

AGE-OF-ACQUISITION RATINGS FOR 2816 DUTCH FOUR- AND FIVE-LETTER NOUNS

Mandy GHYSELINCK, Wendy DE MOOR, and Marc BRYLSBAERT
Ghent University

Studies on object and word naming have shown that the age at which words are acquired is an important factor in processing times. Research on the issue in Dutch has been hampered by the fact that only teacher ratings were available about which words should be known by 6-year-olds. As a supplement to these teacher ratings, we conducted a large-scale study in which 558 students rated the age-of-acquisition of 2816 four- and five-letter nouns. Reliability of the ratings is high, and correlations with word frequency and word imageability are in the same order as those reported for English.

Virtually all psycholinguistic researchers agree that word frequency (i.e., the number of times a person is likely to come across a particular word) is an essential variable in word processing: High frequency words are easier to process than low frequency words. This is true for all sorts of word processing tasks (e.g., word naming, lexical decision, perceptual identification) and no model of word recognition has a chance of being accepted in the literature if it does not account for the frequency effect (for a review, see Monsell, 1991). Also, experiments on the effects of other variables in word processing are unlikely to be published if word frequency has not been taken into account. As a result, researchers have invested major efforts to collect frequency norms for their language. In Dutch, for a long time the frequency norms of Uit den Bogaart (1975) were used. These were based on a corpus of 720,000 written words. Nowadays, the Uit den Bogaart corpus has been replaced by the electronic Celex Database (Baayen, Piepenbrock, & Van Rijn, 1993), which is based on a corpus of 42,380,000 written words. Similarly, in English the old Kuçera and Francis (1967) measures are currently being replaced by the Cobuild frequencies from the Celex Database. These frequency measures have their limitations (e.g., they are nearly all based on written corpora; and there are always choices to be made about which texts to include, which types of word derivatives to combine in the frequency measures, etc.), but in general it is thought that the

This research was made possible by a BOF-project granted to the authors by the Research Council of Ghent University. We thank three anonymous reviewers for their helpful suggestions on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Mandy Ghyselinck, Department of Experimental Psychology, Ghent University, H. Dunantlaan 2, 9000 Ghent. Electronic mail may be sent to mandy.ghyselinck@rug.ac.be

existing measures are reliable enough for practical purposes (at least in Dutch and English) and that further gains would not outweigh the efforts needed to collect additional data.

During the last five years or so, evidence is rapidly growing that the robust frequency effects in word processing tasks are actually a compound of two variables: the frequency with which words are encountered in texts (i.e., the usual interpretation of the word frequency effect), and the age at which the words have been acquired by individuals. Although it is true that most high frequency words have been acquired early in life and most low frequency words have been acquired late in life, giving rise to a strong positive correlation between word frequency and word age-of-acquisition (see below), there are exceptions (e.g., in English *puppy* is a low frequency word that is known by infants, whereas *income* is a high frequency word that is unlikely to be known by children). Using such words, Morrison and Ellis (1995) were able to disentangle the effects of word frequency and word age-of-acquisition (AoA), and they presented evidence that at least part of the usual frequency effect in word naming and lexical decision is due to AoA. These results have been replicated by Gerhand and Barry (1998, 1999a) and Turner, Valentine and Ellis (1998) for English, by Yamazaki, Ellis, Morrison, and Lambon-Ralph (1997) for Japanese, and by Brysbaert (1996) and Brysbaert, Lange, and Van Wijnendaele (2000) for Dutch.

Inspired by these findings, authors have started to examine AoA effects in other tasks. The importance of word AoA in addition to word frequency had already been suggested a long time ago for picture naming (Carroll & White, 1973), and new experiments with better controlled materials have indeed established that a large proportion of the variability in picture naming latencies is due to the age at which the object names have been acquired. This is not only true for English (Barry, Morrison, & Ellis, 1997; Ellis & Morrison, 1998), but also for Spanish (Cuetos, Ellis, & Alvarez, 1999) and French (Alario & Ferrand, 1999). Using a speeded naming task in which participants were instructed to name visual words much faster than they usually do, Gerhand and Barry (1999b) showed that the effect of AoA is stronger than in a normal naming task. Investigating the importance of AoA in the semantic system, van Loon-Vervoom (1989) and Brysbaert, Van Wijnendaele, and De Deyne (in press) reported significant AoA effects in a word association task in which participants were asked to produce the first associate that came to their mind when they saw a stimulus word. Brysbaert et al. additionally reported an AoA effect in a task in which participants had to decide whether a word belonged to the category of nouns with a definable meaning or to the category of first names. Finally, Lewis (1999) reported independent effects of frequency and AoA in a face categorisation task. In this task, participants had to indicate to which TV-soap pictures of characters

