Aphasia and age-of-acquisition: Are early-learned words more resilient?

Marc Brysbaert [1] Andrew W. Ellis [2]

[1] Ghent University, Belgium

[2] University of York, UK

Keywords: aphasia, age of acquisition, word production, word retrieval, word frequency

To be published in Aphasiology

Address: Marc Brysbaert
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Gent
Belgium
Tel. +32 9 264 94 25
Fax. +32 9 264 64 96
E-mal: marc.brysbaert@ugent.be

Abstract

We first review the empirical evidence indicating that age of acquisition (AoA) predicts which words remain accessible in acquired aphasia: words learned early in life are more likely to be retained than words learned later in life. Then we review the origins of the AoA effect and indicate why it is likely to be accompanied by a frequency effect. We argue that the effects of AoA and frequency are a result of the mind and brain acquiring information over time in an incremental manner, improving with practice, and possibly showing a decline in plasticity as the brain ages and more information becomes stored. In addition, AoA has a unique effect on the ease with which the correct verbal response can be retrieved in object naming. This may be due to the organisation of the semantic system, such that early acquired meanings are richer, more accessible and more robust against brain damage, or it may be due to the way in which semantic information is translated in verbal output. We also pay attention to the practical and methodological considerations aphasia researchers should take into account if they want to advance the knowledge of AoA and aphasia, rather than obscure it by using suboptimal word measures and empirical designs.

Aphasia and age-of-acquisition: Are early-learned words more resilient?

Many aphasic patients can retrieve some words quite reliably when speaking but have difficulty retrieving other words. So what makes one word consistently easy to retrieve and another consistently difficult? Lately it has become clear that age of acquisition (AoA) is a powerful predictor of the retention or loss of individual words in acquired aphasia – arguably the best predictor – and that it may also affect relearning. In this paper we first review the evidence for AoA effects, both in aphasia and more widely. We then discuss the underlying mechanisms that have been proposed to explain why healthy adults find early-learned words and concepts easier to process than later-learned words and concepts, and why words learned early in life are more robust to the effects of brain damage than words learned later (not just in aphasia but in various forms of dementia too). After a brief discussion of the neural basis of AoA effects, we continue with an in-depth consideration of the factors researchers should take into account if they want to further advance our knowledge of AoA and neuropsychological disorders. Which measures of AoA and word frequency should be used, which designs are the most powerful, and which confounds should be avoided?

We begin with a review of the eleven studies we are aware of that have explored the role of AoA in the retention and loss of vocabulary in aphasia.

Evidence for an AoA effect in aphasia

To the best of our knowledge, Rochford and Williams (1962) were the first to link vocabulary loss following brain damage to AoA. The authors assessed 32 patients referred for "showing clinical evidence of speech disorder in association with organic pathology" (p. 224). The patients were asked to name drawings of seven objects. If a patient was unable to name an object, up to four cues were given. The percentage of patients able to name different objects, with cues if necessary, ranged from 19% (buckle) to 81% (comb). Rochford and Williams (1962) noticed that the percentage of patients who could name the pictures correlated almost perfectly with the age at which children were first able to name the same pictures (data collected by the authors around the same time). The authors concluded (p. 227) that "order of difficulty [in aphasia] shows a close relationship with the age of acquisition in children: the names first learned in childhood are those last lost".

Rochford and Williams's (1962) study did not have much impact, partly because other authors started to favour word frequency as the factor affecting word finding problems in aphasia, arguing that what makes words easy or difficult for aphasic patients to retrieve is whether they occur frequently or infrequently in spoken and written language (e.g., Howes, 1964; Newcombe, Oldfield, & Wingfield, 1965). AoA and word frequency are highly intercorrelated (early-learned words tend to be more common in adult language than later-learned words). It can be difficult in practice to disentangle the two variables. In addition, word frequency had a higher status because the variable could be measured in an objective way, values were available for all words, and influential theoretical models of word recognition and production proposed mechanisms for how frequency of encounter with different words might influence their processing (e.g., Morton, 1969). After the publication of the seminal works by Preston (1935) and Howes and Salomon (1951), word frequency rapidly became the prime variable in word recognition research more widely. Indeed, Rochford and Williams (1965) in a later publication endorsed word frequency, suggesting that (p. 412) "[the apparent effects of AoA are] an artefact in that earlier learnt responses tend also to be the most frequently used responses throughout life. Indeed, they are learnt early because of their greater communication value and high frequency of usage and this is the factor mainly responsible for their easy accessibility."

It took 20 years from Rochford and Williams (1965) for interest in the role of AoA in aphasia to re-surface. Feyereisen, Van der Borght, and Seron (1988) examined whether patients with aphasia would be better at naming pictures of objects they can interact with (such as apple and needle) than objects they cannot act upon (whale or cloud). Having found some evidence for an advantage of manipulable objects, the authors subsequently realised these words had been acquired earlier and hypothesized that AoA could be the true origin of the manipulability effect. They found evidence for an effect of AoA on object naming and when they controlled both AoA and the rated familiarity of different objects, the effect of manipulability disappeared.

Nickels and Howard (1995) introduced a new approach. They used more stimulus materials and performed regression analyses on the naming scores. In a first study, they presented a set of 104 pictures of objects five times to 12 aphasic patients (so that the scores for different items varied from 0 = not named at all, to 5 = named on every occasion). A multiple regression analysis found significant effects of AoA, word length, and manipulability on naming accuracy, but no effect of word frequency, the visual complexity of the pictures, word familiarity, imageability or concreteness. In a second study, 130 line drawings were shown (once each) to a new set of 15 patients. The significant predictors now were AoA, word length and imageability. Manipulability and word frequency were again not significant. The results of the two analyses produced somewhat inconsistent results, but AoA emerged as significant in both.

Ellis, Lum, and Lambon Ralph (1996) used a similar methodology. They presented 139 object pictures to six aphasic patients for naming on three separate occasions. Five predictor variables were used: object familiarity, AoA, word frequency, imageability, and word length. For the group of six patients as a whole, AoA correlated most strongly with naming accuracy and was the only variable to make a significant independent contribution to predicting naming performance. This finding was replicated in Spanish-speaking aphasic patients by Cuetos, Aguado, Izura, and Ellis (2002) who asked 16 patients to name 140 black-and-white line drawings of objects on three separate occasions. Two more control variables were added - the visual complexity of the drawings and animacy (living or nonliving). Significant effects (in decreasing order) were found for AoA, word frequency, visual complexity and object familiarity. There was no effect of animacy.

Kittredge, Dell, Verkuylen, and Schwartz (2008) tried to disentangle the effects of AoA and word frequency. They asked 50 patients to name the 175 pictures of the Philadelphia Naming Test and used hierarchical linear modelling to identify the variables affecting naming performance. The predictors were AoA, word frequency, word density (number of phonologically similar words), word length, word imageability, and name agreement. The authors found significant effects of both AoA and word frequency on percentage correct responses.

Rossiter and Best (2013) focused on the fact that some objects are more typical than others members of the category to which they belong. For example, robins are more typical for birds than swans or penguins. They were interested to know if typicality would predict word naming capabilities in aphasic adults. To address the issue, the authors asked 20 people with acquired aphasia to name 200 drawings of objects. The set of objects used included both living (animals, vegetables) and non-living items (furniture, tools). Data was also obtained on the frequency, AoA, imageability, manipulability and length of the objects and their names. In a regression analysis, typicality was found to exert a significant independent effect on naming accuracy ($p < .04$) but that variable was outperformed by (in decreasing order) AoA, frequency, manipulability and word length.

Bastiaanse, Wieling, and Wolthuis (this issue) analysed the performance of 54 aphasic individuals on the naming of 50 pictures of objects and 50 pictures of actions. For both types of stimuli, they observed significant effects of AoA and imageability. Word frequency only influenced object naming; no effect was observed in verb naming (this was also true in two other tests that asked the participants to fill in verbs in sentences).

