Are two words recalled or recognised as one? How age-of-acquisition affects memory for compound words

Elsherif, Mahmoud M.; Catling, Jonathan C.

DOI:
10.1016/j.jml.2023.104449

Journal of Memory and Language 132 (2023) 104449

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Are two words recalled or recognised as one? How age-of-acquisition affects memory for compound words

Mahmoud M. Elsherif [a,b,*,†], Jonathan C. Catling [a]

[a] School of Psychology, University of Birmingham, United Kingdom
[b] Department of Vision Sciences, University of Leicester, United Kingdom

ARTICLE INFO

Keywords:
Age of acquisition, free recall, recognition, compound word
Memory
Semantics
Morphology

ABSTRACT

The age at which a person acquires knowledge of an item is a strong predictor of item retrieval, hereon defined as the Age of Acquisition (AoA) effect. This effect is such that early-acquired words are processed more quickly and accurately than late-acquired items. One account to explain this effect is the integrated account, where the AoA effect occurs in the early processes of lexical retrieval and hence should increase in tasks necessitating greater semantic processing. Importantly, this account has been applied to lexical processing, but not, to date, memory tasks. The current study aimed to assess whether the integrated account could explain memory tasks, using compound words, which differ from monomorphemic words regarding ease of mapping and semantic processes. Four-hundred-and-eighty participants were split into four groups of 120 participants for each of four experiments. Participants were required to recall unspaced and spaced compound words (Experiments 1 and 2, respectively) or make a recognition decision for unspaced and spaced compound words (Experiments 3 and 4, respectively). This approach allowed us to establish how semantic processing was involved in recalling and recognising the items. We found that (AoA) was related to all tasks such that irrespective of space, early-acquired compound words were recalled more accurately than late-acquired compound words in free recall. In recognition memory, late-acquired compound words were recognised more accurately than early-acquired compound words. However, the slope for the AoA was semantic processing influenced free recall to a greater extent than the recognition memory, with the AoA effect being larger in free recall than recognition memory. In addition, the AoA effect for the compound word was larger in spaced compound words than unspaced compound words. This demonstrates that the AoA effect in memory has multiple sources.

What is the AoA effect?

The age at or order wherein a person learns words, phrases and objects is known as the Age of Acquisition (AoA) effect. In lexical processing, it is a truism that early-acquired stimuli are processed more quickly and accurately than late-acquired stimuli (e.g. Elsherif et al., 2023). The AoA effect has been demonstrated in many tasks including recognition without identification (e.g., Catling et al., 2021), word naming (e.g. Ellis & Morrison, 1998), phrasal lexical decision (Arnon et al., 2017) and progressive demasking (Ploetz & Yates, 2016). For more information about the AoA effects in lexical tasks and neurodiverse populations such as aphasic patients see reviews by Brysbaert and Ellis (2016) and Elsherif et al. (2023).

Historically, it has been debated whether the subjective ratings of AoA is a genuine variable or the result of a variable not controlled for (e.g. cumulative frequency and/or frequency trajectory; Zevin & Seidenberg, 2002, 2004), though, in the main, these concerns have been quashed, as it has been demonstrated that the AoA effects cannot be simply reduced to these other variables or any other lexical-semantic variables (e.g., Brysbaert, 2017). There is, however, evidence indicating that there is a potential common origin shared between AoA and frequency (i.e. lexical-semantic level), as the magnitude of both predictors increase with greater semantic processing (see review by Elsherif et al., 2023).

Theories of the AoA effect

There are three predominant theories that explains the nature of the

* Corresponding author at: School of Psychology, University of Birmingham, 52 Pritchatts Road, Birmingham B15 2TT, United Kingdom.
E-mail address: mahmoud.medhat.elsherif@gmail.com (M.M. Elsherif).
† Mahmoud Elsherif has moved to Department of Vision Sciences, University of Leicester, University Road, Leicester LE1 7RHA, United Kingdom.

https://doi.org/10.1016/j.jml.2023.104449
Received 10 January 2023; Received in revised form 23 June 2023; Accepted 26 June 2023
Available online 4 July 2023

AoA effect: the representation theory, the mapping theory and the integrated account of the AoA effect. According to the representation theory (Van Loon-Vervoorn, 1989 cited in Brysbaert et al., 2000), the order in which concepts are acquired influences structure of the conceptual knowledge. At the hub or centre of the network are early-acquired concepts. Early-acquired concepts have a greater number of semantic connections to other concepts, affording them richer semantic representations than late-acquired concepts (Henry & Kuperman, 2013; Steyvers & Tenenbaum, 2005). In turn, early-acquired words are less likely to be forgotten, are more easily processed and more accessible (see reviews by Henry & Kuperman, 2013; Marful et al., 2016; Steyvers & Tenenbaum, 2005). Put simply, within this account, the AoA effect occurs as a result of the semantic representation gradually developing over time. Evidence in favour of this account has demonstrated that the AoA effect increases with greater semantic processing, as the AoA effect is observed to be smaller in tasks such as lexical decision compared with visual word naming (see review by Elsherif et al., 2023 but see the opposite pattern in logographic languages; e.g. Chang & Lee, 2020; Havelka & Tomita, 2006). One explanation for this is that lexical decision necessitates access to the lexical-semantic representation in order to recognise words and reject nonwords, while word naming accesses letter-to-sound corresponding without necessarily accessing semantics (Cortese et al., 2018; Kuperman, 2013).

Alternatively, there is the mapping theory, which is not incompatible with the representation theory (e.g. Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; P. Monaghan & Ellis, 2010). In the mapping theory, prior to the entry (learning) of any words, the neural network has rich resources that enables early-acquired words to be well established within the mental lexicon. As a result of early-acquired words entering the neural network, the connections between input and output representations are modified. In turn, Late-acquired words items have to fit into the network's structure, despite the fact that the resources for consolidating the lexical representation of late-acquired items are not as rich. If the letter-sound correspondence is systematic and regular (e.g. visual word naming), late-acquired words can benefit from the structure formed by early-acquired words, whereas when the relationship is arbitrary (e.g. spoken and written picture naming), late-acquired words face a processing cost (Zevin & Seidenberg, 2002). Supporting evidence has shown that visual word naming produces smaller AoA effects than spoken picture naming, as the relationship for the former has a regular letter-sound correspondence and the latter has a more arbitrary relationship, thus cannot gain from any general rules built up due to regularity of input–output coding (e.g. Catling & Elsherif, 2020; Lambon Ralph & Ehsan, 2006).

The final theory is a hybrid approach of the representation theory and mapping theory (Chang et al., 2019; Chang & Lee, 2020; Cortese et al., 2020; Dirix & Duyck, 2017a,b; Elsherif & Catling, 2021; Elsherif et al., 2023; Menenti & Burani, 2007). This account argues that the AoA effect occurs as a result of a combination of the formation of the semantic representation and changing plasticity in the neural network throughout development (Chang et al., 2019). Chang and Lee (2020) noted the AoA effect is smaller in lexical decision tasks than character naming tasks. The former was less influenced by semantic processing than the latter, and also the latter showed an interaction of regularity and AoA such that the AoA effect was stronger in irregular than in regular and systematic words. The integrated account of the AoA effect argues that early-acquired words are afforded a processing advantage, as they have more connections and greater accessibility than late-acquired words, thus the processing advantage is shaped gradually by the experience of learning during development (Brysbaert & Ellis, 2016; Dirix & Duyck, 2017). This theory has been limited to monomorphemic items and can be strengthened by generalising the pattern of findings to morphologically complex words such as compound words. By assessing compound words, one can assess the involvement of lexical and semantic elements during lexical and lexemic processing.

Is the AoA effect observed in compound words?

Compound words (e.g. drugstore) are formed from two lexemes: the modifier (e.g. drug in drugstore) and head (i.e. store in drugstore). The head lexeme contains information relating to the semantic category and morphosyntactic features of the whole compound and makes the whole compound a noun. Different languages vary as a function depending on whether compounds are left- or right-headed. In English, the majority of compounds are right-headed (Plag, 2018; Williams, 1981). The second lexeme is the modifier that narrows the semantic domain (i.e. meaning) of the head noun (e.g. drugstore is a store that sells drugs; Benczes, 2005, 2014, 2015; Günther et al., 2020). Compared to monosyllabic words, disyllabic words such as compound words have more irregular spelling-to-sound correspondence, thus readers need more time to process the pronunciation (Cortese & Schock, 2013). Despite this difference in terms of processing, we know that the subjective ratings of AoA contributes to lexical retrieval of compound words (Bonin et al., 2022; Elsherif et al., 2020; Elsherif & Catling, 2021; Juhasz et al., 2015, Juhasz, 2018). Compound word processing easily fits under the integrated account in that compound words are more irregular in terms of letter and sound, thus more time is required to process pronunciation. As a result, semantic processing becomes more apparent and begins to influence the processing of the compound words (Elsherif & Catling, 2021).

It should be noted that compound words cannot only be presented as one lexical item but can have a space between lexemes to induce morphological decomposition. Using the morphological decomposition route, processing would lead to the meaning of the separate lexemes within a compound word to be compared with the meaning of the compound word. This extra semantic step that participants are likely to make in this paradigm result in semantic variables such as imageability and familiarity having more influence, as participants name the modifier before the head has been fully processed. Elsherif et al. (2020) investigated the properties of combinatorial naming and word naming. The authors observed that AoA contributed to word naming latencies. However, when the same compounds were presented with a space, imageability, AoA and familiarity of the compound word, together with the AoA and imageability of the modifier and head lexeme affected the speed to name the compound word. The authors argued that when a space is included, participants process the meaning of the separate lexemes within a compound word and compare it to the meaning of the whole compound word, otherwise the separate lexemes would compete with the meaning of the compound word. This is followed by the integration of the two meanings, leading to greater semantic processing during combinatorial processing. Despite this explanation, these findings and the integrated account of the AoA effect has been limited to lexical processing. They have not been extended to other cognitive domains such as memory, where the AoA effect is more unclear.

