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

Julia M. H. Birchenough*

Oxford Brookes University

Robert Davies

Lancaster University

Vincent Connelly

Oxford Brookes University

Author Note

Julia M. H. Birchenough, Department of Psychology, Social Work and Public Health, Oxford Brookes University; Robert Davies, Department of Psychology, Lancaster University; Vincent Connelly, Department of Psychology, Social Work and Public Health, Oxford Brookes University.

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.

Correspondence concerning this article should be addressed to Julia M. H. Birchenough, Department of Psychology, Social Work and Public Health, Oxford Brookes University, Headington Campus, Gipsy Lane, Oxford, OX3 0BP. E-mail: 12085385@brookes.ac.uk.

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). Whilst 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. Words were presented in lists of 140 words and participants rated the age in years that they first learned each word. A split-half correlation testified to a high internal reliability. There were high correlations with rated AoA values for sub-sets of items collected in previous studies in German and English. Age and gender were found to influence ratings very weakly: older and male participants tended to give slightly higher age ratings. Education, multilingualism and frequent usage of additional languages other than German did not exert an influence on rating values. These new ratings will extend current existent norms available for language and reading research across languages and will provide researchers with a wider choice of word stimuli. Ratings are available in two measurements, age-in-years and AoA rated on a 7-point-Likert scale.

Keywords: *Age-of-acquisition, AoA, ratings, norms, estimates*

Words (abstract): 198

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

Age-of-acquisition (AoA) denotes the age at which a word and its meaning is first learned (Carroll & White, 1973). There is accumulating evidence that AoA exerts an influence on cognitive processes. Early acquired words are processed faster compared to 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 & Schock, 2013), lexical decision (Cortese & Khanna, 2007), in natural reading paradigms employing eye-tracking (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 late acquired words (De Bleser & Kauschke, 2003; Weekes, Davies, Wadey, & Bradley, 2004; for a recent review also see Brysbaert & Ellis, 2015). In Alzheimer patients, words which are learned earlier in life have also been found to remain intact for a longer time for word production (Cuetos, Gonzalez-Nosti, & Martínez, 2005) and word recognition (Cuetos, Herrera, & Ellis, 2010). AoA is now a widely accepted effect evident in lexical processing (Juhasz, 2005) and learning mechanisms more generally (Catling, Dent, Preece, & Johnston, 2013; Stewart & Ellis, 2008). AoA is thus critical to the investigation of language behaviour 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. Some of these have been very extensive ratings collections for 30,000 words in English (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) and 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 & Vanderwart (1980) picture set 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, Gissuraadóttir, & Jónsson, 2000), Spanish (Cuetos, Ellis, & Alvarez, 1999), and German (Schröder, Kauschke, & De Bleser, 2003). Finally, a selection of 824 rated AoA estimates is available in German (Schröder, Gemballa, Ruppin, & 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 neighbours than words which are learned later in life (Morrison, Chappell, & Ellis, 1997; Stadthagen-Gonzalez & Davis, 2006). The high inter-correlation 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 & Ellis (1995) reported faster word naming for early acquired than late acquired words (AoA effect) when frequency was taken into account, but 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 & Schock, 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 compared to high-frequency words (Bonin, Chalard, Méot, & Fayol, 2001; Gerhand & Barry, 1999). Importantly, Brysbaert & 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, & Ellis (1997) found both AoA and frequency effects in picture naming which 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 for example Chalard & Bonin, 2006). The semantic locus hypothesis (Brysbaert et al., 2000; van Loon - Vervoorn, 1989) proposes that AoA effects in lexical processing occur because the semantic system is organised according to when meanings are acquired. Early-acquired words thus have an advantage and produce faster meaning associations and faster categorisation than later acquired words. This is consistent with the model of semantic growth presented by Steyvers & Tenenbaum (2005) in which AoA functions as an organising 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 & Lambon Ralph (2000) presented a series of simulation studies, where a connectionist system was presented with patterns for learning in a cumulative and interleaved fashion. The simulations reported by Ellis and Lambon Ralph (2000), and later by P. Monaghan & 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 favour. 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 learnt less securely than earlier acquired items.

However, in a review of the computational evidence for AoA effects in reading development, Zevin & Seidenberg (2002) identified the arbitrariness of the input-output patterns deployed in the Ellis & 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 (2002) 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 (2000) because input-output patterns were random and input-output mappings were arbitrary. Zevin and Seidenberg (2002) 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 behavioural studies, J. Monaghan & Ellis (2002) showed that the AoA effect in naming was greater for words with inconsistent compared to consistent spelling-sound mappings. P. Monaghan and Ellis (2010) 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 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 where 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, & 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, 2012; 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 which 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 discussed in the foregoing, 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). Whilst picture naming ages are considered to be more objective estimates, they are also more difficult to obtain for non-picturable 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 judgement. 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. 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 rated acquisition ages were more similar to children's picture naming ages than to spontaneous speech production, suggesting that rated AoA does not indicate veridical first acquisition age but reflects a more advanced form of lexical knowledge. Comparing rated and objective AoA as predictors of picture naming latencies, Chalard et al. (2003) found that although both rated and objective AoA were predictive of latencies, objective AoA accounted for more variability than rated AoA. Morrison & 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 conceptualise rated AoA as reflecting the order of acquisition. Kuperman et al. (2012) compared their AoA ratings with an objective measure which 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 three years of life, and highlighted that this is congruent with infantile amnesia. Kuperman et al. also observed a tendency of 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 actual ages when words are learned. In this light, recent research on AoA tends 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 ratings studies have examined the impact 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 ratings from young participants (Cuetos et al., 2012). In contrast, Schröder et al. (2012) compared their German ratings of older participants with that of the younger group and did not find an age-related difference. Mixed results have also been found with regard to gender. Whilst in Kuperman's sample, female respondents tended to rate words as later learned than male respondents, the reverse was found in other studies conducted in German (Schröder et al., 2003) and English (Winters Jr, Winter, & Burger, 1978). With regard to the influence of education level, Kuperman and colleagues did not find a main effect of education, but noted that respondents with higher education tended to give earlier word learning ages than participants with lower education, although the effect size was very small. Multilingualism was also assessed by Kuperman et al. (2012), but few participants had been raised with more than one language and ratings did not differ from the rest. It thus appears that AoA ratings weakly correlate with age and may be influenced - with varying direction - by gender, but seemingly not by education or multilingualism. With the aim to ensure utmost comparability to other ratings, this study collected rater characteristics and their relative influence on rating values were determined.