belonged. The variables that were manipulated were the time since the first appearance in the soap (AoA) and the average exposure time during an episode (frequency). Both variables had an effect.

Other research has indicated that AoA is unlikely to be a confound of a third variable. For instance, both Brysbaert, Lange, and Van Wijnendaele (2000) and Coltheart, Laxon, and Keating (1988) showed that the AoA effect in visual word processing is not due to the imageability of the words. Although AoA and imageability are intercorrelated (see below), the effect of AoA on word naming and lexical decision remained significant when stimulus lists were matched on both word frequency and word imageability.

Ghyselinck and Brysbaert (submitted) examined the correlations between word familiarity, word frequency, and word AoA, in an attempt to solve a long-lasting debate in the literature on word recognition. In 1984, Gerbsbacher showed that the subjective measure of word familiarity (based on the question: "how often have you come across this word?") explained significantly more variance in word processing times than word frequency. Ever since, there has been a debate about what exactly word familiarity measures in addition to word frequency (see e.g., Balota, 1994). Ghyselinck and Brysbaert's series of experiments strongly suggest that the missing variable is AoA, as the variance in word processing times explained by word familiarity equals the variance explained by word frequency plus the variance explained by word AoA.

Finally, there are theoretical reasons to expect AoA as a crucial variable in word processing in addition to word frequency. Lewis (1999), for instance, pointed out that AoA and frequency are two different measures of how often a person has encountered a particular word. The amount of experience with a word that has been acquired early in life will on the average be greater than the amount of experience with a word that has been acquired only recently. According to Lewis, it may very well be that not only the recent experiences with a word counts, but the cumulative frequency of all encounters with a particular stimulus (which would be evidence for an instance-based organisation of the mental lexicon). This line of reasoning agrees with Ghyselinck and Brysbaert's finding of the relationship between word familiarity, word frequency, and word AoA. On the basis of simulations with connectionist networks, however, Ellis and Lambon Ralph (in press) reported that the effect of AoA is likely to go beyond mere cumulative frequency. They argued that words learned first by a network, have a privileged status in the network because they can be implemented in all possible units and connections between units, and because early training makes for larger weight changes than later learning. The activation function in a typical connectionist network follows a sigmoid curve with small changes towards the extremes and large changes in the middle of the curve.

Due to this characteristic, words that are learned when the network is still in the middle range (i.e., the initial, untrained state) have a larger impact than words learned when the connection weights are already shifted to one of the extremes. This loss of plasticity in connectionist networks results in more dispersed representations and higher error scores for later acquired words than for earlier acquired words, and the effect remains, despite considerable differences in frequency of occurrence between words. This analysis also suggests that AoA effects will be present in all processing stages that are based on connectionist learning principles (in the first studies, researchers tended to limit the effect of AoA to either the speech output system – Morrison & Ellis, 1995; Gerhand & Barry, 1998 – or to the semantic system – van Loon-Vervorm, 1989; Brysbaert, Van Wijnendaele, & De Deyne, in press).

Thus there are both empirical and theoretical reasons to believe that AoA is of equal importance in word processing as word frequency. This puts researchers in a similar situation as in the 1950s when the frequency effect was first demonstrated (Howes & Solomon, 1951): How to determine the age at which words are acquired?

Different AoA Measures

A first problem with AoA measures is that words are unlikely to be acquired upon their first encounter (see, e.g., Elbers & van Loon-Vervorm, 1998). Hence, there may be different measures of AoA, such as the age at which a child understands the meaning of a word in a command or a question, the age at which the child speaks the word, or the age at which the child is able to give a definition of the word. Also we may wonder whether we must make a distinction between the auditory (hearing, speaking) and the visual modality (reading, writing).