The AoA effect within the memory domain

Although the AoA effects in lexical processes have been reliably documented, the effects of AoA within the memory domain have been equivocal. For instance, several studies find the the AoA scores to be positively related to recall or recognition memory, where later-acquired words were associated with better performance (e.g., Cortese et al., 2010; Cortese et al., 2015; Dewhurst et al., 1998; Lau et al., 2018; Macmillan et al., 2021; Morris, 1981 but see Kilecioğlu et al., 2020; Momenian, 2022; Raman et al., 2018; Volkovyskaya et al., 2017 who observed early-acquired words were related to better performance), while others have found no significant effects of AoA on both free recall and recognition memory (e.g., Coltheart & Winograd, 1986; Gilhooly & Gilhooly, 1979; Rubin, 1980). However, the results of previous studies, excluding recall and recognition memory megastudies (Cortese et al., 2010; Cortese et al., 2015; Lau et al., 2018; Macmillan et al., 2021) were limited by small sample sizes for both participants and items, together with substantial methodological variance, coupled with participant heterogeneity, this all potentially adds noise to the data and may obscure AoA effects.

In two megastudies on recognition memory, Cortese and colleagues (2010, 2015) used regression analyses on recognition of approximately 2500 monosyllabic and bisyllabic words, respectively. Cortese and colleagues observed that late-acquired words were recognised better than early-acquired words, reflecting a reverse AoA effect. However, the findings were limited to recognition memory, and thus were unable to disentangle task-specific and task-general item-level properties. Using the same approach as Cortese et al., Lau et al. (2018) conducted recognition and recall tasks on 532 concrete words. The authors replicated the findings of Cortese et al. such that late-acquired words were recognised better than early-acquired words. However, the AoA effect was noted not to be present in the free recall task.

Taken together, depending on the type of memory task, lexical predictors (e.g. frequency and AoA) do predict memory performance. Furthermore, the influence of lexical-semantic predictors (e.g. word frequency) are noted to be larger in free recall than in recognition tasks, while sublexical and structural factors such as word length show the opposite pattern (Lau et al., 2018). One explanation is that free recall requires processes that involve reconstruction of degraded word traces after encoding or searching for the lexical item in memory during effortful retrieval (e.g. Roodenrys et al., 1994). As a result, free recall may be less prone to low-level properties of words such as the orthographic features (Hunt & Elliot, 1980). Recognition memory, however, is sensitive to low-level processes such as perceptual, orthographic and phonology properties, which impact familiarity processes underlying recognition memory performance (Lau et al., 2018; Wilson et al., 2018). In turn, semantic processing is less likely to be involved in this task, and the structural properties of words (e.g. word length) are more likely to affect recognition memory (Lau et al., 2018).

Although these findings are informative in understanding the AoA effect relating to memory processes, these studies do require further generalisability. Although Cortese et al.'s (2010, 2015) and Lau et al. (2018) are an improvement over previous research, as participants were presented to all words in a recognition memory experiment with a large sample size in terms of items, and better measures of word frequency, AoA, and other lexical-semantic properties, they are not optimal for studying the AoA effects within morphemes, as they only focused on the lexical properties, as opposed to lexical and morphemic properties of memory. As a result, our understanding is limited to the surface-level predictors of morphologically complex words, such as compound words, but does not allow access to deeper level processes. Compound words allow us to assess the involvement of lexical-semantic item properties during lexical and lexemic processing. Finally, an important issue that remains to be addressed is the mutual influence of semantic and perceptual predictors on the common and task-specific processes in the two types of memory tasks.

We applied signal detection theory (SDT; Green & Swets, 1966) in the recognition memory task to analyse and dissociate two conceptually distinct measures of performance. The first measure is sensitivity/discriminability (d'). The d' is an indicator for a participant to make a binary distinction between signal (i.e. word recognised in study phase) and noise (i.e. word not recognised in study phase) to detect words in different contexts. The d' is computed based on the difference between hit rates (i.e. correct detection of the item when they are present) and false alarms (false detection of the items when they were absent), providing an unbiased measure for bottom-up processes. The second measure of SDT is Response bias (c). The response bias is a measure of the likelihood a participant favours one response over the other, reflecting top-down strategic component of the decision that is orthogonal to the ability to discriminate the target among noise (Stanislaw & Todorov, 1999). In contrast to previous and more traditional approaches for SDT, which used general linear models with responses aggregated across trials for each participant to produce hit rates, miss rates, or false alarm rates for each participants, using by-subject and by-item analysis (see Elsherif et al., 2017 as an example on emotion recognition), disregards variance across and within participants and stimuli (Rabe, 2018). Previous studies (e.g. Arsal et al., 2021; Poon, 2020; Rabe, 2018) used trial level data to model fixed and random effects within a signal detection model for categorical response data. Within the current study, we used general linear mixed models to consider all data points, while modelling psycholinguistic characteristics of the specific item with d' and c as dependent measures.

The AoA literature in respect to memory is dominated by the representation theory. However, structural properties are also involved in recognition memory (e.g. Lau et al., 2018). The AoA effect in memory could have multiple sources; however, to date, this has not been investigated. The present study used the integrated account of the AoA effect, to assess the role of AoA in recognition memory and free recall with and without a space inserted between lexemes. This study assessed the following predictions:

Based on previous findings (e.g. Lau et al., 2018; Macmillan et al., 2021) and the item-noise model (McClelland and Chappell, 1998), it is suggested that memory is the result of storing individual features that each have a specific probability of being stored. Their features can exist at any orthographic, phonological and/or semantic level. At test, the features present in the word are compared to the stored memory representation. When there is high overlap between the two, then the item is more likely to be judged as 'old' than if there is a low overlap. In recognition memory, the memory representations are going to be weaker for earlier-acquired words, as they share orthographic, phonological and semantic features with many competitors than later-acquired words. Relative to early-acquired words, later-acquired words might be less noisy (i.e. have a higher probability of being stored, aiding recognition), thus are more likely to be distinctive, leading late-acquired words to be recognised more accurately than early-acquired compound words. As a result of fewer features being involved to cue the items, we predict that the AoA effect would not be demonstrated in free recall.

According to the representation theory, the memory representation is weaker for earlier-acquired words, as they share semantic features and have more connections to other words and semantic concepts, leading to richer semantic representations than late-acquired words. As a result, the semantic features of later-acquired words might be less noisy. First, we would predict a reversed AoA effect such that late-acquired words are recognised more accurately than early-acquired words. In addition, compared to recognition memory, free recall requires degraded word traces to be reconstructed after encoding for the lexical item in memory during effortful retrieval (e.g. Ballot et al., 2021; Hunt & Elliot, 1980; Poirier & Saint-Aubin, 1995; Lau et al., 2018). We should therefore expect to observe the AoA effect in free recall *and* recognition memory task. In addition, the magnitude of effect sizes should be larger in free recall than recognition memory. The representation theory would predict that the magnitude of the AoA effect should be independent of whether (or not) a space is inserted between lexemes, as free recall allows the reader to ignore the perceptual properties of the word. However, recognition memory is more likely to be depend on the low-level perceptual properties, and the space between words is more likely to cause access to semantic processing and encourage access to the varying semantic pathways at the lexical and morphemic level (Elsherif et al., 2020; Elsherif & Catling, 2021).

According to the mapping theory, the representation for early-acquired words is weaker than that of later-acquired words, as early-acquired words would enter the neural network and become well established such that the input and output representations are modified based on the structure formed by the early-acquired words. This leads to

Table 1
Descriptive statistics for word target characteristics for compound word.

Predictors	Compound word			Modifier lexeme			Head lexeme		
	M	SD	Range	M	SD	Range	M	SD	Range
Word length	8.61	1.34	6–13	4.42	1.02	2–8	4.19	0.84	2–7
Frequency (out of 7)	2.67	0.72	0.696–4.48	4.50	0.79	2.69–6.78	4.69	0.82	2.38–7.42[1]
Familiarity (out of 7)	5.77	1.15	1.57–7.00	4.72	1.66	1–6.41	4.98	1.50	1–6.43
Imageability (out of 7)	4.28	1.61	1.05–6.95	4.95	1.47	1.40–7.00	5.07	1.44	1.40–6.90
AoA (out of 7)	4.70	1.22	1.93–7.00	3.54	0.86	1.70–6.10	3.35	0.82	2–6.30
ST (out of 7)	4.59		1.33		1.6–6.71		NA		NA
LMD (out of 10)	5.17		1.42		1.47–8.67		NA		NA

Note. AoA = Age-of-Acquisition, ST = Semantic Transparency and LMD = Lexeme Meaning Dominance. Although there is a discrepancy between the maximum and range of values shown in this table, this discrepancy is from the van Heuven et al.'s SUBTLEX-UK online database and the Likert scale used (scale of 1–7 discussed on their website). We had used the word "to", which had a Zipf scale of 7.42 and function words tend to go beyond the maximum score.

richer and more well-established lexical representations for early-acquired than late-acquired words. In turn, late-acquired words should show an advantage over early-acquired words, as late-acquired words have a distinct lexical representation and would thus be less noisy. However, the inclusion of a space for a compound word is also more likely to make the perceptual properties of the item in recognition memory more salient and distinctive, the mapping between letter and sound would be more irregular, thus incurring a processing cost. In turn, this should lead to a larger and a steeper slope for the AoA effect in spaced compound words than unspaced compound words. In addition, the AoA of the lexemes should be more likely to have an impact in the recognition memory task with a space than that without a space, as participants would be encouraged to use the morphemic pathway, as opposed to the lexical, pathway (Elsherif et al., 2020). Finally, the mapping theory would also predict that the AoA effect should be observed in free recall tasks, as longer and more complex words such as compound words have less regular letter-sound mapping, thus semantic processing is more likely to contribute to memory performance of such words.