In the following 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 which resembled as closely as possible that employed by Kuperman et al. (2012) to collect the current largest AoA dataset of 30,000 ratings in English 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 been previously highlighted by other researchers (Schmidtke, Schröder, Jacobs, & Conrad, 2014) and the collection of German AoA ratings reported in the present article endeavoured to broaden researchers' stimuli choice.

Method

Ratings were collected using online questionnaires (Qualtrics, 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 to attract further participants.

Participants

Eligibility for participation was confined to native German speakers, who were aged 18 years or over, and at the time of participation 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 only complete the ratings section, if they so wished. First of all, participants were asked to indicate their age and gender. Education level was probed twofold, both as number of years receiving education, including primary schooling, and as a formalised 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 post-graduate studies), *nicht hier aufgeführt* (not listed). Respondents' language environment was assessed by asking to indicate if 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, respondents were invited to take part in a prize draw to win one of five EUR25 – 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 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 (Duden, http://www.duden.de/suchen/dudenonline/allein). Words were 2-17 letters long (M = 6.73), and had 1-7 phonological syllables (M = 2.25). Frequency ranged between 0 and 2,839.84 (M = 33.31) occurrences per million (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 (Võ 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.

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 ANOVA showed that questionnaires did not differ in terms of log10 frequency, $F(1,3238) = 0.07$, $p = .79$. In each on-line questionnaire, the 120 items were distributed over 12 consecutive question pages. In order to vary the items that were presented together on one question page, we created three pseudo-randomized 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 randomisation occurred at two levels. The order of the question pages within each questionnaire and the order of words within each question page were randomised for each participant.

The 20 control words (see Appendix 1) were chosen from the original word set for all participants to rate, and appeared in all questionnaires. Kuperman et al. (2012) had also used a set of 52 control words to be rated by all participants. In the current study, ratings for control words were collected to have ratings given by all participants on the same 20 items, which could then be used to 1) identify outliers and 2) to investigate the influence of rater differences on AoA ratings. Items were selected to mirror a range of acquisition ages, whilst

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 & Davis (2006) and Kuperman et al. (2012); instructions are reproduced in Appendix 2. 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, we report how 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 reliability. 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 participants who contributed to the present ratings. Lastly, we examine whether raters' characteristics influenced AoA values and explored 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 which were used within data analysis. We distinguish between raw ratings, which denote a single observation of one participant (one rating from one participant for one word), and AoA ratings/estimates, which are all averaged raw ratings from all participants for that

word (our actual AoA norms). We will also refer to control word means (CWMs), which are the average of 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 non-valid entries. Nine raw ratings were deleted because they did not clearly indicate an age. Where the age of learning had been given but additional text had been 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, 179 "x" (unknown) responses were recorded, indicating that overall 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 age-of-acquisition was not known, and not that words were unknown. 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 the two values were neighbouring integers, i.e. 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 was

applied for 57 raw ratings. For six further raw ratings, where 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 be have been interpreted to mean 'lentil' or to refer to an optical lens, and it was assumed that these double entries had been made to capture the acquisition of different word meanings.

Due to an input error, one word was lost and 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), for some participants we removed all responses from the dataset because the AoA values they supplied appeared to be outliers in comparison to the responses supplied by other participants. The detection of outliers was based on data collected using control words. All participants had been asked to rate 20 control words, and for all but three participants (who had not provided 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 (CMWs). We then compared these to the sample's overall CWM. We identified four participants whose CMWs correlated below .4 with the mean. By removing these four participants' contributions, 459 raw ratings were excluded. Although the cut-off of

.4 is 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 of 537 participants to determine the average AoA ratings per word. The average correlation between the overall sample's CWM and each participants' CWM was $r = .83$.

Data Analysis

Age-of-acquisition ratings

Words (excluding control words) were rated on average 19.57 times (range = 7 to 27), a similar raw ratings per word ratio to that recorded for the Kuperman et al. (2012) and Schröder et al. (2012) datasets. (The twenty 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 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 10 words which were rated as 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 which were the most often marked as unknown were *Quorum* (quorum), *Pomp* (pomp), *Alkoven* (alcove), *Charta* (charter), *Jade* (jade), *Noblesse* (nobility) and *Viola* (viola).