A second problem is how to define the population: Do children in the East of Flanders have the same experiences as children in the West of Flanders? Is there a difference between Flanders and the Netherlands (see below)? What about the influence of dialects spoken at home?

A third problem is related to the actual measurement: How to assess the age at which children in a certain region understand/speak/are able to define a particular word? How to cope with cohort differences? (In a very strict sense, the AoAs established for one group are only valid for that particular group and not, for instance, for a group of students that are 15 years older; see e.g., De Moor, Ghyselinck, & Brysbaert (2000) for such cohort effects).

Finally, a language typically consists of a few tens of thousands of words. This makes it virtually impossible to assess AoAs for all words.

It is clear from the preceding list of issues (which is by no means exhaustive) that one will never be able to conduct empirical research on AoA (or for that matter, on any word variable, see the discussion of word frequency above) if one is not willing to take into account some practical considerations. The first consideration is that, although there may be different definitions of AoA, the resulting measures are likely to be highly intercorrelated. This means that conclusions based on one measure will not be very different from those based on another.

The second consideration is that, at present, researchers are not looking for differences between words that are acquired at the age of, say, 4 years and words that are acquired at the age of 4 years and 1 month, just like researchers are not looking for frequency differences between 1 occurrence per million and 2 occurrences per million. Usually, researchers are comparing words that have been acquired early in life (e.g., before the age of 6) with words that have been acquired late in life (e.g., after the age of 10). This means that some unreliability of the measure can be tolerated.

A third practical consideration is that in English very much the same results have been obtained in the North of Scotland (Aberdeen; Gilhooly & Logie, 1982), in the North of England (York; Morrison & Ellis, 1995), and in Wales (Cardiff; Gerhand & Barry, 1998) all on the basis of the same set of AoA measures (see below). Even though Dutch speakers may be very impressed by the differences in spoken language over a distance of 100 km, we should not forget that the differences between the English speaking parts of the world are likely to be larger. Still, Bell, Davies, Hermann, and Walters (2000) showed a very high correlation ($r = .80$) between their adult ratings of AoA (i.e., American English) on a set of pictures and the objective AoA measures collected by Morrison, Chappell, and Ellis (1997) (i.e., British English). An even higher correlation was found between their ratings and the Morrison et al. adult ratings ($r = .90$). This indicates that when it comes to AoA, regional differences are neglectable; ratings obtained in the East of Flanders are likely to be representative for Flanders and the Netherlands.

Finally, the English research has shown that as long as one looks at gross differences in AoA, virtually the same results are obtained when one uses AoA estimates based on student ratings as when one tries to collect more detailed and objective measures. Probably the most ambitious study in this respect was done by Morrison, et al. (1997). These authors tested 280 children from 30 to 131 months individually and asked them to name 297 pictures from the Snodgrass and Vanderwart (1980) set. AoA of a word (e.g., *apple*) was defined as the age at which 75% of the children could name the picture correctly (which in the case of apple was 22.1 months). Morrison et al. also asked 20 students to rate the AoA of the words on a 7-point scale, which ran from 1 = learned at the age of 0-2 years, to 7 = learned at the age

of 13 or more (see Gilhooly & Logie, 1980). This resulted, for example, in a score of 1.80 for *apple*. The correlation between both AoA measures was 0.75. In addition, further research showed that both measures had virtually the same correlation with object naming times (Ellis & Morrison, 1998). On the basis of these and other studies (e.g., Gilhooly & Gilhooly, 1980; Lyons, Teer, & Rubenstein, 1978), Morrison et al. (1997) concluded that student ratings are a valid measure of AoA (see also De Moor et al., 2000). Therefore, most work on AoA effects in English has relied on the student ratings published by Gilhooly and Logie (1980). These authors asked 36 students to rate the AoA of different words on the 7-point scale described above. To assess the reliability of their ratings, Gilhooly and Logie divided the participants randomly in two groups, balanced for sex, and found an intergroup correlation of .98.