Finally, a few studies have looked at the recovery of vocabulary in aphasia. Laganaro, Di Pietro, and Schnider (2006) found better recovery trajectories for early- than late-acquired words following therapy in a group of French-speaking aphasic patients. Zhou, Liang, Xu, and Luo (2009) reported a similar pattern in the spontaneous language recovery of a Chinese aphasic patient. Thus, not only are early-learned words less likely to be lost in acquired aphasia than later-learned words, but if an early-learned word *does* become inaccessible as a result of aphasia, it is more likely to be recovered than a late-acquire word is. Word frequency also appears to make a contribution to predicting retention and loss of individual words (see below).

Kay and Ellis (1987) were under the spell of word frequency when they analysed the naming performance of aphasic patient EST (a man who had a long-standing anomia caused by a brain tumour that had been successfully removed). Kay and Ellis identified frequency as a factor influencing EST's naming performance (better production of higher than lower frequency words) but did not consider AoA at all. Ellis (2006) attempted to rectify that anomaly by revisiting EST's object naming results using both AoA and frequency as predictor variables. Both factors were significant predictors. Like many aphasics, EST was most likely to be able to name objects whose names were both early acquired and used with high frequency in the language. He was least likely to be able to name objects whose names were both late acquired and used with low frequency.

Having established that AoA does indeed appear to affect word retrieval in aphasia, in the next section we turn to a consideration of what might cause the AoA effect.

Possible mechanisms underlying the AoA effect

The evidence for an influence of AoA on aphasic word retrieval is now quite compelling, but *why* might early-acquired words survive the effects of brain damage better than later-acquired words in aphasia (and in dementia, as reviewed by Ellis, 2012)? To answer that question we must turn to a wider literature including cognitive research, computational modelling and cognitive neuroscience. We will endeavour to focus on principles and avoid unnecessary details.

As well as affecting the accuracy of word retrieval in aphasic patients, AoA also affects the speed with which healthy adults can retrieve words from their lexicons. Many studies across several different languages have reported that early-acquired words can be accessed and produced faster than later-acquired words in picture naming and other tasks (for reviews, see Johnston & Barry, 2006; Juhasz, 2005). And although this has sometimes been the basis of controversy, we think it is fair to say that where there is an AoA effect (faster, more efficient processing of early- than late-acquired words) there is also a frequency effect (faster, more efficient processing of high- than low-frequency words).

After reviewing the literature as it then existed, Brysbaert and Ghyselinck (2006) argued that the AoA effect is likely to have two components, one linked to word frequency and one independent of frequency. Some tasks are influenced by AoA and frequency to a similar extent: both effects may be small, or both may be substantial, but they are always similar in magnitude. Those tasks include reading written words aloud as quickly as possible (word naming), deciding whether sequences of letters are real words or not (lexical decision) and making decisions based on the meanings of words (semantic categorisation). Brysbaert and Ghyselinck (2006) said that in these tasks the AoA effect is "frequency related".[1] In these tasks, frequency and AoA appear to be two sides of the same coin, working together to influence processing speed and accuracy.

There are other tasks in which both AoA and frequency effects are present, but the influence of AoA is substantially larger than the influence of word frequency. Object picture naming is the clearest example: picture naming speed is influenced by both AoA and word frequency, but across several studies in several languages, the size of the AoA effect is considerably larger than the size of the frequency effect. Brysbaert and Ghyselinck (2006) said that in these tasks the AoA effect is "frequency independent".

[1] Studies in which this was not the case often used suboptimal measures of frequency or AoA, or did not have enough variation in one of the variables (see below).

Frequency-related AoA effects

We will deal first with circumstances in which AoA and frequency effects are yoked together and of similar magnitudes. How might such related effects arise? Ellis and Lambon Ralph (2000) proposed a possible mechanism using a simple "connectionist" model. Such models are quite simple in principle. They have banks of "input units" across which the researcher can create patterns by turning some of the units on and leaving others off. The model must learn to associate different input patterns with patterns expressed across the output units of the model. The model does this by gradually adjusting the strengths of the connections ("weights") between input and output patterns until each input pattern activates the desired output pattern. There is often a third set of "hidden" units that sits between the input and the output units and facilitates learning. Models of this sort have been used to simulate various aspects of human behaviour and can be surprisingly insightful and thought-provoking.

Ellis and Lambon Ralph (2000) trained their model to associate pairs of input and output patterns that were presented to the network with varying frequencies. In addition, some "early" pairs were introduced to the model at the start of training while the introduction of other "late" pairs was delayed until the model had spent time learning the early pairs. Importantly, the early pairs continued to be trained after the late pairs had been introduced (just as early words continue to be experienced and used as you increase your vocabulary by learning later words). Ellis and Lambon Ralph found that by the end of extensive training, the accuracy with which the model could convert input patterns into correct output patterns was influenced by how often individual pairs had been trained (i.e., by frequency) *and* by the point at which each pair was introduced into training (the model's version of AoA). In one simulation, input-output pairs were introduced at four different points in training. The model's performance at the end of extensive additional training neatly reflected the order in which the different pairs had entered training.

Changes to connection strengths in models like these are greatest at the start of training and become progressively smaller as training proceeds. When a model learns its first set of "early" items, it does not know that other items are going to be introduced down the line. The early items colonise the whole of the network and take advantage of the opportunity to make big changes to connection strengths. Later items may want the connections be adjusted rather differently – they may want the network to have a slightly different overall shape – but their attempts to reconfigure the model are resisted by the early items which continue to be trained alongside them. High frequency of training helps both early and late items to stamp themselves effectively upon the network which is why the effects of age (or order) of acquisition and frequency are yoked in these networks.

Ellis and Lambon Ralph made two further observations that are relevant here. First, the cost of being entered late into training is relatively small if the relationships between input and output patterns are relatively systematic, consistent and predictable. An example might be naming written words with regular spelllings. If such words are learned late in childhood, or in adulthood, they can take advantage of the connections between letters and sounds that were established by regular words learned earlier and can, as a consequence, be assimilated into the vocabulary without much difficulty. In contrast, if you encounter a written word like TOMB relatively late in the course of learning to read, your system will struggle to learn to pronounce it correctly against the competition from rivals such as BOMB, COMB, WOMB, etc., particularly if, having encountered it, you do not read TOMB very often subsequently. This intuition was supported when P. Monaghan and Ellis (2010) trained a connectionist model of reading and found that it duly struggled to learn late-acquired, irregular words. The prediction that AoA effects in human reading (word naming) should mostly be carried by late acquired words and low frequency words with irregular spelling-sound correspondences was upheld by J. Monaghan and Ellis (2002; see also Ellis & J. Monaghan, 2002).

Even for irregular words, the relationship between their spelling and their pronunciation is not completely arbitrary. At least the pronunciation of YACHT starts with "y" and ends with "t". The relationship between the object yacht and its name is, in contrast, completely arbitrary. There is no logical reason why one thing is a "yacht", another a "boat" and a third thing a "ship". Lambon Ralph and Ehsan (2006) showed that when a connectionist network is required to learn arbitrary mappings between inputs and outputs of the sort seen between objects and their names, there is a great benefit to being one of the items entered into training at the outset and a huge cost to being an item that only enters late into training. A prediction arising from this is that the task of object naming should show larger effects of AoA than the task of reading aloud, which appears to be the case (Brysbaert & Ghyselinck, 2006; Lambon Ralph & Ehsan, 2006).

Another observation made by Ellis and Lambon Ralph (2000) – one that relates particularly closely to the concerns of this paper – was that if they damaged their trained network by re-setting a proportion of the connections to zero, the consequences were worse for late acquired than for early acquired items. The order of entry of items into training affected their resilience as reflected in their ability to survive network damage. We believe that there is quite a close analogy to be drawn between the resistance to damage of items entered early into training in a connectionist network and the resistance to damage of early-learned words in aphasia and dementia.