Dataa availability

The experimental data and R analysis files can be downloaded via the Open Science Framework at https://osf.io/qu9hc/.

General method overview

Power analysis

To reduce experimenter bias, the data were analysed after all the participants were recruited and a stopping rule was introduced. A power analysis, using Westfall et al.'s (2014) formula for effect size calculation based on Elsherif et al. (2020) was conducted (see Brysbaert & Stevens, 2018, for discussion), indicating that our sample size should exceed the number required to reach the desired level of power of 0.80 (minimum of 13 recommended, while we included the data from 48 participants in the analyses) for a word naming study. However, as accuracy is the main focus in these tasks, this was increased to 120 participants per experiment in order to increase the number of observations and ensure the minimum number had been met.

Experiment 1: Recall memory without a space

The purpose of Experiment 1 was to assess whether AoA influences recall memory when compound words are used. We could not find previous published research that examined this. We therefore used the design of Lau et al. (2018), who found effects of frequency and imageability in free recall in monomorphemic words. We predicted that the AoA of the modifier and the compound word would contribute to the free recall memory performance, as it requires additional semantic processing in order to reconstruct the degraded representations that requires effortful lexical-semantic processing (Lau et al., 2018).

Method

Participants

One-hundred twenty British monolingual undergraduate students, aged 18–27 years (M = 18.53 ± 1.19 years; 98 females). Participants were remunerated with course credits. The experiment was conducted in accordance with the British Psychological Society's ethical guidelines and approved by the University's ethical committee. All participants had normal or corrected-to-normal vision and signed a consent form to participate in the study.

Materials

The stimuli are described in detail elsewhere (Elsherif et al., 2020, Elsherif & Catling, 2021). Briefly, each participant saw 226 words that were noun-noun compounds (see supplementary materials). We extracted word frequencies as Zipf values for the compound word, modifier and head lexemes from the SUBTLEX-UK database (van Heuven et al., 2014)[1]. Letter length of the compound word was obtained by calculating the length of the word. From Juhasz et al.'s (2015) database, we obtained the AoA and imageability (i.e. the ease that a word evokes a verbal or non-verbal mental image) of the compound word, together with the semantic transparency (ST, the extent that the meaning of the compound word is based on the meaning of the head and modifier lexeme e.g. door to doorbell), and lexeme meaning dominance (LMD, whether the meaning of the compound word is contained in the modifier or head lexeme). The AoA, imageability and familiarity of the modifier and head are all based on subjective ratings, and were taken from the respective database (AoA: Cortese & Khanna, 2008; Schock, Cortese & Khanna, 2012; imageability: Cortese & Fugett, 2004; Schock, Cortese, Khanna et al., 2012; familiarity: Balota et al., 2001) (See Table 1 for psycholinguistic properties).

Design

There were two additional concerns that needed to be considered. The number of study lists needed to be a factor of 226 (i.e. total number

[1] According to Van Heuven et al. (2014), any words not found in SUBTLEX-UK were given a word frequency value of 0.696. This was given to only seven words ('oxcart', 'prizefight', 'turtledove', 'castoff', 'carryall', 'filmstrip' and 'campground').

of studied items). Second, we had to ensure the number of items used in the study phase would not produce ceiling and floor effects in free recall task or recognition memory task. In a typical free recall task, a list length of 10 or more study items is typically used. Increasing the number of items within a list would lead to poorer recall performance (e.g. Gillund & Shiffrin, 1984; Grenfell-Essam & Ward, 2012). Considering ceiling and floor effect constraints, a list length of 19 study items was chosen for the free recall task. As a result, for the free recall task, study lists 1–11 consisted of 19 items, whilst for study list 12 consisted of 17 items.

Procedure

The free recall tasks and recognition memory were conducted using E-Prime version 2 (Schneider et al., 2002). The experiments for free recall were pre-registered: https://osf.io/aqx7m and for recognition memory were pre-registered: https://osf.io/u7d9x.

Study Phase. After providing informed consent, participants received instructions for the study phase. Participants were told they would see a series of words and that it was critical to pay attention to each word. They were told to try to memorise as many words as possible. In each study trial, a fixation point appeared for 500 ms, after which a stimulus was presented. For each participant, words were randomly sampled without replacement across all lists. The order of presentation of words within each list was randomised. Following Lau et al. (2018), we used a presentation duration of 1500 ms where each word was presented at the centre of the screen in a sequential manner. Afterwards, a blank inter-trial interval screen followed each stimulus presentation, lasting for 1000 ms.

Distractor phase. Between each study phase and each test phase, a distractor task followed, wherein participants completed simple mathematics verification problems. These instructions occurred both in written and verbal form for all participants. Participants were given the opportunity to ask questions prior to the study. This distractor task involved participants verifying 18 simple mathematic problems (e.g., [6/2] + 4 = 7?). Participants were informed to press the "m" and "z" keys for correct and incorrect solutions, respectively.

Test Phase. In respect to the recall memory task, in the testing phase, participants were given a screen with 50 blank spaces and were instructed to recall as many words as they could from the study list. Participants were given five minutes to recall as many words as possible in any order immediately after the presentation of each 19 word list on the computer. Participants were told that points would be awarded for each correct answer and not to worry about typos or near-correct responses. Accuracy was assessed in the free recall. A correctly recalled items was defined as the items being written down correctly, however, if there was any typo then they were coded as correctly recalled items and only differ depending on orthography. They would be coded as incorrect if the item had two different lexemes that were included (e.g. airfire), or the item was not recalled during the test phase. The experiment took approximately 45 min.

Analysis

All analyses were conducted in R 3.6.1. (R Core Team, 2019), using the "tidyverse" package version 1.3.0 for data processing (Wickham et al., 2019). We analysed the data using linear mixed-effect models (LMM) using the lme4 package, version 1.1.26 (Bates et al., 2015). T values were computed for each variable of interest and a variable was significant at the alpha =.05 level if the absolute t value was greater than 1.96 (Baayen et al., 2008). In model summaries, we report both marginal (R^2m) and conditional R^2 values (R^2c), using the "MuMIn" package version 1.43.17 (Barton & Barton, 2015). The value obtained from R^2m reflects the estimate of the variance explained by fixed factors only, whereas the latter explains the variance of the whole model (i.e. fixed and random factors; Nakagawa & Schielzeth, 2013).

The LMM is conducted on accuracy in each trial (coded as 1 or 0 for correct or incorrect, respectively), as linear regression coefficients are directly interpretable when interactions or fixed effects are included, making LMM safer to use. A LMM was used as opposed to using a general logistic mixed model since the former produces unbiased estimates of effects on binary outcomes (Gomila, 2021). The following LMM analyses were conducted separately for free recall and recognition memory, with item and subject as random factors and all the predictors as fixed factors All the predictors were centred as their means.

We used a benchmark model that included frequency and word length of the compound word and the frequency of the modifier and head (Elsherif et al., 2020; Juhasz et al., 2015; Kuperman, 2013).[2] For all LMM models, collinearity diagnostic analyses showed a variance inflation factor (VIF) of 1.09. In addition, we checked the extent to which the AoA of the compound word and its lexemes could explain the variance beyond the main linguistic processing predictors (without the fear of collinearity). Each variable of interest was added separately to the model. The correlation matrix between the variables is shown in Table 2.

Results

The data and materials for all experiments are available at https://osf.io/qu9hc/.

Recall memory without a space (Experiment 1)

Following the advice of Woods et al. (2023), across the samples, the rate of missingness was less than 1% and was missing at random. The mean recall accuracy for recall memory without a space was 0.406 (SD = 0.491; 95% CI[0.40, 0.41]). In the baseline model for the recall memory without a space, the frequency of the compound word and head lexeme [3]made significant contribution to free recall without a space accuracy (see Table 3). The more frequent the compound word, the more likely it would be recalled, whereas the more frequent the head lexeme of the compound word, the less likely it would be recalled. Familiarity, AoA, imageability and semantic transparency of the compound word together with the imageability, and AoA of the modifier and head lexeme contributed to free recall without a space accuracy (see Table 4). The more familiar, imageable, and semantically transparent the compound word, the more likely they would be recalled. In addition, the earlier the compound words AoA, the more likely they will be recalled. The earlier the modifier and head lexemes of a compound word AoA, the more likely they will be recalled. However, the familiarity of the modifier lexeme, LMD, word length and familiarity of the head lexeme did not significantly contribute to the free recall without a space between lexemes accuracy. These results will be discussed after Experiment 2.

Experiment 2: Recall memory with a space

Experiment 1 found that early-acquired compound words, together

[2] We deviate from the pre-registration for all four experiments, as the inclusion of phonetic complexity as an additional measure can make the outcome of the analysis less stable, making it more difficult to interpret the findings. (e.g. Wysocki et al., 2022). The VIF when phonetic complexity was included 6.15 driven by alveolar initial phoneme onset. Once removed, the VIF produced a value of 1.55. In line with Open Scholarship Principle and to showcase transparency, we have included the analyses for accuracy with phonetic complexity in the supplementary material.

