
- Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, 34, 424–434. doi:10.3758/BF03195471
- Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology*, 50A, 560–585. doi:10.1080/783663595
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4 (R, package version 1.1-7)*. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology*, 24, 560–567.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, 12, 3–11. doi:10.1017/S1366728908003477
- Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, 93, 498–520.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, 33, 73–79. doi:10.3758/BF03195349
- Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2001). Age-of acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Current Psychology of Cognition*, 20, 401–443.
- Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, 15, 208–216. doi:10.3758/BF03197718
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., B lte, J., & B hl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58, 412–424. doi:10.1027/1618-3169/a000123
- Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, 64, 545–559. doi:10.1080/17470218.2010.503374
- Brysbaert, M., & Ellis, A. W. (2015). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology. Advance Online Publication*. doi:10.1080/02687038.2015.1106439
- Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, 13, 992–1011. doi:10.1080/13506280544000165
- Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, 150, 80–84.
- Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, 104, 215–226. doi:10.1016/S0001-6918(00)00021-4
- Cameirão, M. L., & Vicente, S. G. (2010). Age-of-acquisition norms for a set of 1,749 Portuguese words. *Behavior Research Methods*, 42, 474–480. doi:10.3758/BRM.42.2.474
- Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, 12, 563–576. doi:10.1016/S0022-5371(73)80036-2
- Catling, J., Dent, K., Preece, E., & Johnston, R. (2013). Age-of-acquisition effects in novel picture naming: A laboratory analogue. *Quarterly Journal of Experimental Psychology*, 66, 1756–1763. doi:10.1080/17470218.2013.764903
- Chalard, M., & Bonin, P. (2006). Age-of-acquisition effects in picture naming: Are they structural and/or semantic in nature? *Visual Cognition*, 13, 864–883. doi:10.1080/13506280544000084
- 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, 209–245. doi:10.1080/09541440244000076
- Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, 12, 590–606. doi:10.3758/BF03213348
- Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, 36, 371–383. doi:10.3758/BF03195584
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: Erlbaum.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale: Academic Press.
- 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, 1072–1082. doi:10.1080/17470210701315467
- Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, 40, 791–794. doi:10.3758/BRM.40.3.791
- Cortese, M. J., & Schrock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, 66, 946–972. doi:10.1080/17470218.2012.722660
- Cuetos, F., & Barbón, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, 18, 415–436. doi:10.1080/13594320500165896
- Cuetos, F., Ellis, A. W., & Álvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, 31, 650–658. doi:10.3758/BF03200741
- Cuetos, F., González-Nosti, M., & Martínez, C. (2005). The picture-naming task in the analysis of cognitive deterioration in Alzheimer's disease. *Aphasiology*, 19, 545–557. doi:10.1080/02687030544000010
- Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. *Neuropsychologia*, 48, 3329–3334. doi:10.1016/j.neuropsychologia.2010.07.017
- Cuetos, F., Samartino, T., & Ellis, A. W. (2012). Age acquisition norms from elderly Spanish people: Characteristics and the prediction of word recognition performance in Alzheimer's disease. *Psicológica*, 33, 59–76. Retrieved from www.uv.es/psicologica/articulos1.12/4CUETOS.pdf

- Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & Cognition*, 41, 297–311. doi:10.3758/s13421-012-0263-8
- Davies, R., Wilson, M., Cuetos, F., & Burani, C. (2014). Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies? *Quarterly Journal of Experimental Psychology*, 67, 1808–1825. doi:10.1080/17470218.2013.872155
- De Bleser, R., & Kauschke, C. (2003). Acquisition and loss of nouns and verbs: Parallel or divergent patterns? *Journal of Neurolinguistics*, 16, 213–229.
- De Deyne, S., & Storms, G. (2007). Age-of-acquisition differences in young and older adults affect latencies in lexical decision and semantic categorization. *Acta Psychologica*, 124, 274–295. doi:10.1016/j.actpsy.2006.03.007
- Della Rosa, P. A., Caticcalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract–concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, 42, 1042–1048. doi:10.3758/BRM.42.4.1042
- Development Core Team, R. (2014). *R: A language and environment for statistical computing (Version 3.1.0) [Software]*. Vienna: R Foundation for Statistical Computing. Retrieved from www.r-project.org/
- Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1103–1123. doi:10.1037/0278-7393.26.5.1103
- 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, 1049–1054. doi:10.3758/BRM.40.4.1049
- Gerhand, S., & Barry, C. (1999). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, 27, 592–602.
- Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, 115, 43–67. doi:10.1016/j.actpsy.2003.11.002
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12, 395–427. doi:10.3758/BF03201693
- Gilhooly, K., & Watson, F. (1981). Word age-of-acquisition effects: A review. *Current Psychological Reviews*, 1, 269–286.
- Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition*, 13, 884–910. doi:10.1080/13506280544000093
- Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, 133, 238–248. doi:10.1016/j.cognition.2014.06.015
- Juhász, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, 131, 684–712. doi:10.1037/0033-2909.131.5.684
- 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, 846–863.