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

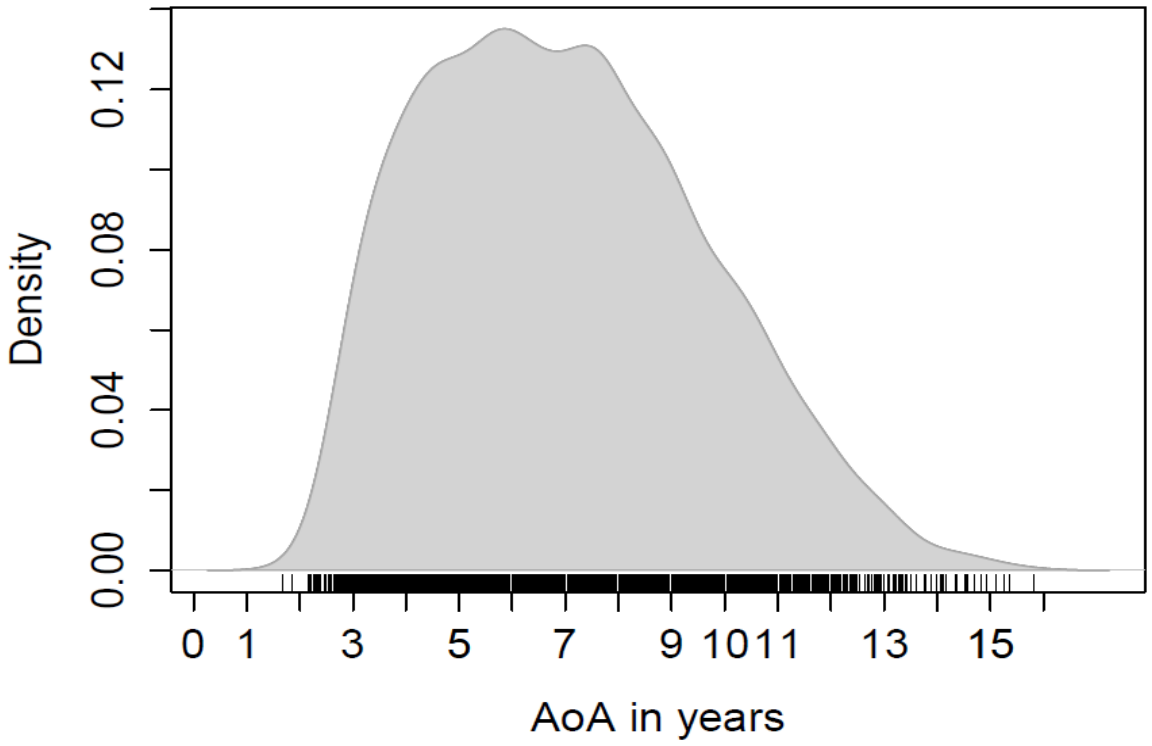

Figure 1. Distribution of AoA estimates in years for 3,259 German words

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. 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 AoA values for these words in the Schröder et al. and in our data were highly correlated, r_s = .91, p < .001. In the present study, 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) compared to Schröder's ratings.

The present ratings were also compared to AoA estimates which had been collected in English. Table 1 shows that 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 & Khanna (2008) norms, and 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 also correlated highly, r_s=.77, p <.001. (Note that we compared Likert-scale AoA estimates from the Bristol and the Cortese & Khanna norms with AoA estimates converted to Likert-scale values in our norms.)

Table 1

Correlation coefficients of present German age-of-acquisition (AoA) norms to extent rating collections

Ratings study/collection	*Language*	*Measurement*	*Number of words in common*	*correlation coefficient Spearman's rho*
Schröder et al. (2012)	German	Likert (1 - 7 scale)	203	.91***
Kuperman et al. (2012)	American English	years	2,683	.74***
Cortese & Khanna (2008)	American English	Likert (1 - 7 scale)	782	.70***
Stadthagen-Gonzalez & Davies (2006; Bristol norms)	British English	Likert (1 - 7 scale)	1,031	.77***

Note: *** p < .001, ** p < .01, *p < .05

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 – 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' production word list and we found a moderate relationship of r = -.53 (p < .001) between the percentage of 20-25-month-olds who produced a word and our AoA estimates for those words (see Figure 2). Words estimated by adults as having been acquired earlier were more likely to have been produced by more children in spontaneous speech.