[3] Additionally, to follow the modelling approaches of previous research such as Elsherif et al. (2019), Elsherif and Catling (2021), Juhasz et al. (2015), Juhasz (2018) and Kuperman (2013), we coded initial phoneme onset onset variables: bilabial, labiodental, dental, labiovelar, postalveolar, alveolar, palatal, palatal.alveolar, glottal, velar and voiced as either a presence of the feature (1) and absence of the feature (-1). Although this is conducted in naming studies to control for initial phoneme characteristics, it was conducted to ensure that the effects in memory were not a result of phonetic complexity not being controlled. In all experiments, we observed that the findings were not due to these effects (see supplementary materials for more detail).

Table 2
Correlations between independent variables.

	CL	CF	CFA	CAOA	CI	ST	LMD	ML	MF	MFA	MAoA	MI	HL	HF	HFA	HAOA
CF	0.01															
CFA	0.07	0.33***														
CAoA	0.04	−0.20**	−0.65***													
CI	0.04	0.04	0.54***	−0.66***												
ST	0.08	−0.08	0.37***	−0.37***	0.45***											
LMD	−0.06	−0.01	0.00	−0.06	0.24***	0.03										
ML	0.78***	−0.02	0.05	−0.02	0.04	0.04	−0.07									
MF	−0.08	0.27***	0.22***	−0.11	−0.10	0.11	−0.10	−0.11								
MFA	0.06	0.09	0.19**	−0.17**	0.13^{+}	0.21**	−0.09	0.08	0.49***							
MAoA	0.23***	−0.09	−0.15*	0.30***	−0.12^{+}	−0.16*	0.00	0.28***	−0.59***	−0.50***						
MI	0.10	−0.13^{+}	0.03	−0.24***	0.42***	0.22***	0.08	0.14*	−0.13^{+}	0.18**	−0.31***					
HL	0.65***	0.04	0.14*	0.09	0.01	0.08	0.01	0.03	0.00	0.00	0.03	−0.01				
HF	−0.12^{+}	0.15*	0.00	−0.01	−0.15*	0.09	−0.17**	0.02	0.21***	0.13^{+}	−0.06	−0.19**	−0.21***			
HFA	0.01	0.09	0.10	−0.02	−0.01	0.23***	0.06	0.03	0.05	0.09	−0.02	−0.01	−0.02	0.55***		
HAoA	0.17*	0.01	−0.11	0.08	−0.12^{+}	−0.08	−0.14*	0.02	−0.05	−0.05	0.11	−0.07	0.24***	−0.55***	−0.41***	
HMI	0.03	−0.09	0.14*	−0.20***	0.48***	0.15*	0.38***	−0.07	−0.19**	−0.05	−0.07	0.34***	0.14*	−0.25***	0.13*	−0.34***

Note. CL = compound word length, CF = compound word frequency, CFA = Compound word familiarity, CAoA = Compound age-of-acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance, ML = length of the modifier, MF = frequency of the modifier, MAoA = age-of-acquisition of the modifier, MFA = familiarity of the modifier, MI = imageability of the modifier, HL = length of the head, HF = frequency of the head lexeme, HFA = familiarity of the head, HAoA = age-of-acquisition of the head, HI = imageability of the head. $^{+}$ p <.10, * p <.05, ** p <.01 and *** p <.001.

Table 3
Linear mixed effects regression results for the baseline model for Experiment 1.[a]

Values	CL	CFreq	MFreq	HFreq
β	0.001	0.040	−0.005	−0.015
SE	0.006	0.007	0.007	0.007
2.5%CI	−0.011	0.026	−0.019	−0.029
97.5% CI	0.013	0.054	0.009	−0.001
t-value	0.23	6.14*	−0.73	−2.36*
R^2m	0.007			
R^2c	0.120			

Note. Coefficients and standard error are presented for each variable in the baseline model: compound word length (CL), compound word frequency (CFreq), modifier frequency (MFreq) and Head frequency (HFreq). CI = Confidence interval and SE = Standard error. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

with modifier and head lexemes were more likely to be recalled than late-acquired compound words, modifier and head lexemes. The purpose of Experiment 2 was to examine whether AoA influences free recall when a space is inserted between the lexemes. Although this method of presentation may appear trivial it induces morphological decomposition and forces readers to travel down the morphological decomposition route, thus the effects of the morphemes should become evident (Frisson et al., 2008). If morphological decomposition is involved, the effects of AoA, imageability and frequency should be larger, as semantic processing is more likely to be taxed on (Brooks & de Garcia, 2015; Elsherif et al., 2020). This occurs, as the meaning of the lexemes within a compound word, depending on semantic transparency, cannot easily be reassembled to give the meaning of the compound as a whole. Using a lexical decision task, Libben et al. (2003) noted that when compounds were presented as two separate words, opaque compound words took longer than transparent compound words. In turn, the meaning of two lexemes are accessed, which compete with and are compared to the meaning of the compound word, followed by the integration of the two meanings, leading to greater semantic processing during combinatorial processing (Brooks & de Garcia, 2015; Kuperman, 2013). This additional step in semantic processing allows semantic variables to become more influential, together with the individual lexeme. This finding has been demonstrated in word naming (Elsherif et al., 2020; Elsherif & Catling, 2021) but as of yet has not been shown in free recall.

Method

Participants

One-hundred twenty British monolingual undergraduate students, aged 18–24 years (M = 18.92 ± 0.99 years; 100 females). Participants were remunerated with course credits. The experiment was conducted in accordance with the British Psychological Society's ethical guidelines and approved by the University's ethical committee. All participants had normal or corrected-to-normal vision and signed a consent form to participate in the study. None of the students participated in the previous experiment.

Materials and procedures

The same materials and procedures were used as in Experiment 1 with the following exception: a space was inserted between the two lexemes of the compound (e.g. air plane). The insertion of the space was included in the study and test phase. Participants were informed that they would be simultaneously presented with two lexical strings which they had to recall as one word.

Table 4
Linear mixed effects regression results for the baseline model plus the additional 11 variables included for each separate model for Experiment 1.

Values	CFA[a]	CAoA[a]	CI[a]	ST[a]	LMD[a]	MFA[a]	MAoA[a]	MI[a]	HFA[a]	HAoA[a]	HI[a]
β	0.044	−0.042	0.039	0.020	−0.001	0.007	−0.002	0.012	−0.001	−0.017	0.010
SE	0.007	0.006	0.006	0.006	0.006	0.007	0.008	0.004	0.008	0.007	0.005
2.5%CI	0.030	−0.054	0.027	0.008	−0.013	−0.007	−0.018	0.004	−0.017	−0.031	0.0002
97.5% CI	0.058	−0.030	0.051	0.032	0.011	0.021	0.014	0.020	0.015	−0.003	0.020
t-value	6.65*	−7.10*	6.52*	3.18*	−0.18	0.93	−2.08*	2.85*	−0.14	−2.18*	2.16*
R^2m	0.013	0.014	0.013	0.008	0.007	0.007	0.008	0.008	0.007	0.008	0.008
R^2c	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.120	0.120

Note. Coefficients and standard error are presented for each variable when added separately to the baseline model containing compound word length, frequency, modifier frequency and head frequency. CI = Confidence interval; SE = Standard error; CFA = Compound word familiarity; CAoA = Compound word Age of acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance; MFA = Modifier familiarity, MAoA = Modifier age of acquisition, MI = Modifier imageability, HFA = Head familiarity, HAoA = Head age of acquisition and HI = Head imageability. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

Table 5
Linear mixed effects regression results for the baseline model for Experiment 2.[a]

Values	CL	CFreq	MFreq	HFreq
β	−0.006	0.042	−0.006	−0.015
SE	0.008	0.009	0.009	0.009
2.5%CI	−0.02	0.024	−0.024	−0.033
97.5% CI	0.01	0.060	0.012	0.003
t-value	−0.78	4.86*	−0.69	−1.74.
R^2m	0.007			
R^2c	0.122			

Note: Coefficients and standard error are presented for each variable in the baseline model: compound word length (CL), compound word frequency (CFreq), modifier frequency (MFreq) and Head frequency (HFreq). CI = Confidence interval and SE = Standard error. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

Results and discussion

Recall memory with a space (Experiment 2)

Following the advice of Woods et al. (2023), across the samples, the rate of missingness was less than 1% and was missing at random. The mean recall accuracy for recall memory without a space was 0.406 (SD = 0.491; 95% CI[0.40, 0.41]), while the mean recall accuracy for recall memory with a space was 0.404 (SD = 0.491; 95% CI[0.398, 0.41]. There was no significant difference between these conditions ($b = 0.002$, $t = 0.14$, $p = .89$). In the baseline model for the recall memory with a space between lexemes (see Table 5), the frequency of the compound word and head lexeme made significant contribution to free recall with a space accuracy. The more frequent the compound word, the more likely it would be recalled, whereas the more frequent the head lexeme of the compound word, the less likely it would be recalled. Familiarity, AoA, imageability and semantic transparency of the compound word together with the imageability, AoA and familiarity of the modifier lexeme contributed to free recall with a space accuracy (see Table 6). The imageability of the head lexeme contributed to free recall with a space between the lexemes' accuracy. The more familiar, imageable, and semantically transparent the compound word, the more likely they would be recalled. In addition, the earlier the compound words are acquired, the more likely they will be recalled. The earlier the modifier of a compound word is acquired, the more likely they will be recalled. The more imageable the head of the lexeme of a compound word, the more likely they will be recalled. However, the LMD, word length, familiarity and AoA of the head lexeme did not significantly contribute the free recall with a space between lexemes accuracy.