Frequency-independent AoA effects

Returning to Brysbaert and Ghyselinck (2006) and the relationship between AoA and frequency effects, those authors noticed that while the magnitude of AoA and frequency effects are yoked in many tasks, there are other tasks were the impact of AoA is much greater than the impact of word frequency. The clearest example is picture naming in healthy participants where the AoA effect is more than 100 ms larger than expected on the basis of the size of the frequency effect. Other tasks in which the AoA effect is substantially larger than the frequency effect, are word associate generation (Brysbaert, Van Wijnendaele, & De Deyne, 2000), naming instances from a semantic category (Catling & Johnston, 2005), and retrieving a word as a response to a definition (Navarrete, Pastore, Valentini, & Peressotti, 2015). This suggests that there is a second source of AoA effects that plays more of a part when verbal responses must be given on the basis of semantic information than in tasks like reading words aloud or lexical decision.

Brysbaert and Ghyselinck (2006) explored the possibility, first proposed by van Loon-Vervoorn (1989), that AoA could affect the way in which meanings are represented and the way in which the semantic system itself is organised (see also Brysbaert et al., 2000). According to this view, the first learned meaning representations (concepts) are more easily accessible in semantic memory than later acquired representations. Evidence for this idea was reported by Dent, Catling, and Johnston (2007), who observed that pictures with late acquired names must be presented for a longer time before they can be identified, and by Catling, Dent, and Williamson (2008), who noticed that pictures with late acquired names take exceptionally long to name when they are visually degraded. As a corollary of the semantic hypothesis, the first acquired representations are also more impervious to damage than the later acquired representations. Whilst not ruling out the possibility that the frequency with which concepts are experienced and activated might also influence their semantic representations, this account suggests that the impact of order of acquisition within the semantic system is greater than the impact of frequency of use.

One account of how AoA could influence the organisation of the semantic system was presented by Steyvers and Tenenbaum (2005). They started from the observation that the semantic system is most likely organised in a network structure, consisting of nodes (concepts) that are connected to other nodes. Such a network is unlikely to have a random organisation or an organisation in which all nodes are connected to each and every other node. Rather, the most common structure observed in real-world networks that have grown up spontaneously is a so-called ‘small world’ organisation in which a small number of central nodes (hubs) are connected to a cluster of dedicated nodes and, in addition, to the hubs of other clusters. This is seen, for example, in the organisation of airports: many local airports connect to a nearby larger airport – the hub – from which long-distance flights are offered. Steyvers and Tenenbaum (2005) argued that in any system organised along such lines, the nodes established first are more likely to become hubs than are nodes established later. To make this more concrete, the mammals you learn early in life (e.g., cat, dog, sheep, lion) should form the core of your understanding of what mammals are, with animals learned later being understood in relation to those early concepts. In relation to this possibility, Storms, De Boeck, and Ruts (2000) provided evidence that stimulus categorisation (e.g., deciding whether a newly encountered animal is a mammal or a bird) is partly based on the comparison of the new stimulus to a few prominent exemplars of the existing categories. Also, Holmes and Ellis (2006) found a significant correlation ($r = -.434$) between the age of acquisition of familiar objects and the degree to which they were rated as typical of the category to which they belonged. Early-acquired objects were considered more typical exemplars of their categories than later-acquired object. It is not unreasonable to assume that the first learned instances are the most prominent exemplars of a semantic category. So, deciding whether a newly encountered moose is a mammal or a bird would be done by comparing the moose to a cat, dog, sheep, lion on the one hand, and to a robin, sparrow, tit on the other hand.

If early-acquired concepts are more accessible and form more central hubs in the semantic system, in all likelihood they will be more resistant to network damage than later acquired concepts. This prediction is supported by the finding that early words are more preserved not only in acquired aphasia (see above) but also in Alzheimer's disease (Cuetos, Arce, Martínez, & Ellis, in press; Cuetos, Rodriguez-Ferreiro, Sage, & Ellis, 2012; Ellis, 2012; Forbes-McKay, Ellis, Shanks, & Venneri, 2005; Sailor, Zimmerman, & Sanders, 2011) and semantic dementia (Lambon Ralph, Graham, Ellis, & Hodges, 1998; Woollams, 2012; Woollams, Cooper-Pye, Hodges, & Patterson, 2008). Here too it is observed that early-acquired words are most resistant to semantic memory loss.

Neuroimaging studies have also reported effects of AoA on the degree to which various brain areas become engaged in different tasks (Hernandez & Li, 2007). Brain areas associated with semantic processing seem to be particularly prone to such effects. An fMRI study by Fiebach, Friederici, Müller, von Cramon, and Hernandez (2003) reported that early acquired words activate more semantically related brain areas than late acquired words. One of the brain areas most strongly associated with semantic processing is the left anterior temporal lobe which has been linked to the establishment of conceptual or semantic representations and is the area crucially implicated in semantic dementia, a condition involving a relatively pure and selective loss of semantic knowledge (Patterson, Nestor, & Rogers, 2007; Visser, Embleton, Jefferies, Parker, & Lambon Ralph, 2010). Ellis, Burani, Izura, Bromiley, and Venneri (2006) observed a stronger response to early- than late-acquired objects in the left anterior temporal lobe in a picture naming task.

Ellis et al. (2006) also observed a stronger response to early than late acquired objects in visual processing areas of the brain that they struggled to explain. Urooj, Cornelissen, Simpson, Wheat, Woods, Barca, & Ellis (2014) revisited the issue in a magnetoencephalography (MEG) study of object naming. The advantage of MEG is that it gives a much more detailed picture of brain activity in time, so that there is information not only about which brain areas are activated during the performance of a task but also about when the areas become active and when their activation is affected by factors like AoA. The results of the study by Urooj et al. (2014) suggested that the initial analysis of object forms in visual cortex is not influenced by AoA, but that a fastforward sweep of activation from the occipital to the left anterior temporal cortex results in stronger activation of semantic representations for early- than late-acquired objects. Activation then feeds back to visual areas which show a second burst of activation that is greater for early than late acquired items. The modulation of visual activity by AoA, therefore, occurs later in time and can be interpreted as a top-down by-product of the stronger semantic response to early acquired objects.

The importance of the anterior temporal cortex for picture naming has also been observed in lesion studies. Bell, Davies, Hermann, and Walters (2000) investigated the factors influencing naming decline in patients who underwent left anterior temporal lobectomy for the treatment of epilepsy. They used the Boston Naming Test (consisting of 60 line drawings) to examine declines in naming performance after the surgery and examined the stimulus characteristics predictive for the decline. Of the variables word length, word frequency, AoA, and semantic category (living or nonliving), only AoA predicted the percentage of patients who could no longer name the object. A follow-up study by Ruff, Swanson, Hammeke, Sabsevitz, Mueller, and Morris (2007) used the same test and examined patients with left and right anterior temporal lobotomy before and after the surgery. The authors found the same results as Bell et al. (2000): the words most likely to be lost after left lobotomy were late acquired words. Unfortunately, Ruff et al. (2007) did not include word frequency as a variable in their analyses.

Although a semantic contribution to the exceptionally large AoA effect in picture naming has a firm grounding, it should be noted that not all evidence supports the idea. For a start, AoA does not seem to have a larger effect than word frequency in semantic judgment tasks, such as deciding whether a stimulus refers to a naturally occurring or a man-made object (Catling & Johnston, 2006, 2009; Holmes & Ellis, 2006; Menenti & Burani, 2007). A second problem is that the semantic hypothesis seems to make the wrong prediction about the AoA effect in second language word processing. According to the semantic hypothesis, the AoA effect in a newly acquired language should be the same as the AoA effect in the initially acquired language (as the same meanings are involved). This is not what has been found. In general, the AoA effect in second language processing is related to the order in which the words of the second language were acquired rather than the order in which the meanings of the words were acquired (Hirsh, Morrison, Gaset, & Carnicer, 2003; Izura & Ellis, 2004). A third problem is that the effect of AoA on EEG data seems to be rather late (Laganaro, Valente, & Perret, 2012; Perret, Bonin, & Laganaro, 2014), and different from the category typicality effect found in semantic decision (Räling, Holzgrefe-Lang, Schröder, & Wartenburger, 2015). Finally, in Steyvers and Tenenbaum's (2005) network simulations, frequency of occurrence also affected which nodes were likely to become hubs, and the impact of frequency versus order of acquisition could be modulated by changing a parameter of the model. So, if AoA affects the organisation of the semantic system, frequency of occurrences is likely to do so too (possibly to a different degree). Research by Hoffman, Rogers, and Lambon Ralph (2011) suggests that the effect of frequency on semantic organisation could be particularly strong when a high frequency of occurrence entails that the word or concept is encountered in a large diversity of contexts. Frequency of occurrence would have less impact if the entity is always encountered in the same context.