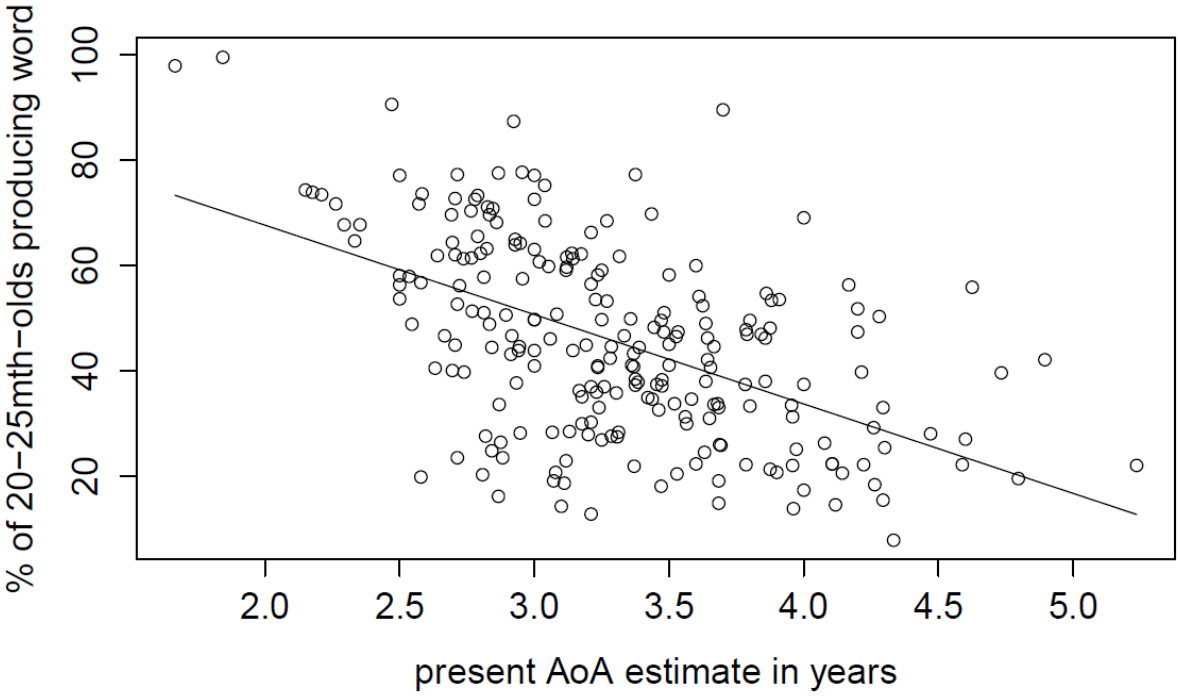

Figure 2. Scatterplot showing relationship between percentage of children aged 20-25 months producing a word and our present AoA ratings

Relationship of AoA with other word characteristics

As mentioned in the foregoing, previous research has shown that early acquired words tend to be shorter, more frequent and to have more orthographic neighbours. In order to inspect these relationships for our AoA estimates, a number of correlations with known word properties were performed. Frequency estimates (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 to capture the orthographic similarity neighbourhood characteristics of 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. As Coltheart's N has been argued to be insensitive to orthographic similarity neighbourhoods for longer words, for each word we also calculated the Orthographic Levensthein 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 neighbours. Both measures were computed using the R package vwr (Keuleers, 2013) based on the Clearpond word list of 27,751 DELETE most frequent DELETE END German words (Marian, Bartolotti, Chabal, & Shook, 2012). As expected, words that were acquired earlier tended to be shorter, more frequent, and to have more orthographic neighbours (see Table 2). Both length measures were only weakly correlated with AoA.

Table 2

Correlation coefficients for multiple comparisons of present age-of-acquisition (AoA) norms for German words to other word characteristics

	AoAestimate	log frequency	length in letters	syllable number	old20
AoAestimate					
log frequency	-0.53***				
length in letters	0.29***	-0.26***			
syllable number	0.36***	-0.26***	0.75***		
old20	0.41***	-0.40***	0.81***	0.66***	
ColtN	-0.35***	0.33***	-0.49***	-0.42***	-0.58***

Note: log frequency (SUBTLEX-DE frequency estimates; Brsybaert et al., 2011), old20 (Orthographic Levesthein distance; Yarkoni et al, 2008), ColtN (Coltheart's N; Coltheart et al, 1977)