The estimate value for familiarity and imageability of the compound word, which has been argued to be a measure of lexical/semantic processes in compound word processing (e.g. Juhasz et al., 2015), was also numerically larger for the recall without a space than the recall with a space task (Tables 3–6), while the AoA, familiarity and imageability of the lexeme was stronger in recall with a space between lexemes than those without lexeme, indicating that lexical-semantic effects are stronger for recall without a space, while lexeme-semantic effects are stronger for recall with a space.

In Experiment 1 and 2, we found an effect of AoA on free recall. Participants were more likely to recall early-acquired compound words than late-acquired compound words. This finding replicates previous studies that have shown the AoA effect in free recall (e.g. Almond & Morrison, 2014; Raman et al., 2018). This demonstrates that the AoA effect is a robust variable that impacts free recall. This aligns with the integrated account (Chang & Lee, 2020; Elsherif et al., 2023), free recall requires processes that entail the reconstruction of degraded word representations, thus inducing an effortful search to locate a lexical item by activating a semantic representation of the word. This is demonstrated, as we have shown effects of imageability, frequency and familiarity within the free recall tasks such that high-frequency, imageable, familiar and semantically transparent items are more likely to be recalled than low-frequency, low imageable, unfamiliar and semantically opaque items. In addition, as semantics is likely to be involved, the AoA effect is likely to be magnified, as semantic processing is more likely to induce more arbitrary mappings between input and output (Chang et al., 2019). This indicates that the integrated account can explain the findings of the AoA effect in free recall. Also, based on these findings, we can conclude that the AoA effect contributes to free recall in compound words.

Experiment 3: Recognition memory without a space

Recognition memory is ideal to assess the generalisability of the AoA effects, as the task demands in recognition memory differs from free recall (e.g. Lau et al., 2018). Recognition memory is more susceptible to low-level processes such as orthographic sensitivity; thus we use recognition memory to assess the extent to which AoA effects within memory in compound words are due to differences in accessing the meaning of early- and late-acquired words (i.e. semantic processing) and/or differences in sensitivity to low-level processes (i.e. orthographic processing). Several papers have examined AoA on recognition memory. Cortese et al. (2010, 2015) and Lau et al. (2018) found that AoA predicted recognition memory. Experiment 3 was designed to assess whether the AoA affects recognition memory in compound words without a space inserted between lexemes. Based on these studies, we predict that there should be a late-acquired word advantage, as these items are semantically and orthographically distinct (Cortese et al., 2010, 2015).

Method

Participants

One-hundred twenty four British monolingual undergraduate students, aged 18–23 years (M = 18.76 ± 0.85 years; 106 females). Participants were remunerated with course credits. The experiment was conducted in accordance with the British Psychological Society's ethical guidelines and approved by the University's ethical committee. All

Table 6
Linear mixed effects regression results for the baseline model plus the additional 11 variables included for each separate model for Experiment 2.

Values	CFA[a]	CAoA[a]	CI[a]	ST[a]	LMD[a]	MFA[a]	MAoA[a]	MI[a]	HFA[a]	HAoA[a]	HI[a]
β	0.050	−0.049	0.046	0.028	−0.003	0.026	−0.022	0.023	−0.005	−0.012	0.011
SE	0.009	0.008	0.008	0.008	0.008	0.009	0.010	0.006	0.010	0.010	0.006
2.5%CI	0.030	−0.065	0.030	0.012	−0.019	0.008	−0.042	0.011	−0.0246	−0.032	−0.001
97.5% CI	0.068	−0.033	0.062	0.044	0.010	0.044	−0.002	0.035	−0.002	−0.002	0.001
t-value	5.57*	−6.37*	5.74*	3.43*	−0.31	2.81*	−2.13*	4.15*	−0.51	−1.22	1.87.
R^2m	0.015	0.017	0.015	0.010	0.007	0.009	0.008	0.012	0.007	0.008	0.008
R^2c	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122	0.122

Note. Coefficients and standard error are presented for each variable when added separately to the baseline model containing compound word length, frequency, modifier frequency and head frequency. CI = Confidence interval; SE = Standard error; CFA = Compound word familiarity; CAoA = Compound word Age of acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance; MFA = Modifier familiarity, MAoA = Modifier age of acquisition, MI = Modifier imageability, HFA = Head familiarity, HAoA = Head age of acquisition and HI = Head imageability. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

participants had normal or corrected-to-normal vision and signed a consent form to participate in the study. None of the students participated in the previous experiment. However, we excluded four participants, as they had a mean error rate above 60%.

Design

For the recognition memory task, we separated the words into two lists (i.e. Lists 1 and 2). These two lists were created to counterbalance the items across participants and to reduce target repetition but allow data collection for all targets. Each target was presented once in each list. The two lists were matched on all relevant variables. To ensure that both lists had the targets occurring as 'old' or 'new' equally often, half of the participants were randomly assigned to List 1 to be studied and List 2 to be unstudied words, and vice versa for the remaining participants. However, to collect responses for all items, there were two sessions, in which the participants who received List 1 in Session 1 received List 2 in Session 2 and the other half participants who received List 2 in Session 1 received List 1 in Session 2. The mean difference in timings of these sessions was three to seven days (M = 5.93).

Materials

The same materials were used as in Experiment 1. However, we noted that the number of items presented for free recall is likely to result in ceiling effects for recognition task (e.g. Lau et al., 2018). Cortese et al. (2010, 2015) had participants study 50 words per block and obtained an average score of.73. As a result, we increased it to 38 studied word to reduce ceiling effect and floor effects[4].

Procedures

The study and distractor phase from Experiments 1 and 2 were the same for the recognition memory task except that during the study phase, participants were first presented with 38 words randomly sampled without replacement from the 'old' list, with each word being presented for 1500 ms at the centre of the screen. The main difference was in the test phase, where participants were told they would again see a list of words and that points would be awarded for correctly stating if the item was old or new. Participants received the presentation of all 38 old words along with 38 new words (i.e. foils), pseudo-randomly interspersed within the testing block. Participants were asked to respond by pressing the "m" key labelled as old, if they had seen the item in the previous list or the "z" key labelled as new, if this is the first time a word occurred in this list. Participants were told that each word would occur precisely twice for both sessions, and thus there would be an equal chance of a stimulus being "Old" or "New". In a given trial, participants saw a stimulus word presented in black print on the centre of the screen for 1500 ms. If no response was made within 1500 ms, the experiment continued automatically. Following the response or timeout, there was an inter-trial interval of 1000 ms between their response and the onset of the next stimulus. Participants tracked their position in the experiment via a progress bar at the top of the screen. The experiment took approximately 45 min.

Analysis

The same analyses were conducted as Experiments 1 and 2. Also for the recognition memory tasks, we would favour using signal detection metrics such as *d*' and *c*, as we believe that as there is a trade-off such that any psycholinguistic effect for memory is spread between hits or false alarms and is captured only imperfectly in either measures[5]. *d*' is conducted by calculating the proportion of yes and no responses when a target is present or absent (Green & Swets, 1966; Macmillan & Creelman, 2005). A d' value of zero indicates chance-level performance, a positive *d*' value indicates better discriminability index and a negative value is indicative of performance below chance. When contrast coding is explicitly noted, then there is no requirement to include post-hoc testing to determine the directionality of effects, so model coefficients reflect the effect magnitude independently of other predictors' levels (see Brehm & Alday, 2022 for a detailed discussion of contrast coding in LMM). All categorical variables were deviation coded (+1, −1). Sensitivity was calculated by including IsOld (unstudied vs. studied, as −1 and 1, respectively). The slope for IsOld indicates the differences in response probability between studied and unstudied trials, and interactions with IsOld indicate changes in sensitivity (Rabe, 2018).

The second measure is response bias (β or criterion *c*), which is an indicator of a participant tendency to select one of the response alternatives. A *c* value of 0 provides a measure that it is an unbiased response. If $c > 0$, target detection is very like the result of a very conservative approach indicating participants were more likely to respond that the item was not studied, whereas if $c < 0$, target detection is likely the outcome of a very liberal approach, suggesting the item was studied. Put simply, the higher the response bias, participants were more likely to report the target as absent, leading to lower false alarms and hits, while the converse is demonstrated for the lower the response bias (Elsherif et al., 2017; Stanislaw & Todorov, 1999). Parameter estimates where IsStudied is absent, reflects response bias. Trial-level information was

[4] It is important to note that even though the list length in free recall is half of that of recognition memory, thus accuracy should be higher in the former than the latter. This was not the case, as accuracy was around 80% in the latter and 40% in the former. If the list lengths were the same between both tasks, the difference between tasks would have become larger.

[5] We deviate from the pre-registration for recognition memory by focusing solely on signal detection metrics, as they have been argued to be a better measure over solely hits and false alarm rates. However, in line with Open Scholarship Principle and to showcase transparency, we have included the analyses for accuracy with phonetic complexity in the supplementary material.

Table 7
Linear mixed effects regression results for the baseline model for Experiment 3 in *d*' (sensitivity) [a].

Values	CL	CFreq	MFreq	HFreq
β	−0.002	0.022	−0.033	−0.010
SE	0.003	0.004	0.004	0.003
2.5%CI	−0.010	0.014	−0.041	−0.016
97.5% CI	0.004	0.030	−0.025	−0.004
t-value	−0.58	6.20*	−9.34*	−2.86*
R^2m	0.005			
R^2c	0.080			

Note. Coefficients and standard error are presented for each variable in the baseline model: compound word length (CL), compound word frequency (CFreq), modifier frequency (MFreq) and Head frequency (HFreq). CI = Confidence interval and SE = Standard error. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

Table 9
Linear mixed effects regression results for the baseline model for Experiment 3 in c (response bias) [a].