The studies done in Dutch so far all made use of a different AoA measure. In 1981, Kohnstamm, Schaeerlaekens, de Vries, Akkerhuis, and Froominckx published a book that indicated for 6,785 words to what extent each word should be known by 6-year-olds (see also Schaeerlaekens, Kohnstamm, & Lejaegere, 1999, for a very recent update of the list). The estimates were obtained by asking a representative sample of teachers from the Netherlands and the Dutch speaking part of Belgium to mark for each word whether it should be known by a 6-year-old or not. For the Belgian data, 40 teachers of kindergarten and 41 teachers of the first year of primary school from all over the Dutch speaking region took part in the assessment. The advantages of the Kohnstamm et al. measure, defined as the percentage of teachers indicating that a word should be known by a pupil who starts primary school, are that the measure is based on persons who have daily experience with the children they are judging, and that the measure has been obtained some 20 years ago (when the undergraduates of current studies were born). In addition, van Loon-Vervoom (1989) obtained a correlation of -0.92 between the Kohnstamm et al. measure and student ratings on an 8-point scale, based on 44 nouns.

The Kohnstamm et al. ratings have proven very useful in several studies, but they are limited because they only provide information about one moment in time: the transition from kindergarten to primary school. There is no information about whether a word that should be known by a 6-year-old, has been acquired at the age of 2 or at the age of 5; similarly, a word that is not known at the age of 6 according to the teacher ratings, can be acquired at the age of 7 or at the age of 15 and later. Also, the different measures used in English and in Dutch studies make it difficult to interpret deviating findings between both languages. Therefore, we decided to collect student ratings of AoA for the Dutch language as well.

Method

Stimulus Materials and Rating Procedure

We selected 2816 four- and five-letter nouns from the CELEX Database (Baayen et al., 1993). The selection was limited to four- and five-letter nouns for practical reasons and because we typically use words of these lengths in our studies of visual word processing. Nouns were chosen partly because they can be used in picture naming experiments. Words with a frequent noun interpretation and nouns with multiple frequent meanings or with a frequency less than 1 per 42,380,000 were excluded.

Of the stimulus set, we created 10 lists of 281 or 282 words that were matched on frequency and word length. Because our previous experience had shown that rating scales like Gilhooly and Logie's (1980) were sometimes confusing for the participants, we simply asked them to indicate for each word from which age they estimated they knew the word. If they did not know the word, they could write an "N". Of each list, three different permutations were made to minimise sequence effects. The lists were handed out at the beginning of a course, and completion of a list took about half an hour. Each participant completed but one list.

Participants

Participants were 558 undergraduates (310 females and 248 males). They were students from the faculties Political and Social Sciences, Criminal Sciences, Philosophy or Moral Sciences at Ghent University. All were native Dutch speakers. Average age was 19 (range 16-42).

Results and Discussion

The AoA data are shown in the Appendix. The full matrix of AoA, frequency, Log(frequency) and % of answers is available on the internet http://allserv.rug.ac.be/~hnaessen/vakgroep/documents_data_etc.

For each list we computed the correlation between the individual ratings and the mean AoA measures. Sixteen participants who correlated less than .60 with the means were excluded from further analyses. All in all, the minimal number of raters per list was 50. To assess the reliability of our ratings, we calculated the intraclass correlation of Shrout and Fleiss (1979). For this analysis, words that were not known by at least 80% of the

participants were excluded and the missing AoA values for the remaining words were estimated on the basis of the means of the rows and the columns. The reliability of the individual lists varied from .95 to .98, and the reliability of the total stimulus set was .98.

To further check the reliability and validity of our AoA measure, we computed the correlation between our figures and the figures that had been collected by Lange¹ in 1995. She asked 43 psychology students from the University of Leuven to indicate for 180 four- and five-letter words at which age they thought they had acquired them, using Gilhooly and Logie's (1980) rating scale. Reliability of Lange's AoA ratings was .96. Figure 1 shows the correlation between our measure and hers for the 121 words in common. The correlation amounted to .90, which is nearly the maximum that can be obtained, given the measure reliabilities of .98 and .96. This analysis shows (1) that the student ratings are stable over time and over geographical location, and (2) that the rating scale we used, returns the same data as Gilhooly and Logie's.