*** p < .001, ** p < .01, *p < .05

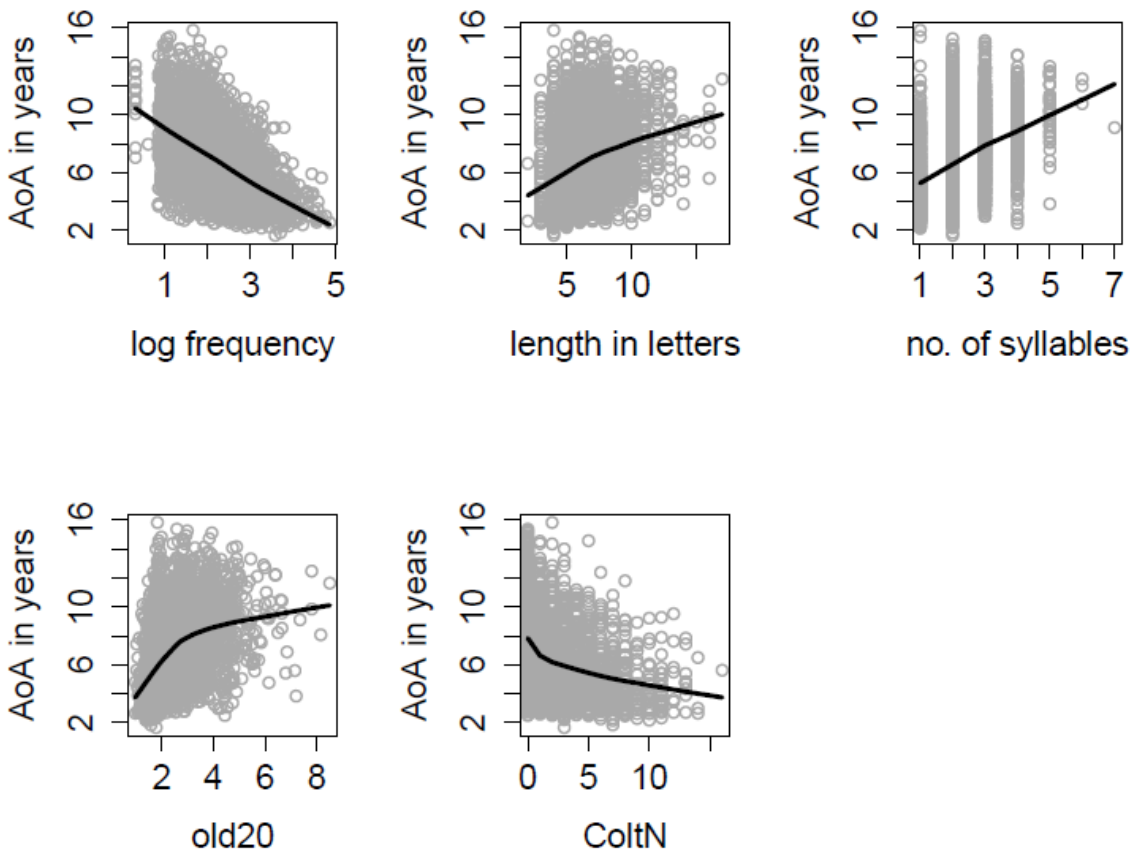

Figure 3. Scatterplots of AoA with lexical variables log frequency (SUBTLEX-DE frequency estimates; Brsybaert et al., 2011), length in letters, number of syllables, old20 (Orthographic Levensthein distance; Yarkoni et al, 2008), and ColtN (Coltheart's N; Coltheart et al, 1977). Lowess smoothers are shown.

A regression analysis showed that variance in German age-of-acquisition 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 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 variable word length in letters 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 earlier acquired when they were more frequent, shorter and more orthographically similar to other words.

When inspecting the scatterplots in Figure 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 was retained in the model because higher-order terms (like the quadratic term) cannot be estimated accurately without including lower order terms (Cohen, Cohen, Aiken & West, 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 syllable numbers. 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 fact that the inclusion of quadratic terms for all word characteristics was beneficial for the model means that although early acquired words tend to be shorter and have more neighbours, this relationship is attenuated for later acquired words.[1]

Table 3

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

Simultaneous Regression Analysis with present German age-of-acquisition (AoA) norms as outcome variable. Model 1 includes word characteristics log frequency, number of phonological syllables and orthographic similarity measure Orthographic Levensthein distance (old20) as predictor variables. Model 2 additionally adds quadratic terms for old20 and number of syllables.

	Model 1				*Model 2*			
	standardized coefficient	coefficient	*SE*	t-value	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 number²					0.13	0.08	0.04	2.23***
old20²					-0.55	-0.23	0.03	-8.76*
R^2		.34				.36		

Note: *SE* is Standard Error for coefficient. *** $p < .001$, ** $p < .01$, * $p < .05$

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 Figure 4. The age group between 18 and 20 years old comprised 6.21% of all participants. The largest participant group were 21 – 30-year-olds with 44.63%, followed by 19.4% who were between 31 and 40 years old. 14.5% were aged between 41 and 50 years, 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 aged 71 or older.

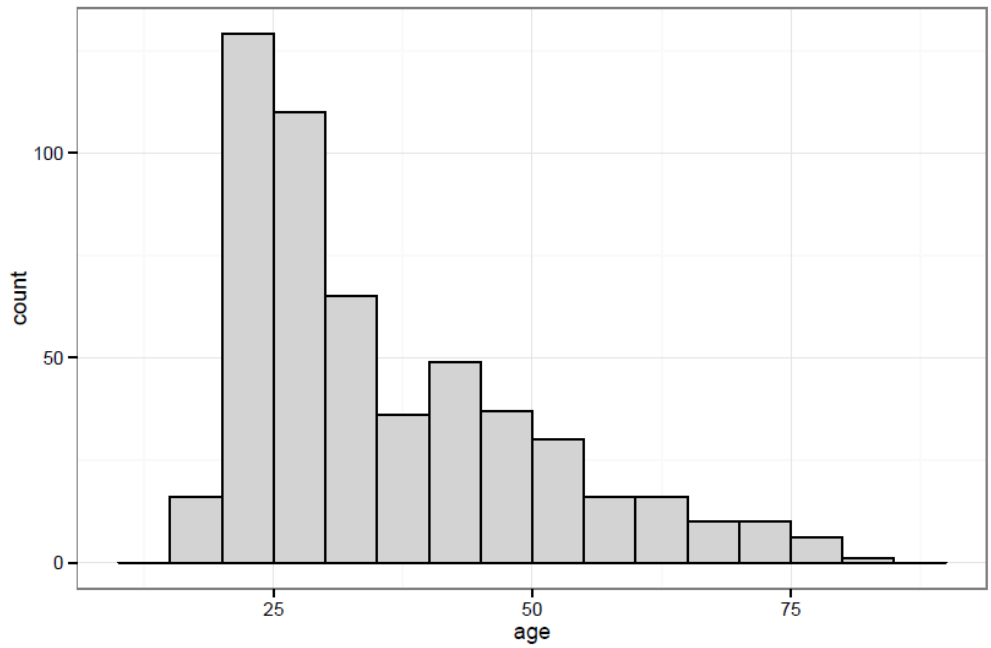

Figure 4. Age distribution of 531 participants who gave information about their age

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 secondary education or had not completed post-school 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 to post-graduate level. Finally, 4.66% felt that their education level was not represented by the choices given.