Because of the problems with the semantic organisation hypothesis, Belke, Brysbaert, Meyer, and Ghyselinck (2005) ventured that the extra effect of AoA in picture naming might (additionally) be due to the selection of output words (lemmas) rather than to the organisation of the semantic system. One of the characteristics of word production in picture naming is that several word candidates become activated by the semantic system and compete with each other for production. So, a picture of a lion not only activates the word candidate 'lion', but also 'tiger', 'cat', 'badger', 'bear', and so on. According to Belke et al. (2005), the unexpectedly large AoA effect in picture naming could be due to the fact that early acquired lemmas are stronger competitors in word production than later acquired lemmas. As a result, late acquired words would have to compete extra hard against earlier acquired competitors. Other authors have found evidence too suggesting that AoA has a particularly strong effect on the selection of the correct word in a situation when several word candidates are activated and compete with each other (Dent, Johnston, & Humphreys, 2008; de Zubicaray, Miozzo, Johnson, Schiller, & McMahon, 2012; Navarrete, Scaltritti, Mulatti, & Peressotti, 2013; Woollams, 2012).

Independent of whether the frequency-unrelated AoA effect in picture naming comes from the organisation of the semantic system or from the translation of the activated semantic concepts into the right word, it is clear that AoA has a particularly strong effect when it comes to produce words on the basis of semantic information. Both components make that late acquired words are particularly vulnerable to brain damage resulting in acquired aphasia. Research with aphasia patients could even be critical to decide between both views: Do patients fail to name pictures because semantics are no longer able to activate the right word or because there is interference from other, easier response candidates?

On the importance of good measures and good research designs

Although the literature in hindsight is quite convergent about the contributions of AoA and word frequency to performance in a variety of tasks, research has been characterised by several false starts and needless discussions, which could have been avoided if all researchers had used good research designs. This is another message we want to convey. Research on factors like AoA and frequency is inevitably limited by the quality of the research design, including the quality of the measures used. Our measures of when people of different ages learned different words and our measures of how often they encounter and use those words in everyday life are both inherently imperfect. In some ways it is surprising that the measures we have predict human performance across a range of tasks as well as they do. Still, it requires us to use the best available measures and research designs if we want to move forward.

Measures of word frequency

It cannot be denied that some of the inconsistencies in the studies reported above (and in other studies not included in our review) are due to the use of suboptimal measures of AoA and/or word frequency. Indeed, the growth of AoA research has triggered a critical evaluation of which word frequency measure to use in psychological research (Brysbaert & New, 2009; Zevin & Seidenberg, 2002). It rapidly became clear that not all frequency measures are equally good. In particular, the much-used word frequencies of Kucera and Francis (1967) turned out to be limited in value, because of the small size of the corpus (only 1 million words), because of the dated sources, and because the measure was based entirely on written texts in which, for example, the names of domestic objects such as food and clothing may be under-represented. If AoA and word frequency both affect performance, then using a poor measure of one will cause the influence of the other to be over-estimated (Brysbaert & Cortese, 2011).

At the moment, the best frequency measure for use with student populations is based on television and film subtitles, or a combination of subtitles and written sources such as fiction, newspapers and magazines (Brysbaert & New, 2009; Brysbaert, New, & Keuleers, 2012; Dimitropoulou, Duñabeitia, Avilés, Corral, & Carreiras, 2010; Mandera, Keuleers, Wodniecka, & Brysbaert, 2015; Soares, Machado, Costa, Iriarte, Simões, de Almeida, Comesaña, & Perea, 2015).[2] In addition, we know that country-specific frequency measures are better than word frequencies obtained in another country. So, British researchers are advised to use word frequencies based on British materials, whereas American researchers should use American measures (van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

An extra complicating factor in relation to aphasia research is that many studies are done with patients who are much older than typical students. Cuetos et al. (2012) found that for elderly Alzheimer patients, word frequencies based on slightly dated written language samples predicted object naming accuracy better than word frequencies based on (newer) film subtitles (see Adelman, Marquis, Sabatos-DeVito, & Estes, 2013, for a similar observation). So, it is a good idea to work with several word frequency measures (e.g., in the regression analyses discussed below) to see which one works best for the participant sample tested.

Measures of AoA

Measures of the age of acquisition of different words are also non-optimal. Much of the AoA research in the literature has been done using adult ratings of when participants (often students) believe that they and others learned different words. Such a process seems questionable: How much insight are adults likely to have into the age at which different words are learned? To tackle the problem, AoA researchers have invested energy in validation studies, to see how well adult ratings correlate with data obtained from children. The largest study was reported by Morrison, Chappell, and Ellis (1997), who obtained AoA ratings for a large set of object names from young adults (undergraduate students) and compared them against data showing how well children ranging in age from 2 years 6 months to 10 years 11 months could name those objects. They found a correlation of .75 between the student AoA ratings and a measure of objective AoA derived from the real-life naming results. Similar correlations ranging from .55 to .75 have been reported between ratings and objective AoA measures in Spanish, Italian, French and Icelandic (Álvarez & Cuetos, 2007; Chalard, Bonin, Méot, Boyer, & Fayol, 2003; Lotto, Suiran, & Job, 2010; Pérez & Navalón, 2005; Pind, Jónsdóttir, Gissurardóttir, & Jónsson, 2000). In addition, objective AoA tends to predict naming speed in young adults better than rated AoA (Chalard et al., 2003; Ellis & Morrison, 1998; Lotto et al., 2010; Pind & Trygvadottir, 2002), indicating that AoA research would produce even stronger effects if objective measures could be used for all stimuli. In the absence of such objective norms, however, subjective AoA ratings are valid.

Because the Morrison et al. (1997) study is almost 20 years old and apparently not known or trusted by all researchers[3], we sought two new validation analyses in English. A first one is provided by a website for American teachers (https://www.flocabulary.com/wordlists/; retrieved on June 6, 2015), which provides lists of words to be taught in various classes (going from Kindergarten to Grade 8). According to the website, these lists have been created by first compiling words from grade-appropriate novels and basal readers. The researchers

[2] In all the analyses we did, subtitle-based frequencies also outperform frequencies based on spoken corpora (Brysbaert & New, 2009; New, Brysbaert, Veronis, & Pallier, 1997; see also Ernestus & Cutler, 2015, who compared subtitle frequencies with spoken frequencies for auditory lexical decision). The reason for this is that spoken corpora tend to be rather small and restricted in the variety of contexts sampled and topics talked about.
[3] Among whom one of our reviewers.

then analysed how often these vocabulary words appeared on US state tests. For each reading level, they looked at state tests at that level and two grade levels above. So the words taught in Grade 5 were defined as those words that are both found in 5th grade reading material and are most likely to appear on state tests in 5th, 6th and 7th grades. For 1,441 of these words, AoA ratings could be found in the list made available by Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012). For these words we also had the word frequencies based on American subtitles (Brysbaert & New, 2009). The correlation between Grade advised and rated AoA was $r = .78$ ($N = 1{,}441$, $p < .0001$). The correlation between Grade advised and word frequency was $r = -.59$ ($N = 1{,}441$, $p < .0001$); AoA and word frequency correlated $r = -.67$ with each other. Regression analyses indicated that if AoA was entered first, it explained 60% of the variance in Grade advised, with 1% more explained by frequency. If frequency was entered first, it explained 35% of the variance, with 26% explained extra by AoA. So, certainly for words taught in primary school and middle school, adult raters seem to have a good sense of the order in which the words are taught.