Values	CL	CFreq	MFreq	HFreq
β	0.005	−0.013	0.015	−0.015
SE	0.003	0.003	0.003	0.003
2.5%CI	−0.0008	−0.019	0.009	−0.021
97.5% CI	0.011	−0.007	0.021	−0.009
t-value	1.71.	−4.07*	4.61*	−0.27
R^2m	0.005			
R^2c	0.080			

Note. Coefficients and standard error are presented for each variable in the baseline model: compound word length (CL), compound word frequency (CFreq), modifier frequency (MFreq) and Head frequency (HFreq). CI = Confidence interval and SE = Standard error. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

preserved, thus random variation among participants (i.e. by-subject random effects) and items (i.e. by-item random effects).

Results and discussion

Sensitivity

Following the advice of Woods et al. (2023), across the samples, the rate of missingness was less than 1% and was missing at random. In this analysis, we assess *d*' and *c*, each of those terms represent either a grand mean or condition effect on either response bias or sensitivity. The fixed effects (see Tables 7 and 8) capture the grand means of *d*' for the psycholinguistic effects in recognition memory without a space between lexemes. Sensitivity was unsurprisingly high on average ($b = 0.044$, $t = 13.14$, $p < .001$). In the baseline model for recognition without a space model, participants showed higher sensitivity to high-frequency compound words than low-frequency compounds, while there was higher sensitivity to low-frequency modifiers and head lexemes of the compound word than high-frequency modifiers and lexemes within the compound word. In addition, participants showed lower sensitivity to early-acquired, less imageable and less familiar compound words than late-acquired, highly imageable and familiar compound words. Participants showed higher sensitivity to early-acquired modifiers and head lexemes within compound words than late-acquired modifiers and head lexemes, whereas imageable and familiar head lexeme within the compound word showed higher sensitivity than less imageable and more unfamiliar head lexeme within the compound word. However, the imageability and familiarity of the modifier lexeme did not contribute significantly to recognition memory performance, while semantic transparency and lexeme meaning dominance did not contribute significantly to recognition memory performance.

Response bias

The fixed effects (see Tables 9 and 10) capture the grand means of *C* for the psycholinguistic effects in recognition memory without a space between lexemes. Participants were more likely to be very conservative in recognition memory without a space ($b = 0.808$, $t = 77.68$, $p < .001$). In the baseline model, participants showed a stronger bias to recognise infrequent compound words than frequent compound words, while there was a stronger bias to recognise frequent modifier lexemes within a compound word than infrequent modifier lexeme within a compound word. However, there was no effect of word length of a compound word or the frequency of the head lexeme on recognition performance. There was a stronger bias to recognise unfamiliar and low imageable compound words than familiar and high imageable compound words. There was a weaker bias for familiar and imageable head lexemes within compound words than unfamiliar and less imageable head lexemes within compound words. However, there was no effect of the AoA lexeme meaning dominance and semantic transparency of the compound word on response bias. There was no familiarity, imageability and AoA of the modifier lexemes within a compound word on response bias. There was no effect of AoA of the head lexeme within a compound word on response bias. These results will be discussed after Experiment 4.

Experiment 4: Recognition memory with a space

Several papers have examined AoA on recognition memory. Cortese et al. (2010, 2015) and Lau et al. (2018) found that AoA predicted recognition memory. Experiment 4 was designed to assess whether the AoA affects recognition memory in compound words with a space inserted between lexemes. Recognition memory, however is more likely to be affected by the space between lexemes than free recall, thus the structural properties of words would be exaggerated, allowing semantic processing to be more likely to be involved in recognition memory (Cortese & Schock, 2013). We predict that having a space between words would have a bigger impact in the recognition memory task

Table 8
Linear mixed effects regression results for the baseline model plus the additional 11 variables included for each separate model for Experiment 3 in *d*' (sensitivity).

Values	CFA[a]	CAoA[a]	CI[a]	ST[a]	LMD[a]	MFA[a]	MAoA[a]	MI[a]	HFA[a]	HAoA[a]	HI[a]
β	0.020	0.007	0.017	−0.006	−0.005	0.007	−0.009	−0.34	−0.016	−0.011	0.008
SE	0.004	0.003	0.004	0.003	0.003	0.004	0.004	0.002	0.004	0.004	0.002
2.5%CI	0.010	0.001	0.009	−0.012	−0.011	−0.001	−0.017	−0.344	−0.0238	−0.019	0.010
97.5% CI	0.028	0.013	0.025	−0.0001	0.0009	0.015	−0.001	−0.336	−0.002	−0.002	0.028
t-value	5.02*	2.14*	4.95*	−1.64	−1.31	1.85.	−2.17*	−1.44	−3.91*	−2.58*	3.39*
R^2m	0.006	0.006	0.006	0.005	0.005	0.006	0.005	0.005	0.006	0.006	0.006
R^2c	0.081	0.081	0.081	0.080	0.080	0.081	0.080	0.080	0.081	0.080	0.081

Note. Coefficients and standard error are presented for each variable when added separately to the baseline model containing compound word length, frequency, modifier frequency and head frequency. CI = Confidence interval; SE = Standard error; CFA = Compound word familiarity; CAoA = Compound word Age of acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance; MFA = Modifier familiarity, MAoA = Modifier age of acquisition, MI = Modifier imageability, HFA = Head familiarity, HAoA = Head age of acquisition and HI = Head imageability. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

Table 10
Linear mixed effects regression results for the baseline model plus the additional 11 variables included for each separate model for Experiment 3 in *c* (response bias).

Values	CFA[a]	CAoA[a]	CI[a]	ST[a]	LMD[a]	MFA[a]	MAoA[a]	MI[a]	HFA[a]	HAoA[a]	HI[a]
β	−0.014	−0.007	−0.012	0.001	0.009	−0.049	−0.003	0.029	−0.011	−0.003	−0.006
SE	0.003	0.003	0.003	0.003	0.003	0.003	0.004	0.002	0.004	0.004	0.002
2.5%CI	−0.020	−0.013	−0.018	−0.005	0.003	−0.055	−0.011	0.025	−0.0188	−0.011	−0.010
97.5% CI	−0.008	−0.001	−0.006	0.007	0.010	−0.043	0.004	0.033	−0.002	−0.002	0.001
t-value	−4.11*	−0.23	−3.80*	0.31	0.30	−0.14	0.91	1.41	−3.18*	0.94	−2.95*
R^2m	0.006	0.006	0.006	0.005	0.005	0.006	0.005	0.005	0.006	0.006	0.006
R^2c	0.081	0.081	0.081	0.080	0.080	0.081	0.080	0.080	0.081	0.080	0.081

Note. Coefficients and standard error are presented for each variable when added separately to the baseline model containing compound word length, frequency, modifier frequency and head frequency. SE = Standard error; CFA = Compound word familiarity; CAoA = Compound word Age of acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance; MFA = Modifier familiarity, MAoA = Modifier age of acquisition, MI = Modifier imageability, HFA = Head familiarity, HAoA = Head age of acquisition and HI = Head imageability. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

Table 11
Linear mixed effects regression results for the baseline model for Experiment 4 in d' (sensitivity).[a]

Values	CL	CFreq	MFreq	HFreq
β	0.012	0.001	0.006	0.009
SE	0.003	0.004	0.004	0.004
2.5%CI	0.010	−0.007	−0.002	0.001
97.5% CI	0.018	0.009	0.014	0.017
t-value	3.53*	0.18	1.62	2.50*
R^2m	0.007			
R^2c	0.068			

Note. Coefficients and standard error are presented for each variable in the baseline model: compound word length (CL), compound word frequency (CFreq), modifier frequency (MFreq) and Head frequency (HFreq). CI = Confidence interval and SE = Standard error. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

compared to the free recall task as the former may be more prone to low-level properties of words such as orthographic features than the latter (Hunt & Elliot, 1980). In addition, it has been shown that the addition of a space between lexemes is likely to affect lexical decision tasks such that spaced compounds take longer to recognise than unspaced compounds (e.g. Juhasz et al., 2005). Based on previous findings that showed creating a space between lexemes should encourage further resources within semantic processing, we would expect to observe larger semantic processing within recognition memory with a space between lexemes, thus larger AoA effects should be noted.

Method

Participant

One-hundred twenty four British monolingual undergraduate students, aged 18–22 years (M = 18.85 ± 0.86 years; 101 females). Participants were remunerated with course credits. The experiment was conducted in accordance with the British Psychological Society's ethical guidelines and approved by the University's ethical committee. All participants had normal or corrected-to-normal vision and signed a consent form to participate in the study. None of the students participated in the previous experiment. However, we excluded four participants, as they had a mean error rate above 60%.

Analysis, materials and procedures

The same analysis, materials and procedures were used as in Experiment 3 with the following exception: a space was inserted between the two lexemes of the compound word (e.g. air plane). The insertion of the space was included in the study and test phase. Participants were informed that they would be simultaneously presented with two lexical strings which they had to recognise whether they had studied the item. The mean difference in timings of these sessions was four to seven days (M = 5.20).

Results and discussion

Rt

Following the advice of Woods et al. (2023), across the samples, the rate of missingness was less than 1% and was missing at random. We analysed RTs for studied items only in recognition memory with and without a space. We removed incorrect responses and missed/late responses only for RT data. Outliers, including responses faster than 200 ms and greater than or below 2.5SD from the group mean were also removed, leading to the removal of 27.3% of responses for recognition memory RT (Brysbaert & Stevens, 2018; see supplementary materials for psycholinguistic predictors on reaction time data and data transformation). Spaced items (*M* = 1014, *SD* = 371; 95% CI[1009, 1019]) took longer to recognise as studied items than unspaced items (*M* = 879, *SD* = 372; 95% CI[874, 884]) (*b* = -0.16, *t* = -3.44, *p* <.001).