Participants had also been 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, $p < .001$). As more respondents had specified their formal level of education than number of years, the former measure was used for all further analysis.

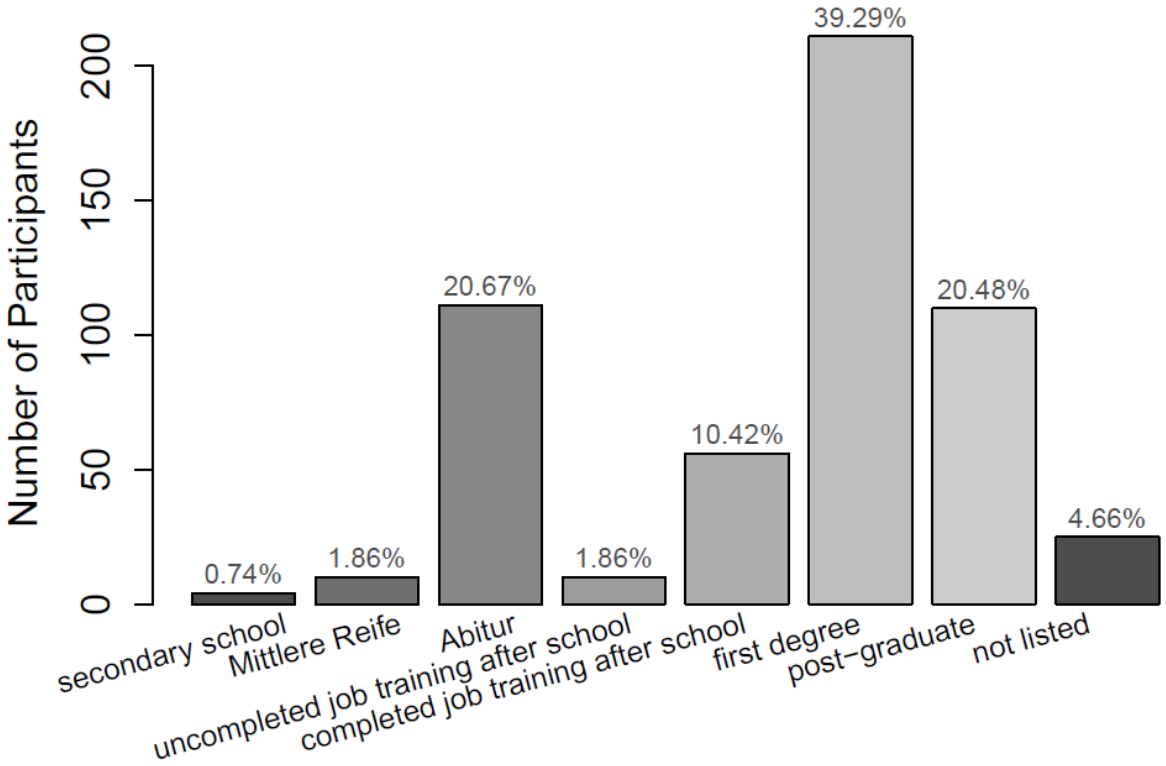

Figure 5. Percentage of 537 participants within formal levels of education. *Mittlere Reife oder Äquivalent* (school qualification at age 16 or equivalent), *Abitur oder Äquivalent* (school qualification at age 18 or equivalent).

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

Did ratings vary with rater characteristics?

In order 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 responses of participants who had rated all 20 control words and also given information 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 slightly later learned than younger participants. This concurs with findings by Kuperman and colleagues in their American English sample. We investigated this weak but significant correlation further and examined the correlation between age and raw control word ratings (the original AoA values for each control word given by each participant). Correlations varied between r = -.15 and .38. Control words which showed the strongest positive association with age were *Karriere* (r=.38, career), *Hubschrauber* (.33, helicopter) and *Urlaub* (.24, holidays), whilst 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. In order to examine whether the age correlation would disappear if 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 (control word means) 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 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 level (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 later acquired (see Table 4). However, there were very small participant numbers included in our sample for these education levels. A post-hoc Tukey multi-comparison between groups was conducted, where group was defined by education level and the observations were the CWMs (average AoA ratings of control words per participant). Participants in different education groups did not significantly differ in terms of CWMs.

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
Mittlere Reife or equivalent	7	1.55	6.64	1.86
Abitur 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
post-graduate	88	19.43	5.64	1.61
not listed	21	4.64	6.23	1.34

Note: N = number of participants, total *N* = 453; % = percentage of 453 participants; CWM = Control Word Mean (averaged AoA ratings in years of 20 control words of 453 participants)

Participants with multilingual ($N = 40$, $M = 5.75$, $SD = 1.72$) and monolingual upbringing ($N = 413$, $M = 5.62$, $SD = 1.46$) did not differ in the rating values they gave to control words (CWMs, $t(44.667) = 0.46$, $p = .644$). Equally, whether respondents regularly 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 on CMWs ($t(239.928) = -1.21$, $p = .227$).