The second analysis addressed the question whether adults can give valid AoA estimates for words learned early in life. Adults remember very little from their childhood, a phenomenon known as *infantile* or *childhood amnesia*, which has been investigated extensively by psychologists (Bauer & Larkina, 2013; Tustin & Hayne, 2010). Indeed, in AoA rating studies as well, very few words receive AoA estimates of younger than three years (Kuperman et al., 2012). This raises the question whether adults can distinguish the order in which words are acquired by toddlers. Arguably the most extensive preschool database is provided by the MarArthur-Bates Communicative Development Inventories. In this widely used test to detect delays in language development, parents of children from 8 months to 30 months are asked to indicate which words their child speaks. On the basis of the norming data, Goodman, Dale, and Li (2008) calculated the age of production for 562 words as the age (in months) at which 50% of the children were able to say the word. [4] For 531 of these words Kuperman et al. (2012) gave AoA ratings. A further advantage of the dataset is that it includes all types of words (nouns, adjectives, verbs, function words), which Goodman et al. (2008) showed are not all acquired at the same time. On average, nouns are acquired first, followed by verbs, adjectives, and function words. In a regression analysis with Type of word (6 different types, as differentiated by Goodman et al., 2008), rated AoA, and Childes word Frequency[5] as predictors and age of production as the dependent variable, Type of word had a significant effect ($F(5,523) = 32.11$, $MSe=259.7$, $p < .0001$), together with word frequency ($F(1,523) = 20.0$; $MSe = 161.6$, $p < .0001$), as reported by Goodman et al. (2008). In addition, rated AoA had a significant impact as well ($F(1, 523) = 143.6$, $MSe = 1161.0$, $p < .0001$). The significance of AoA remained when the interaction between word type and word frequency was included in the model ($F(1,518) = 125.9$, $MSe = 994.3$, $p < .0001$; see Goodman et al., 2008, for evidence that the interaction of word type and word frequency may matter). Further analyses indicated that nearly all of the systematic variance was accounted for by word type ($R^2 = .12$) and AoA (increase in variance explained equal to $R^2 = .26$). Childes word

[4] The authors thank Philip S. Dale for kindly sharing the data with them.
[5] This word frequency measure based on child directed speech was used by Goodale et al. (2008) and is more representative for the world of young children than frequencies based on adult directed materials.

frequency and its interaction with word type accounted only for 4% extra variance (total explained variance: $R^2 = .42$). This shows that even for the earliest words, AoA ratings from adults provide a valid estimate of the order in which the words were acquired. At the same time, the findings show that the absolute AoA values given by adult raters tend to overestimate the age at which the first words are learned (ratings varied between 3 and 6 years). It is fair to say that AoA ratings provide a good estimate of the order in which the words are acquired but not of the exact age at which they are learned (see also Kuperman et al., 2012).

It might seem surprising that adults are able to give pretty good estimates of when early childhood words have been acquired, given that they remember little or nothing of those days. The task of generating AoA ratings and the task of recalling events from early childhood are, however, very different. Childhood amnesia is an inability to remember specific events from early childhood. It relies on what psychologists call *episodic memory* – memory for individual episodes in life – and is certainly weak for early childhood. But when adults are asked to provide AoA ratings they are not asked to recall the particular occasion when they first heard or used the word FAIRY or the word FUNDAMENTALIST. They are simply asked to indicate their belief that they and other people probably learned the word FAIRY when they were (say) two years old and the word FUNDAMENTALIST when they were (say) fourteen. The knowledge of what words mean forms part of what psychologists call *semantic memory*. People who become amnesic as a result of brain damage can no longer remember events from their lives (episodic memory) but often are still able to understand and use a wide range of words in conversation because their semantic memories remain intact. In addition, it has to be taken into account that when adults rate the AoA of words, they have other information to rely upon than their own memories; they are likely to make use of knowledge from interactions with young relatives or from children they encounter through babysitting or as part of their work in youth organisations.

Related to the distinction between episodic and semantic memory, it may be interesting to know that word frequency has opposite effects in semantic and episodic memory tasks. In semantic memory tasks like object naming or lexical decision, participants are faster to recognise and produce high than low frequency words. But if they are shown a list of words and at some later time asked to recall the words from the original list or pick out the words from a longer list with decoys (classic episodic memory tasks), they usually remember the more distinctive low frequency words from the list better than the high frequency words. Similar things may happen with AoA: you respond more quickly to early than late acquired words in semantic memory tasks but in episodic memory tasks late acquired words may be easier to remember than early acquired words. The dissociations between the effects of frequency and AoA in episodic and semantic memory have been investigated by Dewhurst, Hitch, and Barry (1998), Cortese, Khanna, and Hacker (2010), Cortese, McCarty, and Schock (2015).

AoA ratings provided by students are fine for young participants, but may not be the best estimates for older participants. Two studies have reported that AoA measures based on ratings obtained from elderly participants are better than AoA ratings from students at predicting performance of healthy older adults and of patients with Alzheimer's disease (Cuetos et al., 2012; De Deyne & Storms, 2007). This is understandable given that the learning histories of older and younger participants are likely to differ because of a cohort effect. For example, words relating to modern technology and food introduced in recent decades may be rated as early acquired by young adults but as late acquired by older people. For the same reason, objective AoA estimates collected in recent years are likely to be more applicable to young participants than to older ones.[6] AoA ratings from older participants are currently available for relatively few words, but they can easily be collected for the stimuli one wants to present.

To collect AoA ratings, we recommend that you ask participants to indicate the age when they think they learned each word in years. Do not use Likert scales (e.g., 1 = younger than three years, 2 = three to four years, …, 7 = older than twelve years). People find it easier to write that they acquired a word at the age of 4 years than to find the corresponding age on a scale.

Research designs

Another factor that has made research on AoA messier than it might have been is the tendency of authors to make use of small-scale factorial designs. In such designs, small samples of 'carefully selected' stimuli are compared with each other. For instance, in a study comparing the effects of AoA and frequency there are often four experimental conditions that orthogonally vary frequency (low vs. high) and AoA (early vs. late) in a 2x2 design. In each condition, stimuli are chosen (e.g., words or pictures) that differ as much as possible on the variables of interest while being matched as closely as possible on other, control variables (imageability, typicality, word length, manipulability, …).

There are several problems with such designs. The first is that stimuli taken from the extremes of a scale may be unrepresentative of words as a whole. For instance, the first acquired words are all related to the world of a toddler. This is particularly the case for early acquired words with low adult frequencies which can, if you are not careful, be dominated by words like fairy, giant and dragon. Similarly, high frequency, late acquired words are often related to politics, finance, university studies and science. A second issue is that because of the controls, it is often difficult to have a big difference between the levels of a variable, so that for instance the early acquired words may have been learned on average around the age of 7 and the late acquired words around the age of 9. Similarly, the difference between the low-frequency and the high-frequency words may be limited (e.g., between 10 per million words and 50 per million words). Furthermore, in such designs it is impossible to compare the ranges of the two variables, so that in one study frequency seems to be the more important variable and in the other AoA when the strength of the manipulation of the two factors may in fact be the underlying cause. Finally, the many constraints imposed by the control variables make it impossible to find a reasonably large number of stimuli, so that many experiments end up with 20 or less stimuli per condition (from which far-reaching theoretical conclusions are drawn).

[6] So, the Morrison et al. (1997) objective AoA estimates are becoming increasingly interesting for studies with older participants and will remain so for at least another 50 years.

In many respects, a better approach is to present a large set of items and to analyse the data using multiple regression which estimates how well performance can be accounted for by a series of predictor variables. In this approach, stimuli from the entire range of a variable (frequency, AoA, or any other variable) can be used, as long as researchers make sure that the correlation between the variables is not too high. Correlations below .60 are generally safe (you can find more information by looking under the term 'multicollinearity'). It is advisable to include as many items as practically possible (we recommend a minimum of 200).

In addition, researchers should not feel limited to traditional 'linear' regression. Other, more advanced and more powerful techniques have been developed recently such as generalized additive mixed models with non-linear relationships (see Baayen, 2013, for a general introduction; also see Cuetos et al., in press, and Kuperman, Estes, Brysbaert, & Warriner, 2014, for examples of the application of such analyses). Another advantage of regressions is that, having run one analysis, researchers can easily replace a suboptimal measure (e.g., of frequency or AoA) by a better one, without violating the factorial design.