Sensitivity

We used signal detection metrics such as *d*' and *c* to provide more accurate and reliable measures of memory. In this analysis, we assess *d*' and *c*, each of those terms represent either a grand mean or condition effect on either response bias or sensitivity. The fixed effects (see Tables 11 and 12) capture d' as well as the psycholinguistic effects in recognition memory with a space between lexemes. Sensitivity was unsurprisingly high on average (*b* = -0.06, *t* = 17.87, *p* <.001). Participants were significantly more sensitive to compound words with a space between lexemes than those without a space between lexemes (*b* = 0.017, *t* = 3.55, *p* <.001).

In the baseline model for recognition with a space model, participants showed higher sensitivity to longer words than smaller words, whereas high-frequency head lexeme of a compound showed higher sensitivity than low-frequency head lexeme within a compound word. However, there was no significant effect of the frequency of the compound word nor its head lexeme on recognition memory performance. In addition, participants showed higher sensitivity to familiar, highly imageable and early-acquired compound words than unfamiliar, low imageable and late-acquired compound words. Furthermore, participants showed higher sensitivity in recognition memory following familiar, highly imageable and early-acquired head lexemes within compound word than unfamiliar, low imageable and late-acquired head lexemes. However, there was no effect of semantic transparency and lexeme meaning dominance of the compound word, together with the imageability, familiarity and the AoA of the modifier within the compound word did not contribute to recognition memory performance.

The estimate value for frequency, which has been argued to be a measure of lexical/semantic processes in compound word processing (e.g. Juhasz et al., 2015), was also numerically larger for the recognition memory without a space than the task with a space, while word length predictor was numerically larger for recognition memory with a space than without a space (Tables 7, 8, 11 and 12). This indicates that lexical-semantic effects are stronger for the recognition memory without a

Table 12
Linear mixed effects regression results for the baseline model plus the additional 11 variables included for each separate model for Experiment 4 in d' (sensitivity).

Values	CFA[a]	CAoA[a]	CI[a]	ST[a]	LMD[a]	MFA[a]	MAoA[a]	MI[a]	HFA[a]	HAoA[a]	HI[a]
β	0.010	−0.010	0.011	−0.002	−0.003	0.006	−0.007	0.004	0.008	−0.018	0.008
SE	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.004	0.004	0.003
2.5%CI	0.002	−0.018	0.003	−0.010	−0.011	−0.002	−0.015	0.0005	0.0002	−0.026	0.002
97.5% CI	0.018	−0.002	0.019	0.006	0.005	0.014	0.001	0.008	−0.002	−0.002	0.001
t-value	2.55*	−2.95*	3.10*	−0.47	−0.84	1.47	−1.52	1.58	2.14*	−4.26*	3.07*
R^2m	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
R^2c	0.068	0.068	0.069	0.068	0.068	0.068	0.068	0.068	0.068	0.069	0.069

Note: Coefficients and standard error are presented for each variable when added separately to the baseline model containing compound word length, frequency, modifier frequency and head frequency. CI = Confidence interval; SE = Standard error; CFA = Compound word familiarity; CAoA = Compound word Age of acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance; MFA = Modifier familiarity, MAoA = Modifier age of acquisition, MI = Modifier imageability, HFA = Head familiarity, HAoA = Head age of acquisition and HI = Head imageability. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

Table 13
Linear mixed effects regression results for the baseline model for Experiment 4 in c (response bias).[a]

Values	CL	CFreq	MFreq	HFreq
β	−0.001	0.001	−0.007	−0.015
SE	0.003	0.004	0.004	0.004
2.5%CI	−0.007	−0.007	−0.015	−0.023
97.5% CI	0.005	0.009	0.001	−0.007
t-value	−0.30	0.20	−1.94.	−4.07*
R^2m	0.007			
R^2c	0.068			

Note. Coefficients and standard error are presented for each variable in the baseline model: compound word length (CL), compound word frequency (CFreq), modifier frequency (MFreq) and Head frequency (HFreq). SE = Standard error. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

space than with a space between lexemes, while recognition memory with a space may tap into the structural processes. In terms of lexemic processing, the AoA, familiarity and imageability of the modifier was stronger in the recognition memory without a space, while the imageability of the modifier was stronger in the recognition memory with a space. This suggests that recognition memory without a space taps into the lexemic-semantic process, while recognition memory with a space may tap into the lexemic process level at lexemic level.

Response bias

The fixed effects (see Tables 13 and 14) capture the grand means of *C* for the psycholinguistic effects in recognition memory with a space between lexemes. Participants were more likely to be very conservative in recognition memory with a space ($b = 0.809$, $t = 85.67$, $p < .001$). There was no significant difference in response bias for spaced versus unspaced recognition memory ($b = -0.001$, $t = -0.08$, $p = .94$).

In the baseline model, participants showed a stronger bias to recognise infrequent head lexemes within compound words than frequent head lexemes within compound words. However, there was no effect of word length of a compound word or the frequency of the compound word and modifier lexeme within a compound on recognition performance. In addition, participants showed stronger bias to unfamiliar, low imageable and late-acquired compound words than familiar, high imageable and early-acquired compound words. Furthermore, participants showed weaker bias in recognition memory following familiar, highly imageable and early-acquired head lexemes within compound word than unfamiliar, low imageable and late-acquired head lexemes. However, there was no effect of semantic transparency and lexeme meaning dominance of the compound word, together with the imageability, familiarity and the AoA of the modifier within the compound word did not contribute to recognition memory performance.

In Experiments 3 and 4, we found an effect of AoA on recognition memory but in different directions. In Experiment 3, participants were more likely to recognise late-acquired compounds more accurately than early-acquired compound words, while in Experiment 4, participants were more likely to recognise early-acquired compound words more accurately than late-acquired compound words. The former replicates previous findings such that the reverse AoA effect is observed in recognition memory (e.g. Dewhurst et al., 1998; Macmillan et al., 2021). The results are also in line with the previous regression studies (e.g. Cortese et al., 2010, 2015; Lau et al., 2018). The findings of Experiment 4 do not replicate the previous results, as we observe an AoA effect such that early-acquired compound words are recognised more accurately than late-acquired words. One possible explanation is that recognition memory is sensitive to low-level processes such as perceptual, orthographic and phonological properties, which influences the familiarity processes underlying the performance. As a result of the space being included, thus structural properties of the word may be exaggerated, as demonstrated with spaced items taking longer to process than unspaced items. The space encourages the reader to recognise the compound word as two short words, which are more orthographically distinct (e.g. Cortese et al., 2020), leading to reverse word length effect such that long words are recognised more accurately than short words (e.g. Hendry &

Table 14
Linear mixed effects regression results for the baseline model plus the additional 11 variables included for each separate model for Experiment 4 in c (response bias).

Values	CFA[a]	CAoA[a]	CI[a]	ST[a]	LMD[a]	MFA[a]	MAoA[a]	MI[a]	HFA[a]	HAoA[a]	HI[a]
β	−0.016	0.010	−0.011	−0.001	0.003	−0.004	0.001	0.001	−0.009	0.010	−0.005
SE	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.004	0.004	0.003
2.5%CI	−0.024	0.002	−0.019	−0.009	−0.005	−0.012	−0.007	−0.003	−0.0168	0.002	−0.011
97.5% CI	−0.008	0.018	−0.003	0.007	0.011	0.004	0.009	0.005	−0.002	−0.002	0.001
t-value	−3.96*	2.79*	−2.99*	−0.36	0.75	−1.04	0.30	0.52	−2.14*	2.60*	−1.97*
R^2m	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
R^2c	0.068	0.068	0.069	0.068	0.068	0.068	0.068	0.068	0.068	0.069	0.069

Note: Coefficients and standard error are presented for each variable when added separately to the baseline model containing compound word length, frequency, modifier frequency and head frequency. SE = Standard error; CFA = Compound word familiarity; CAoA = Compound word Age of acquisition, CI = Compound word imageability, ST = semantic transparency, LMD = lexeme meaning dominance; MFA = Modifier familiarity, MAoA = Modifier age of acquisition, MI = Modifier imageability, HFA = Head familiarity, HAoA = Head age of acquisition and HI = Head imageability. *Significant at the α =.05 level. [a] This model did not converge with the variable of interest. We only included a random subject- and item intercept.