In order to verify these results based on the means of 20 control words, we entered the full raw data set into a mixed-effects model using lme4 (Bates, Maechler, Bolker & Walker, 2014). Eliminating empty fields resulted in reducing 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 with the Likelihood Ratio Test (LRT). The following shows the order in which predictor groups were added. Results comparing each model to its predecessor are given in brackets:

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

Model 3: syllable numbers^2 and old20^2 (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: age^2 (LRT, $\chi^2 = 9.5151$, 1*df*, $p = .002$);

Model 6: random slope for frequency (LRT, $\chi^2 = 1782.2$, 2*df*, $p < .001$)

In the final model (see Table 5, Model 6), the lexical variables log frequency, old20 and their quadratic terms syllable numbers^2, old20^2, and the participant characteristics gender, age as well as age^2 emerged as significant predictors. Adding a term to capture the effect of random participant variation in frequency 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 model fit was significantly improved by the inclusion of the random frequency effect of participants indicates that there is unexplained or unmeasured variation among 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 estimated to have a negative-signed coefficient, while the quadratic age effect was positive. As the inclusion of age^2 signifcantly improved the model, this may indicate that participant age only affected ratings in the older participants rather than all age groups. Whilst older participants showed a slight tendency to rate words as later learned, there did not seem to be a systematic variation amongst the younger participants. This finding concurs with our previous results when inspecting control word means. However, the coefficient estimates for age are 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 (Likelihood Ratio Test (LRT), $\chi^2 = 5.7562$, 7*df*, $p = .569$). Adding the interaction after the inclusion of the age^2 term into Model 5 also did not improve 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 non-significant when the age:formal_education 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 non-ordinal 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 possibly include participants who have spent less time in education than those who chose the preceding education level "*Abitur* or equivalent (A-levels or equivalent)". Consequently, formal education choices may then 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 a reduced number of 503 participants indicated that the effects of predictor variables remained the same while 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 learned earlier. Interestingly, gender lost its predictive power when we added an age:gender interaction term to Model 4. However, as there was no age:gender interaction effect, the interaction term was not kept in the model.

Table 5

Final mixed model summary table of word and participant characteristics, including quadratic terms. The model specified random intercepts for both words and participants, as well as random slopes for frequency over participants.

Fixed effects				Wald confidence intervals	
	Estimate	*SE*	*t*-value	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					
Groups			Variance	Std. Dev	Corr
Word (Intercept)			4.059	2.015	
Participant (Intercept)			5.995	2.448	
lgSUBTLEX			0.235	0.485	-0.87
Residual			3.796	1.948	

Note. Number of obs: 72836, groups: Word, 3259; Part, 528; For categorical fixed effects variables, estimates indicate effect with regard to the reference variable, which were "Version 1" (for Version of questionnaire), "male" (for gender), "secondary school" (for formal.education.level), "yes to multilingual upbringing" (for multilingual.upbringing), and "yes to frequent.other.language.usage" (for frequent.other.language.usage).

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, high internal and external reliability of the current norms have been 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. 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 the correlation.

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 production ages, this association nevertheless indicates the validity of the current 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 two years only. Furthermore, this lends support to the concept of rated AoA as an order of acquisition.

We used regression to explore which lexical variables and which participant characteristics accounted for variance in AoA estimates. In accordance with previous research (Stadthagen-Gonzalez & Davis, 2006), the current rated AoA estimates were influenced by word frequency and orthographic neighbourhood so that early learned words tended to be more frequent and to be more similar in spelling with other words. The curvilinear effect of word length in syllables indicates that later learned words tend to be longer than early learned words, and that this relationship becomes more pronounced as

words become lengthier. 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 impact of age, gender, education and general language usage on ratings. This was first examined by analysing participant characteristics effects on control word means (CWMs).

Consistent with 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. had 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 had only employed a small sample size. Walley & Metsala (1992) demonstrated that 5-year-old children give similar AoA estimates to adults.

In the current sample, age was confirmed as a predictor of AoA ratings, when added to the mixed-effects model on raw estimates. The fact that model fit was improved by the inclusion of the quadratic age term, corresponding to the nonlinear 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 later learned.

In the current sample, the correlation of participant age on AoA control word means disappeared when 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 patients. Cuetos and colleagues found that although ratings from elderly and young adults correlated highly, ratings also differed greatly for some words (e.g. robot, television), indicating that the

day-to-day vocabulary people use changes with time. In the present sample it may have been possible that some of the later acquired words (career, helicopter) were driving 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, as De Deyne and Storms (2007) found that participants, who were in their 50s, rated words as significantly later learned than young participants in their 20s, and that this was true for both common words and modern words. As suggested by an anonymous reviewer, the question of whether the small age effect present in the current data was due to a genuine tendency of older participants to rate words as later acquired or due to a generational difference in word acquisition (common versus 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 paper.

We found that men tended to rate words as a little later acquired than women. This is consistent with Winters Jr 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 current children's usage and knowledge of words. The inclusion of an age:gender interaction did not significantly improve the model. We note that the gender effect we observed is opposite in direction to that observed by Kuperman et al. (2012), while 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 current analysis, education level did not affect ratings, in line with Kuperman's findings, but we have highlighted that participant numbers were very unevenly distributed over education categories. We also examined if respondents' language environment and

usage influenced 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 current sample the frequent use of languages other than the mother tongue did not influence ratings given.