Researchers have been fond of small-scale factorial designs, (a) because they assume that such designs provide a better view of the size of the main effects and the interactions between variables, and (b) because they think that an experimental design allows them to draw causal conclusions. As for the former assumption, sizes of main effects and interactions (in ms) can easily be obtained from the graphs current statistical regression packages provide (see Figure 3 in Kuperman et al., 2014, for an example). As for the latter conviction, researchers should be aware of the fact that selecting words for a factorial design does not give the study the power of an experiment, even though the results are analysed with an analysis of variance (ANOVA). The reason for this is that word characteristics are stimulus-specific. In a true experiment one can assign stimuli to one or the other condition at random or in a counterbalanced design. Under those conditions, one can be sure of the true origin when a significant effect is found.[7] Such control, however, is not possible in studies investigating the properties of words themselves. Researchers cannot assign the words at random to the AoA and frequency conditions; all they can do is select the words in the various conditions and match them on a number of confounds. This in effect turns the study into a correlational study, no matter how the data are analysed (Lewis & Vladeanu, 2006).

For the above reasons, we strongly recommend authors to use regression designs with the largest possible and diverse set of stimulus materials and the best possible measures of word frequency and AoA.

Practical considerations about research on the AoA effect in aphasia

In the previous sections we have provided arguments for the use of good measures and powerful designs in research on AoA. Given that some of these arguments at times became

[7] An example of a true experiment with words is semantic priming. In such studies the same target word (e.g., ‘doctor’) can be assigned at liberty to the semantic priming condition (preceded by the prime ‘nurse’) or to the control condition (preceded by the prime ‘purse’). Finding a difference between both conditions in a properly designed study then allows the researchers to accept the effect, independent of the target words used.

rather detailed, it may be good to repeat them but now applied to the specific case of research on aphasia (and dementia).

First, we mentioned that thought needs to be given to the appropriateness of the norms. This will partly depend on the characteristics of the participants in the study (their age, educational histories, etc.). When the participants are older (as is often the case in research on aphasia and dementia), it is good to search for frequency measures and AoA estimates aimed at this age group. For frequency, this means that the counts are based on materials the participants are likely to have come across (and, for instance, not based on a recently collected corpus of internet websites). Another source to avoid is frequency measures based on academic texts, if the participants do not have a university degree (Brysbaert et al., 2012). If possible, try to find several frequency measures and see how they perform relative to each other (Adelman et al., 2013; Cuetos et al., 2012). As for AoA measures, it is worthwhile to use AoA ratings provided by healthy participants of the same age as the patients. Such ratings are easy to collect and, as we have seen, are valid.

Second, there is no point in doing underpowered research. Thus usually leads to null results and, occasionally, to significant chance findings that cannot be replicated. Doing a study of adequate power involves a substantial number of items (we recommend at least 200). If you present these items to older participants or patients, you will need to take a view on whether they can tolerate a naming test consisting of substantial numbers of items in a single session, or whether it is better to present the items over more than one testing session (close enough to avoid any concerns that the participants may be changing over the testing period). It is also advised to randomise the order of items across patients, to ensure, for example, that different items are affected by end-of-session fatigue in different patients.

Third, it seems to be generally true that properties of objects and words that affect the speed with which healthy adults can respond to them in experimental tasks also affect the accuracy with which people with aphasia or dementia respond to the same items. Stimuli that healthy adults respond to correctly but slowly become stimuli they do not recognise or make errors to following brain damage. This means that it is worthwhile to take into account the factors that influence picture naming in healthy participants (such as visual complexity, typicality, manipulability, name agreement, word length, and other factors mentioned above[8]). It also means that it is good to search for validated stimulus materials, for which norms are available. These can be commercially available tests (such as the Peabody picture vocabulary test or the Boston naming test) or – even better – picture sets collected for psycholinguistic research (such as the International Picture Naming Project [http://crl.ucsd.edu/experiments/ipnp/index.html] or the coloured pictures developed by Rossion & Pourtois, 2004 [e.g., http://wiki.cnbc.cmu.edu/Objects]).

[8] Another variable that may matter and that has not yet been investigated is the number of morphologically derived words and compound words the name of an object has, a variable known as the morphologically family size (Baayen, Feldman, & Schreuder, 2006). Words with a large family size are easier to process than words with a small size.

Fourth, when doing AoA research it is good practice always to include information about word frequency as well. We have argued above that for many tasks AoA effects are likely to be accompanied by frequency effects of the same size (and vice versa), because part of the AoA effect is due to learning in systems that accumulate new information while at the same time refreshing the existing knowledge. In a few other tasks the AoA effect is likely to be considerably larger than the frequency effect. These tasks involve access to semantics and the translation of semantics to verbal output (such as picture naming and naming to definition). Indeed, an interesting question for each neuropsychological study involving the semantic system is to find out whether the AoA effect is larger than the frequency effect. For example, Sage and Ellis (2004) investigated the effects of AoA, word frequency, imageability, word length and number of orthographic neighbours on spelling to dictation in a dysgraphic patient whose spelling impairment would typically be ascribed to a 'graphemic buffer disorder'. The patient's object naming was at or above average for her age and her performance on a range of semantic tasks was normal. However, after her stroke she made a large number of spelling mistakes. Sage and Ellis proposed that the patient's deficit affected her ability to access orthographic representations from semantics rather than affecting the semantic or orthographic representations themselves. Although the authors found significant effects of all the variables investigated, they did not examine specifically whether the AoA effect was larger than predicted on the basis of the frequency effect. There are some data pointing in this direction (Table 3 or Sage & Ellis, 2004), but the information collected is not detailed enough to draw firm conclusions.

Modern regression techniques allow participant and item properties to be combined in a single analysis. This allows the researcher to conduct group analyses which retain a degree of sensitivity to individual differences. That should make it possible, for example, to discover whether the effects of AoA, frequency and other variables of interest relate to patient characteristics (e.g., performance on semantic tests and performance on phonological tests). There is also scope for combining these analyses with data derived from neuroimaging to discover, for example, whether damage or atrophy in different brain areas relates to the influence of different variables on performance (e.g., Harvey & Schnur, 2015). It is furthermore possible to combine 'categorical' variable like whether an object is living or nonliving, or a noun or a verb, with 'continuous' variables like word frequency and AoA. For these analyses, it is not necessary to attempt to create sets of living and nonliving items, or sets of nouns and verbs, that are matched on the many other variables that might affect performance. Living vs. nonliving or noun vs. verb can be entered into an analysis along with the other variables to see if Category has an effect of performance over and above the effects of those other factors. Such methods will allow a fresh assault to be made on the question of whether some aphasic patients perform better on some types of words than other. [9]

If practical issues dictate that the research questions be addressed at the level of single cases rather than groups, thought must be given to power and reliability. Ellis et al. (1996) asked six aphasic patients to name 139 object pictures on three separate occasions over the space of

[9] In this context we note that Bastiaanse et al. (2015) reported an effect of word frequency on retrieval of nouns but not verbs with AoA and imageability controlled.

three or four weeks. They found that when the data were analysed at the level of the single patient, the effects of AoA, frequency and other factors appeared to vary somewhat from session to session. The authors argued that these variations were most likely due to fluctuations in overall performance levels. They recommended that the results should be aggregated across administrations of the same items to reduce this statistical 'noise' (see also Cuetos et al., 2002). With the benefit of hindsight, however, we would recommend now that this type of study is run with more and better measures, more powerful analysis techniques (that take into account the effect of session), and more items.

Conclusions

There can be little doubt that the ability of (most) aphasic patients to retrieve and produce words in situations like object naming is influenced by the age of acquisition of those words and by their frequency of adult use. Words learned early in life and used frequently thereafter are easiest to retrieve and use while those learned late in life and used infrequently thereafter are hardest. There are doubtless other factors that influence aphasic word retrieval, with typicality and manipulability being on the list of candidates, but the contributions of AoA and word frequency seem clear.