- Keuleers, E. (2013). *vwr: Useful functions for visual word recognition research*. Retrieved from <http://CRAN.R-project.org/package=vwr>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978–990. doi:10.3758/s13428-012-0210-4
- Lambon Ralph, M. A., & Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition*, 13, 928–948.
- Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology*, 92, 322–344.
- Lotto, L., Surian, L., & Job, R. (2010). Objective age of acquisition for 223 Italian words: Norms and effects on picture naming speed. *Behavior Research Methods*, 42, 126–133. doi:10.3758/BRM.42.1.126
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, 7, e43230. doi:10.1371/journal.pone.0043230
- Marques, J. F., Fonseca, F. L., Morais, A. S., & Pinto, I. A. (2007). Estimated age of acquisition norms for 834 Portuguese nouns and their relation with other psycholinguistic variables. *Behavior Research Methods*, 39, 439–444. doi:10.3758/BF03193013
- Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 183–206. doi:10.1037/0278-7393.28.1.183
- Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, 63, 506–525. doi:10.1016/j.jml.2010.08.003
- Moreno-Martínez, F. J., Montoro, P. R., & Rodríguez-Rojo, I. C. (2014). Spanish norms for age of acquisition, concept familiarity, lexical frequency, manipulability, typicality, and other variables for 820 words from 14 living/nonliving concepts. *Behavior Research Methods*, 46, 1088–1097. doi:10.3758/s13428-013-0435-x
- 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. *Quarterly Journal of Experimental Psychology*, 50A, 528–559. doi:10.1080/027249897392017
- Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 116–133. doi:10.1037/0278-7393.21.1.116
- Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, 91, 167–180.
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, 20, 705–714. doi:10.3758/BF03202720
- Pind, J., Jónsdóttir, H., Gissurardóttir, H., & Jónsson, F. (2000). Icelandic norms for the Snodgrass and Vanderwart (1980) pictures: Name and image agreement, familiarity, and age of acquisition. *Scandinavian Journal of Psychology*, 41, 41–48. doi:10.1111/1467-9450.00169
- Qualtrics, L. L. C. (2014). *Qualtrics [Software]*. Provo: Author.
- Raman, I. (2006). On the age-of-acquisition effects in word naming and orthographic transparency: Mapping specific or universal? *Visual Cognition*, 13, 1044–1053.
- Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods*, 46, 1108–1118. doi:10.3758/s13428-013-0426-y
- Schröder, A., Gemballa, T., Ruppert, S., & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods*, 44, 380–394. doi:10.3758/s13428-011-0164-y
- Schröder, A., Kauschke, C., & De Bleser, R. (2003). Messungen des Erwerbsalters für konkrete Nomina [Different measures of age-of-acquisition for German nouns]. *Neurolinguistik*, 17, 83–114.

- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215. doi:[10.1037/0278-7393.6.2.174](https://doi.org/10.1037/0278-7393.6.2.174)
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*, 598–605. doi:[10.3758/BF03193891](https://doi.org/10.3758/BF03193891)
- Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age-of-acquisition effect? *Psychonomic Bulletin & Review*, *15*, 70–74. doi:[10.3758/PBR.15.1.70](https://doi.org/10.3758/PBR.15.1.70)
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78. doi:[10.1207/s15516709cog2901_3](https://doi.org/10.1207/s15516709cog2901_3)
- Suchodoletz, W. V. (2010). Daten zum Wortschatz von 20–25 Monate alten Kindern. Retrieved from www.kjp.med.uni-muenchen.de/sprachstoerungen/DzS.php
- van Loon-Vervoom, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse: Swets & Zeitlinger.
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*, 534–538. doi:[10.3758/BRM.41.2.534](https://doi.org/10.3758/BRM.41.2.534)
- Walley, A. C., & Metsala, J. L. (1992). Young children's age-of-acquisition estimates for spoken words. *Memory & Cognition*, *20*, 171–182. doi:[10.3758/BF03197166](https://doi.org/10.3758/BF03197166)
- Weekes, B. S., Davies, R., Wade, A., & Bradley, V. (2004). Verb transformation in non-fluent progressive aphasia. *Brain and Language*, *91*, 108–109.
- Wilson, M. A., Cueto, F., Davies, R., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: The role of item imageability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1842–1859. doi:[10.1037/a0033090](https://doi.org/10.1037/a0033090)
- Wilson, M. A., Ellis, A. W., & Burani, C. (2012). Age-of-acquisition affects word naming in Italian only when stress is irregular. *Acta Psychologica*, *139*, 417–424. doi:[10.1016/j.actpsy.2011.12.012](https://doi.org/10.1016/j.actpsy.2011.12.012)
- Winters, J. J., Jr., Winter, L., & Burger, A. L. (1978). Confidence in age-of-acquisition estimates and its relationship to children's labeling performance. *Bulletin of the Psychonomic Society*, *12*, 361–364.
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. doi:[10.3758/PBR.15.5.971](https://doi.org/10.3758/PBR.15.5.971)
- Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*, 1–29. doi:[10.1006/jmla.2001.2834](https://doi.org/10.1006/jmla.2001.2834)
- Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, *12*, 413–430. doi:[10.1080/09541440050114570](https://doi.org/10.1080/09541440050114570)