Figure 2 shows the correlation between Log(frequency) and AoA ($r = -0.59$). The reported correlation coefficients between frequency and rated AoA in English range from -.40 (Rubin, 1980) to -.71 (Gilhooly & Logie, 1982).

We also correlated the teacher ratings of Kohnstamm et al. (1981) with our AoA ratings. This correlation of -.80 ($N = 756$; see Figure 3) shows that the Kohnstamm et al. (1981) measures are indeed a good alternative for student ratings, as already suggested by van Loon-Vervoom (1989; see above) and our own previous work with the Kohnstamm et al. measure (Brysbart, 1996; Brysbart et al., 2000; Brysbart et al., in press).

van Loon-Vervoom (1985) had all the words of the Kohnstamm et al. (1981) list rated on a 7-point scale for imageability. Thus we could also correlate our AoA ratings with this imageability measure (Figure 4). The correlation we obtained was rather low: $r = -.36$ ($N = 756$). This is very similar to the correlation of -.38 reported by Morrison et al. (1997), but unlike the correlation of -.72 reported by Gilhooly and Logie (1980).

Finally, we compared our AoA ratings with the English measures for words that could be considered as unequivocal translations (e.g., *slang-snake*). The correlation with Morrison et al.'s (1997) AoA measure based on individual tests with children was .60 ($N = 113$), the correlation with Gilhooly and Logie's (1980) student ratings was .71 ($N = 377$). These correlations show that AoA measures in different languages converge, as one could expect from the multiple interactions between the English and the Dutch cultures.

¹ Personal communication, May 19, 1999.

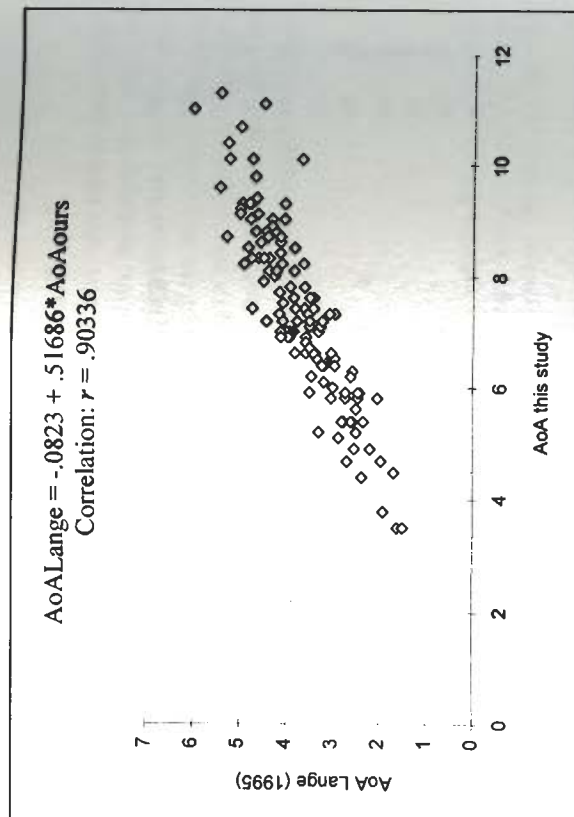


Figure 1. Correlation between AoAs collected by Lange (1995) and AoAs of this study ($N = 121$).