We have argued that in many tasks comparable effects of AoA and frequency will be found if researchers use good measures and examine the entire range of the variables. We have further provided evidence that in some tasks the AoA effects is likely to be substantially larger than the frequency effect. These task all involve the production of verbal output on the basis of semantic information, such picture naming, associate generation, naming exemplars from a category, or retrieving a word as a response to a definition. Certainly with respect to picture naming, there can be no doubt that AoA is a more important variable than word frequency, also in young, healthy participants (Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004; Brysbaert & Ghyselinck, 2006; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005; Szekely et al. 2005).

At the same time, our survey of the research so far suggests that there may been been too much of an emphasis on naming object pictures in the literature, and that our ability to understand aphasic performance will improve if we explore word recognition and production in a wider range of tasks. For example, Cuetos, Herrera, and Ellis (2010) found an effect of AoA in patients with Alzheimer's disease in a ‘lexical selection’ task where they were simply required to look at four strings of letters (a word and three pronounceable nonwords) and indicate which one was the real word. Holmes, Fitch and Ellis (2006) reported an effect of AoA in Alzheimer patients in an ‘object decision’ task where they had to indicate whether line drawings depicted real or imaginary objects. For these tasks too it would be interesting to find out whether the AoA effect is larger than the one predicted by word frequency, and how aphasic patients perform on them. If aphasic patients show extra strong AoA effects in lexical selection and object decision tasks, that could point to a core semantic impairment as the source of the problem. In contrast, if the AoA effect in these tasks is much smaller than in picture naming, this would be extra evidence that the big AoA effect in picture naming is due to the translation from semantics to verbal output. Also, a number of studies suggest that some patients show stronger AoA and frequency effects than others (Cuetos et al., 2002; Ellis et al., 1996; Kittredge et al., 2008; Nickels & Howard, 1995). These estimates can be clouded by 'floor' and 'ceiling' effects (patients whose overall performance is either very good or very poor will perform well or badly on all items in a test and not be in a position to show large effects), but there is still scope for studies which compare the magnitude of AoA and frequency effects in different types of aphasic patients or look at the extent to which the magnitude of those effects are predicted by, for example, performance on tests of semantic or phonological ability.

Finally, further progress on the topic depends on the willingness of researchers to step up their methodological standards. As we have argued, there is little point in mindlessly reaching for the nearest available set of norms. It is worthwhile to spend time optimising the stimulus set to be presented and refining the measures of the stimulus features that will be used. Fortunately, recently developed statistical techniques based on regression rather than factorial design allow researchers to include many more stimuli than traditionally experienced under the tyranny of trying to create sets of items that vary on one factor while being matched on dozens of others. This increases the power (and hence the replicability) of the findings.

References

Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. G., & Estes, Z. (2013). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(4)*, 1037-1053.

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36(1)*, 140-155.

Álvarez, B., & Cuetos, F. (2007). Objective age of acquisition norms for a set of 328 words in Spanish. *Behavior Research Methods, 39(3),* 377-383.

Baayen, R. H. (2013). Multivariate statistics. In R. Podesva and D. Sharma (Eds), *Research methods in linguistics* (pp. 337-372). Cambridge: Cambridge University Press.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55(2)*, 290-313.

Bastiaanse, R., Wieling, M., & Wolthuis, N. (2015). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology*.

Bauer, P. J., & Larkina, M. (2013). Childhood amnesia in the making: different distributions of autobiographical memories in children and adults. *Journal of Experimental Psychology: General, 143(2)*, 597-611.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition, 96*, B45-B54.

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(1)*, 83-92.

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

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

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977-990.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding Part-of-Speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*, 991-997.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*, 215-226.

Catling, J. C., Dent, K., & Williamson, S. (2008). Age of acquisition, not word frequency affects object recognition: Evidence from the effects of visual degradation. *Acta Psychologica, 129(1)*, 130-137.

Catling, J. C., & Johnston, R. A. (2005). Age of acquisition effects on word generation. *European Journal of Cognitive Psychology*, *17(2)*, 161-177.

Catling, J. C., & Johnston, R. A. (2006). Age of acquisition effects on an object□name verification task. *British Journal of Psychology*, *97(1)*, 1-18.

Catling, J. C., & Johnston, R. A. (2009). The varying effects of age of acquisition. *The Quarterly Journal of Experimental Psychology*, *62(1)*, 50-62.

Chalard, M., Bonin, P., Méot, A., Boyer, B., & Fayol, M. (2003). Objective age-of-acquisition (AoA) norms for a set of 230 object names in French: Relationships with psycholinguistic variables, the English data from Morrison et al.(1997), and naming latencies. *European Journal of Cognitive Psychology*, *15(2)*, 209-245.

Cortese, M. J., & Khanna, M. M., & Hacker, S. D. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18(6)*, 595-609.

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology 68(8)*, 1489-1501.

Cuetos, F., Aguado, G., Izura, C., & Ellis, A. W. (2002). Aphasic naming in Spanish: Predictors and errors. *Brain and Language, 82,* 344–365.

Cuetos, F., Arce, N., Martínez, C., & Ellis, A. W. (in press). Visual word recognition in Alzheimer's disease: Effects of semantic degeneration. *Journal of Neuropsychology*.

Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. *Neuropsychologia, 48(11),* 3329-3334.

Cuetos, F., Rodríguez-Ferreiro, J., Sage, K., & Ellis, A. W. (2012). A fresh look at the predictors of naming accuracy and errors in Alzheimer's disease. *Journal of Neuropsychology, 6(2),* 242-256.

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(3),* 274-295.

Dent, K., Catling, J. C., & Johnston, R. A. (2007). Age of acquisition affects object recognition: Evidence from visual duration thresholds. *Acta Psychologica, 125(3),* 301-318.

Dent, K., Johnston, R. A., & Humphreys, G. W. (2008). Age of acquisition and word frequency effects in picture naming: A dual-task investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(2),* 282-301.

de Zubicaray, G. I., Miozzo, M., Johnson, K., Schiller, N. O., & McMahon, K. L. (2012). Independent Distractor Frequency and Age-of-Acquisition Effects in Picture–Word Interference: fMRI Evidence for Post-lexical and Lexical Accounts according to Distractor Type. *Journal of Cognitive Neuroscience, 24(2),* 482-495.

Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24,* 284-298.

Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2009). Subtitle-based word frequencies as the best estimate of reading behavior: the case of greek. *Frontiers in Psychology, 1*.

Ellis, A. W. (2006). Word finding in the damaged brain: Probing Marshall's caveat. *Cortex, 42*, 817–822.

Ellis, A.W. (2012). The acquisition, retention, and loss of vocabulary in aphasia, dementia, and other neuropsychological conditions. In M. Faust (Ed.), *The handbook of the neuropsychology of language* (pp. 637-660) Oxford: Blackwell Publishing Ltd.

Ellis, A. W., Burani, C., Izura, C., Bromiley, A., & Venneri, A. (2006). Traces of vocabulary acquisition in the brain: Evidence from covert object naming. *Neuroimage, 33(3),* 958-968.

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(5),* 1103-1123.

Ellis, A. W., Lum, C., & Ralph, M. A. L. (1996). On the use of regression techniques for the analysis of single case aphasic data. *Journal of Neurolinguistics, 9(3),* 165-174.

Ellis, A.W., & Monaghan, J. (2002). Reply to Strain, Patterson, and Seidenberg (2002). *Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 28(1),* 215-220.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(2),* 515-523.

Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decision. *Quarterly Journal of Experimental Psychology,* .

Feyereisen, P., Van der Borght, F., & Seron, X. (1988). The operativity effect in naming: A reanalysis. *Neuropsychologia, 26*, 401 – 415.

Fiebach, C. J., Friederici, A. D., Müller, K., von Cramon, D. Y., & Hernandez, A. E. (2003). Distinct brain representations for early and late learned words. *NeuroImage, 19(4),* 1627-1637.