In the mixed-effects model, log frequency emerged as the strongest predictor for raw AoA estimates. It is therefore interesting to note that model fit was significantly improved by taking into account the random variation of the frequency effect between participants. This indicates that while 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 age-of-acquisition ratings with high internal and external reliability 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, multi-lingualism or frequent use of other languages. Although a weak correlation with age was found, this association may have been driven by modern words rated to be later acquired by older participants. This may inform sampling for future ratings collections.

Acknowledgement

The authors would like to thank all respondents for their participation.

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.

allein. (n.d.) In Duden Online . Retrieved from http://www.duden.de/suchen/dudenonline/allein

Álvarez, B. & Cuetos, F. (2007). Objective age of acquisition norms for a set of 328 words in Spanish. *Behavior Research Methods*, *39*, 377–383.

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.

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.

Barry, C., Morrison, C. M. & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*, 560–585.

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.

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.

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. & 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.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J. & Böhl, A. (2011). The word frequency effect. *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)*, *58*, 412–424.

Brysbaert, M. & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *The Quarterly Journal of Experimental Psychology*, *64*, 545–559.

Brysbaert, M., & Ellis, A. W. (2015). Aphasia and age of acquisition: are early-learned words more resilient?. *Aphasiology*, 1-24.

Brysbaert, M. & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition*, *13*, 992–1011.

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*(2), 215–226.

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.

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.

Catling, J., Dent, K., Preece, E. & Johnston, R. (2013). Age-of-acquisition effects in novel picture naming: A laboratory analogue. *The Quarterly Journal of Experimental Psychology*, *66*, 1756–1763.

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.

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.

Chumbley, J. I. & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, *12*, 590–606.

Clark, J. M. & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 371–383.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple correlation/regression analysis for the behavioral sciences* (3rd ed.). Mahwah , NY: Lawrence Erlbaum Associates

Coltheart, M., Davelaar, E., Jonasson, T. & Besner, D. (1977). *Access to the internal lexicon*. In S. Dornic (Ed.), Attention and Performance VI (pp.535 – 555). London: 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. *The Quarterly Journal of Experimental Psychology*, *60*, 1072–1082.

Cortese, M. J. & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, *40*, 791–794.

Cortese, M. J. & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology*, *66*, 946–972.

Cuetos, F. & Barbón, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, *18*, 415–436.

Cuetos, F., Ellis, A. W. & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, Instruments, & Computers*, *31*, 650–658.

Cuetos, F., Gonzalez-Nosti, M. & Martínez, C. (2005). The picture-naming task in the analysis of cognitive deterioration in Alzheimer's disease. *Aphasiology*, *19*, 545–557.

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.

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: International Journal of Methodology and Experimental Psychology*, *33*, 59–76.

Davies, R., Barbón, A. & Cuetos, F. (2012). Lexical and semantic age-of-acquisition effects on word naming in Spanish. *Memory & Cognition*, 1–15.

Davies, R., Wilson, M., Cuetos, F. & Burani, C. (2014). Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies? *The Quarterly Journal of Experimental Psychology*, *67*, 1808–1825.

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.

Della Rosa, P. A., Catricalà, 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.

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.

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.

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.

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.

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.

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.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684–712.

Juhasz, 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.

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*(4), 322-344.

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.

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.

Marian, V., Bartolotti, J., Chabal, S. & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, 7(8): 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.

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.

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.

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.

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*, 528–559.

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.

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.

Pind, J., Jónsdóttir, H., Gissurardóttir, H. & Jónsson, F. (2000). Icelandic Norms for the Snodgrass and Vanderwart (1980) Pictures: Name and Image Agreement, Familiarity, and Aage of Acquisition. *Scandinavian Journal of Psychology*, *41*, 41–48.

Qualtrics, L. (2014). Qualtrics [software]. Utah, USA: Qualtrics

Raman, I. (2006). On the age-of-acquisition effects in word naming and orthographic transparency: Mapping specific or universal? *Visual Cognition*, *13*, 1044–1053.

R Development Core Team (2014). R: A language and environment for statistical computing (Version 3.1.0) [software]. Vienna: R foundation for statistical computing. Retrieved from http://www.r-project.org/

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.

Schröder, A., Gemballa, T., Ruppin, S. & Wartenburger, I. (2012). German norms for semantic typicality, age of acquisition, and concept familiarity. *Behavior Research Methods*, *44*, 380–394.

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(2), 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.

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

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.

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.

Suchodoletz, W. v. (2010). *Daten zum Wortschatz von 20-25 Monate alten Kindern.* Retrieved from http://www.kjp.med.uni-muenchen.de/sprachstoerungen/DzS.php

Van Loon - Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse: Swets and Zeitlinger

Võ, M. L., 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.

Walley, A. C. & Metsala, J. L. (1992). Young children's age-of-acquisition estimates for spoken words. *Memory & Cognition*, *20*, 171–182.

Weekes, B. S., Davies, R., Wadey, A. & Bradley, V. (2004). Verb transformation in non-fluent progressive aphasia. *Brain and Language*, *91*, 108–109.

Wilson, M. A., Cuetos, 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.

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.

Winters Jr, J. J., 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.

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.

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*(3), 413-430.

Appendix 1

List of 20 control words

Control Word	**translation**
Milch	Milk
Tier	animal
Stuhl	chair
warm	Wam
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 2

Instructions given to participants. This is a direct translation into German from the instructions used by Stadthagen-Gonzalez & 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 3

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), 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	carrion	carrion	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