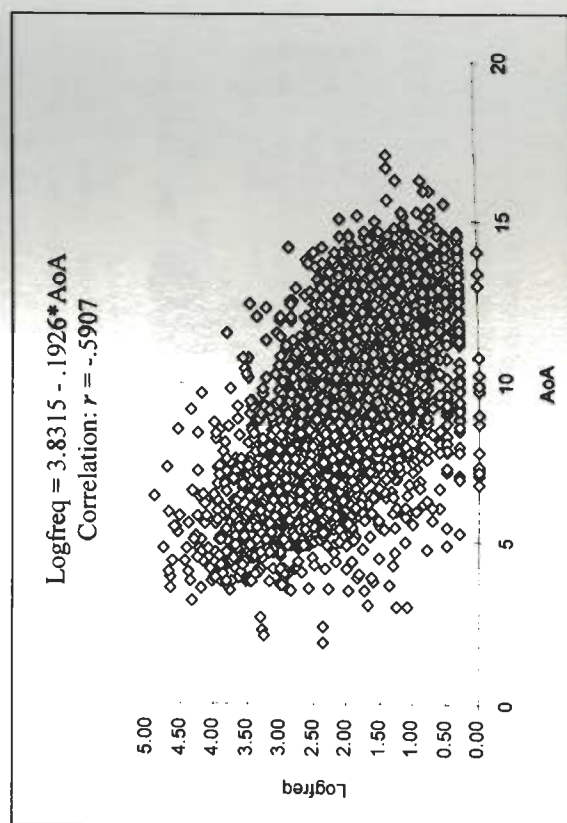


Figure 2. Correlation between Logfrequency and AoAs ratings in this study ($N = 2816$).

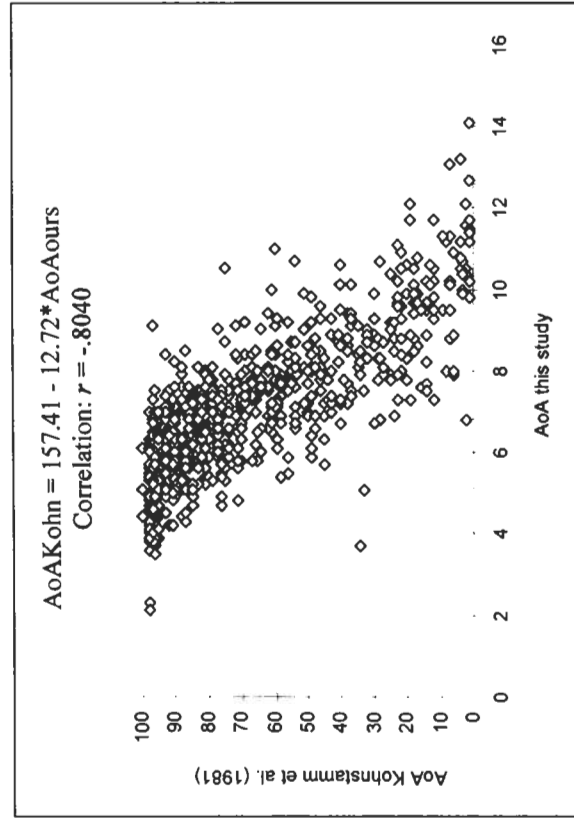


Figure 3. Correlation between the Kohnstamm et al. (1981) and AoAs ratings in this study ($N = 756$).

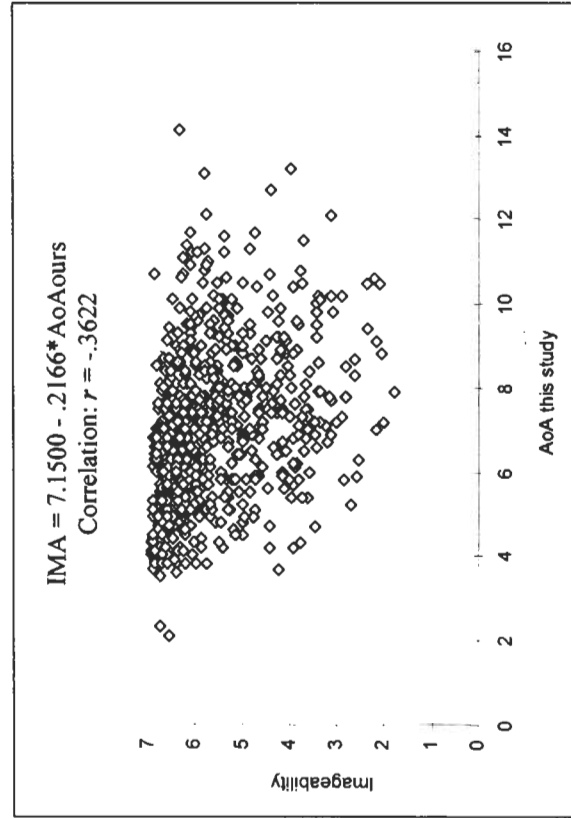


Figure 4. Correlation between imageability ratings and AoAs ratings in this study ($N = 756$).