Forbes-McKay, K. E., Ellis, A. W., Shanks, M. F., & Venneri, A. (2005). The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age related cognitive decline. *Neuropsychologia, 43(11),* 1625-1632.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35(3),* 515-531.

Harvey, D.Y., & Schnur, T.T. (2015). Distinct loci of lexical and semantic access deficits in aphasia: Evidence from voxel based lesion-symptom mapping and diffusion tensor imaging. *Cortex, 67,* 37-58.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: its neural and computational mechanisms. *Psychological Bulletin, 133(4),* 638-650.

Hirsh, K. W., Morrison, C. M., Gaset, S., & Carnicer, E. (2003). Age of acquisition and speech production in L2. *Bilingualism: Language and Cognition, 6(2),* 117-128.

Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience, 23(9),* 2432-2446.

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition, 13(7-8),* 884-910.

Holmes, S. J., Fitch, F. J., & Ellis, A. W. (2006). Age of acquisition affects object recognition and naming in patients with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 28(6)*, 1010-1022.

Howes, D. (1964). Application of the word frequency concept to aphasia. In A. V. S. De Reuck & M. O’Connor (Eds.), *Disorders of language* (pp 47 – 75). Ciba Foundation Symposium. London: Churchill.

Howes, D.H., & Solomon, R.L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology, 41*, 401-410.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language*, *50(2)*, 165-181.

Johnston, R .A. & Barry, C. (2006) Age of Acquisition and lexical processing: A review. *Visual Cognition, 13(7-8)*, 789-845.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131(5)*, 684-712.

Kay, J., & Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia. *Brain, 110*, 613–629.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture - naming errors. *Cognitive Neuropsychology, 25*, 463-492.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence,RI: Brown University Press.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143(3)*, 1065-1081.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods, 44*, 978-990.

Laganaro, M., Di Pietro, M., & Schnider, A. (2006). What does recovery from anomia tell us about the underlying impairment: The case of similar anomic patterns and different recovery. *Neuropsychologia, 44*, 534–545.

Laganaro, M., Valente, A., & Perret, C. (2012). Time course of word production in fast and slow speakers: a high density ERP topographic study. *NeuroImage, 59(4)*, 3881-3888.

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(7-8)*, 928-948.

Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—what matters? *Neuropsychologia, 36(8)*, 775-784.

Lewis, M. B., & Vladeanu, M. (2006). What do we know about psycholinguistic effects? *The Quarterly Journal of Experimental Psychology, 59(6)*, 977-986.

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(1),* 126-133.

Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). Subtlex-pl: subtitle-based word frequency estimates for Polish. *Behavior Research Methods, 47(2),* 471-483.

Menenti, L., & Burani, C. (2007). What causes the effect of age of acquisition in lexical processing?. *Quarterly Journal of Experimental Psychology, 60(5)*, 652-660.

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(1)*, 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(4)*, 506-525.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology: Section A, 50(3),* 528-559.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76(2)*, 165-178.

Navarrete, E., Pastore, M., Valentini, R., & Peressotti, F. (2015). First learned words are not forgotten: Age-of-acquisition effects in the tip-of-the-tongue experience. *Memory & Cognition, 43(7)*, 1085-1103. Advance electronic publication, DOI 10.3758/s13421-015-0525-3.

Navarrete, E., Scaltritti, M., Mulatti, C., & Peressotti, F. (2013). Age-of-acquisition effects in delayed picture-naming tasks. *Psychonomic Bulletin & Review, 20(1)*, 148-153.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*, 661-677.

Newcombe, F., Oldfield, R. C., & Wingfield, A. R. (1965). Object naming by dysphasic patients. *Nature, 207*, 1217.

Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia, 33*, 1281–1303.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8(12)*, 976-987.

Pérez, M. A. & Navalón, C. (2005). Objective-AoA norms for 175 names in Spanish: Relationships with other psycholinguistic variables, estimated AoA, and data from other languages. *European Journal of Cognitive Psychology, 17,* 179-206.

Perret, C., Bonin, P., & Laganaro, M. (2014). Exploring the multiple-level hypothesis of AoA effects in spoken and written object naming using a topographic ERP analysis. *Brain and Language, 135*, 20-31.

Pind, J., & Tryggvadóttir, H. B. (2002). Determinants of picture naming times in Icelandic. *Scandinavian Journal of Psychology*, *43,* 221-226.

Pind, J., Jónsdóttir, H., Gissurardóttir, H., & Jónsson, F. (2000). Icelandic norms for the Snodgrass and Vanderwart (1980) pictures: Name and image agreement, familiarity, and age of acquisition. *Scandinavian Journal of Psychology*, *41*, 41-48.

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *Journal of General Psychology, 13*, 199-203.

Räling, R., Holzgrefe-Lang, J., Schröder, A., & Wartenburger, I. (2015). On the influence of typicality and age of acquisition on semantic processing: Diverging evidence from behavioural and ERP responses. *Neuropsychologia*, *75*, 186-210. Advance publication, DOI:10.1016/j.neuropsychologia.2015.05.031.

Rochford, G., & Williams, M. (1962). Studies in the development and breakdown of the use of names. *Journal of Neurology, Neurosurgery, and Psychiatry, 25*, 222-228.

Rochford, G., & Williams, M. (1965). Studies in the development and breakdown of the use of names. IV. The effects of word frequency. *Journal of Neurology, Neurosurgery, and Psychiatry, 28(5)*, 407-413.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object set: The role of surface detail in basic-level object recognition. Perception, 33, 217-236.

Rossiter, C. & Best, W. (2013). "Penguins don't fly": An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology, 27 (7)*, 784 - 798.

Ruff, I. M., Swanson, S. J., Hammeke, T. A., Sabsevitz, D., Mueller, W. M., & Morris, G. L. (2007). Predictors of naming decline after dominant temporal lobectomy: Age at onset of epilepsy and age of word acquisition. *Epilepsy & Behavior, 10(2),* 272-277.

Sage, K., & Ellis, A. W. (2004). Lexical influences in graphemic buffer disorder. *Cognitive Neuropsychology, 21(2-4),* 381-400.

Sailor, K. M., Zimmerman, M. E., & Sanders, A. E. (2011). Differential impacts of age of acquisition on letter and semantic fluency in Alzheimer's disease patients and healthy older adults. *The Quarterly Journal of Experimental Psychology, 64(12)*, 2383-2391.

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119(2),* 159-187.

Soares, A. P., Machado, J., Costa, A., Iriarte, A., Simões, A., Almeida, J. J., Comesaña, M., & Perea, M. (2015). On the advantages of frequency measures extracted from subtitles: The case of Portuguese. *Quarterly Journal of Experimental Psychology, 68*, 680-696

Steyvers, M., & Tenenbaum, J. B. (2005). The large‐scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29(1)*, 41-78.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language, 42(1)*, 51-73.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... & Bates, E. (2005). Timed action and object naming. *Cortex, 41(1)*, 7-25.

Tustin, K., & Hayne, H. (2010). Defining the boundary: age-related changes in childhood amnesia. *Developmental Psychology, 46(5),* 1049-1061.

Urooj, U., Cornelissen, P. L., Simpson, M. I., Wheat, K. L., Woods, W., Barca, L., & Ellis, A. W. (2014). Interactions between visual and semantic processing during object recognition revealed by modulatory effects of age of acquisition. *NeuroImage, 87,* 252-264.

van Heuven, W. J. B., Mandera, P., Keuleers, M., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176-1190.

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

Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon Ralph, M. A. (2010). The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. *Neuropsychologia, 48(6),* 1689-1696.

Woollams, A. M. (2012). Apples are not the only fruit: The effects of concept typicality on semantic representation in the anterior temporal lobe. *Frontiers in Human Neuroscience, 6*.

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia, 46(10)*, 2503-2514.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and language, 47(1),* 1-29.

Zhou, X., Liang, H., Xu, M., & Luo, B. (2009). Determinants of lexical access in pure anomic recovery: A longitudinal study. *Journal of Zheijian University Science B, 10,* 341–347.