Conclusion

It is becoming increasingly clear that researchers who want to investigate frequency effects have to control their stimulus materials for AoA. In addition, the AoA effect is becoming an interesting issue on its own. Although individual researchers will always be frustrated by the shortage of measures and by the fact that we could not obtain more objective measures (but see De Moor et al., 2000), we hope that the present data will provide Dutch speaking researchers with the same means as their English colleagues to start to clarify these issues.

References

- Alario, F. X., & Ferrand, L. (1999). A set of pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments and Computers*, 31, 531-552.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303-358). San Diego, CA: Academic Press.
- 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: Human Experimental Psychology*, 50A, 560-585.
- Bell, B. D., Davies, K. G., Hermann, B. P., & Walters, G. (2000). Confrontation naming after anterior temporal lobectomy is related to age of acquisition of the object names. *Neuropsychologia*, 38, 83-92.
- Brysbaert, M. (1996). Word frequency affects naming latency in Dutch with age of acquisition controlled. *European Journal of Cognitive Psychology*, 8, 185-193.
- Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, 12, 65-85.
- Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (in press). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*.
- Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 25, 85-95.
- Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods Instruments and Computers*, 31, 650-658.

- Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, 79, 1-12.
- De Moor, W., Ghyselinck, M., & Brysbaert, M. (2000). A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor, & Brysbaert. *Psychologica Belgica*, 40, 99-114.
- Elbers, L., & van Loon-Vervorm, A. (1998). Acquiring the lexicon. In S. Gillis & A. De Houwer (Eds.), *The acquisition of Dutch* (pp. 301-377). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 515-523.
- Ellis, A. W., & Lambon-Ralph, M. (in press). 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*.
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, 113, 256-281.
- Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 267-283.
- Gerhand, S., & Barry, C. (1999a). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, 27, 592-602.
- Gerhand, S., & Barry, C. (1999b). Age-of-acquisition and frequency effects in speeded word naming. *Cognition*, 73, 27-36.
- Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, 71, 105-110.
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures of 1944 words. *Behavior Research Methods & Instrumentation*, 12, 395-427.
- Gilhooly, K. J., & Logie, R. H. (1982). Word age-of-acquisition and lexical decision making. *Acta Psychologica*, 50, 21-34.
- Ghyselinck, M., & Brysbaert, M. (submitted). *Towards a solution of the word-familiarity versus word-frequency debate: The neglected role of age of acquisition*.
- Howes, D. H., & Solomon, R. L. (1951). Visual duration thresholds as a function of word probability. *Journal of Experimental Psychology*, 41, 401-410.
- Kohnstamm, G. A., Schaerlaekens, A. M., de Vries, A. K., Akkerhuis, G. W., & Froominckx, M. (1981). *Nieuwe streeflijst woordenschat*. Lisse: Swets & Zeitlinger.
- Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lewis, M. B. (1999). Age of acquisition in face categorisation: Is there an instance-based account? *Cognition*, 71, 23-39.
- Lyons, A. W., Teer, P., & Rubenstein, H. (1978). Age-at-acquisition and word recognition. *Journal of Psycholinguistic Research*, 7, 179-187.

- Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.
- 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*, 21, 116-133.
- 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 variable. *The Quarterly Journal of Experimental Psychology*, 50A, 528-559.
- Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, 19, 736-755.
- Schaerlaekens, A., Kohnstamm, D., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen: Derde herziene versie gebaseerd op nieuw onderzoek in Nederland en België*. Lisse: Swets & Zeitlinger.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Snodgrass, J. G., & Vandervort, 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.
- Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory and Cognition*, 26, 1282-1291.
- Uit den Bogaart, P. C. (Ed.). (1975). *Woordfrequenties in geschreven en gesproken Nederlands*. Utrecht: Oosthoek, Scheltema & Holkema.
- van Loon-Vervorm, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse woorden*. Lisse: Swets & Zeitlinger.
- van Loon-Vervorm, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse: Swets & Zeitlinger.
- Yamazaki, M., Ellis, A. W., Morrison, C. M., & Lambon-Ralph, M. A. (1997). Two age of acquisition effects in reading of Japanese Kanji. *British Journal of Psychology*, 88, 407-421.