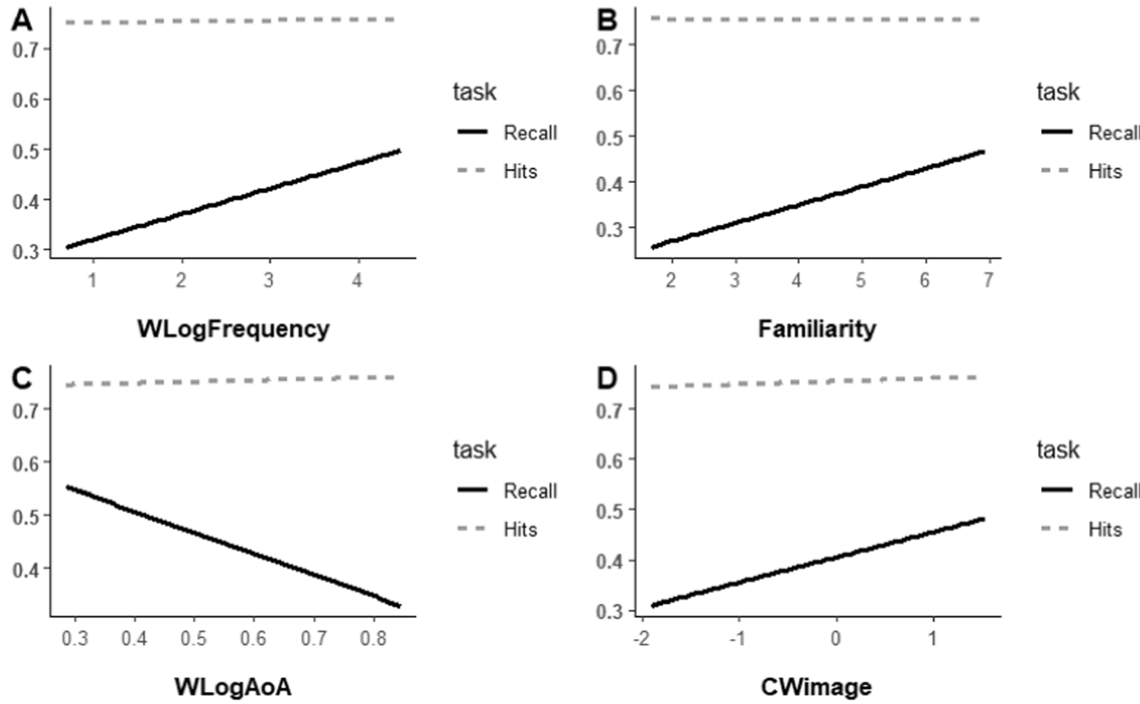

Fig. 1. The interaction patterns between (A) task and frequency, (B) task and familiarity, (C) task and AoA and (D) task and imageability on accuracy.

Tehan, 2005; Lau et al., 2018; Tehan & Tolan, 2007). As a result of this orthographic distinctiveness, late-acquired words are unlikely to benefit from the structure formed by early-acquired words (Ellis & Lambon Ralph, 2000), leading to early-acquired word advantage. Put simply, these findings align with the integrated account (Chang & Lee, 2020; Elsherif et al., 2023), as the space disrupts the low-level processes that allow late-acquired word to benefit from the structure formed by early-acquired words to be recognised (Ellis & Lambon Ralph, 2000). This indicates that the integrated account can explain the findings of the AoA effect in recognition memory. Based on these findings, we can conclude that the AoA contributes to recognition memory of compound words.

Additional data analysis across experiments: Interaction of the AoA of the compound word, task and space.

We assessed whether there was an interaction between frequency, familiarity, imageability or AoA and task. Here we investigated studied items only to make it easier to interpret the findings between task. Task (recognition and recall coded as −1 and 1, respectively) was coded as fixed effects. When contrast coding is explicitly noted, then there is no requirement to include post-hoc testing to determine the directionality of effects, so model coefficients reflect the effect magnitude independently of other predictors' levels (see Brehm & Alday, 2022 for a detailed discussion of contrast coding in LMM). The combined model was associated with $R^2m = 12.98\%$ and $R^2c = 22.13\%$ where task was a significant predictor ($b = 0.350$, SE $= 0.013$, $t = 27.01$), with more hits in recognition task than for free recall. Task by frequency, familiarity or AoA of the compound word was added into the combined model separately as a fixed factor. Adding task × frequency to the model resulted in a significant improvement ($\chi2(1) = 183.38$, $p < .001$), where $R^2m = 12.83\%$ and $R^2c = 22.11\%$ respectively. Adding task × familiarity to the model also resulted in a significant improvement ($\chi2(2) = 408.42$, $p < .001$), where $R^2m = 13.27\%$ and $R^2c = 22.39\%$ respectively. A similar result was obtained for AoA and task ($\chi2(2) = 447.81$, $p < .001$), where $R^2m = 13.29\%$ and $R^2c = 22.40\%$ respectively. In addition, the pattern was noted with imageability ($\chi2(2) = 317.62$, $p < .001$), where $R^2m = 13.23\%$ and $R^2c = 22.32\%$ respectively. The interaction patterns (Fig. 1) showed that all target effects were stronger for free recall than hits and in opposite directions. The more frequent, familiar and imageable the compound words, the more the items were recalled. However, the less frequent, familiar and imageable the compound words, the more hits were obtained. In respect to AoA, earlier-acquired compounds words were recalled more accurately than late-acquired compound words, whereas later-acquired compound words received fewer hits than earlier-acquired compound words.

Additional data analysis across experiments: Psycholinguistic and space interaction

We assessed whether there was an interaction between frequency, familiarity, imageability or AoA and Space in these tasks. Again here we investigated studied items only to make it easier to interpret the findings between task. Task (recognition and recall coded as −1 and 1, respectively) and Spaced (unspaced and spaced coded as −1 and 1, respectively) were coded as fixed effects. When contrast coding is explicitly noted, then there is no requirement to include post-hoc testing to determine the directionality of effects, so model coefficients reflect the effect magnitude independently of other predictors' levels (see Brehm & Alday, 2022 for a detailed discussion of contrast coding in LMM). The baseline model that was used included interaction of task by the range of psycholinguistic predictors. Adding space × frequency to the model did not result in a significant improvement ($\chi2(1) = 0.05$, $p = .82$), neither space × familiarity ($\chi2(1) = 0.23$, $p = .63$) nor space × imageability improvement ($\chi2(1) = 0.07$, $p = .79$). However, adding AoA × space resulted in a significant improvement of the model ($\chi2(1) = 7.91$, $p = .005$), where $R^2m = 13.30\%$ and $R^2c = 22.41\%$ respectively. The slope of the AoA effect on memory tasks is much larger in the spaced compounds than the unspaced compounds (Fig. 2). Nevertheless, the three-way interaction of space and task did not interact with any of these predictors, excluding imageability. Adding Imageability × space × task resulted in a significant improvement of the model ($\chi2(2) = 6.41$, $p = .04$), where $R^2m = 13.24\%$ and $R^2c = 22.32\%$ respectively. In recall,

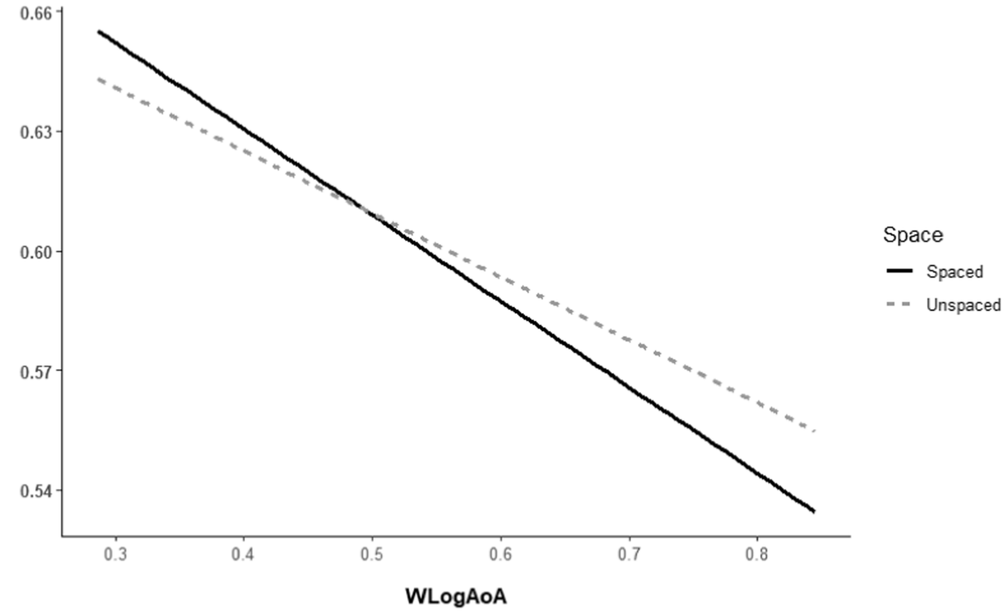

Fig. 2. The interaction patterns between space and AoA on accuracy.

the higher imageability score, the more likely an item would be recalled, irrespective of a space. The more imageable the item, the higher the hit rate (see Fig. 3).

General discussion

The current study is the first to assess the effects of AoA on compound words through old/new recognition and free recall. Across the four experiments, the results showed that the age at which a person learned a compound word significantly impacted recognition memory and recall performance. The present results support previous findings (e.g. Almond & Morrison, 2014; Cortese, Khanna, & Hacker, 2010, 2015) that the AoA of a compound word contributes to free recall and recognition memory. In addition, we observed that the age at which the modifier lexeme was acquired, together with their imageability and familiarity was more likely to affect free recall. One explanation is that the modifier lexeme narrows the semantic domain of the head lexeme, thus any predictors related to the modifier lexeme have been argued to be semantic in nature (Benczes, 2005; Elsherif & Catling, 2021). This indicates that within this task, the AoA effect is semantic in nature. Within Experiments 3 and 4, the predictors related to form and meaning (e.g. word length and the head lexemes) were found to contribute to recognition memory performance, together with semantic predictors of the compound word. This indicates that recognition memory is affected by the form and meaning processes, as opposed to primarily only semantic processes.

The effect of the AoA of the modifier was only observed in free recall, not recognition memory where the AoA of the modifier contributed to its accuracy, supporting Elsherif et al. (2020). Elsherif et al. argued that the presence of lexeme AoA occurs when semantic processing is involved. The AoA of the modifier indicates that semantic processing is involved in free recall, while the AoA effect of the head lexeme in recognition memory suggests that the mapping between low-level processes and

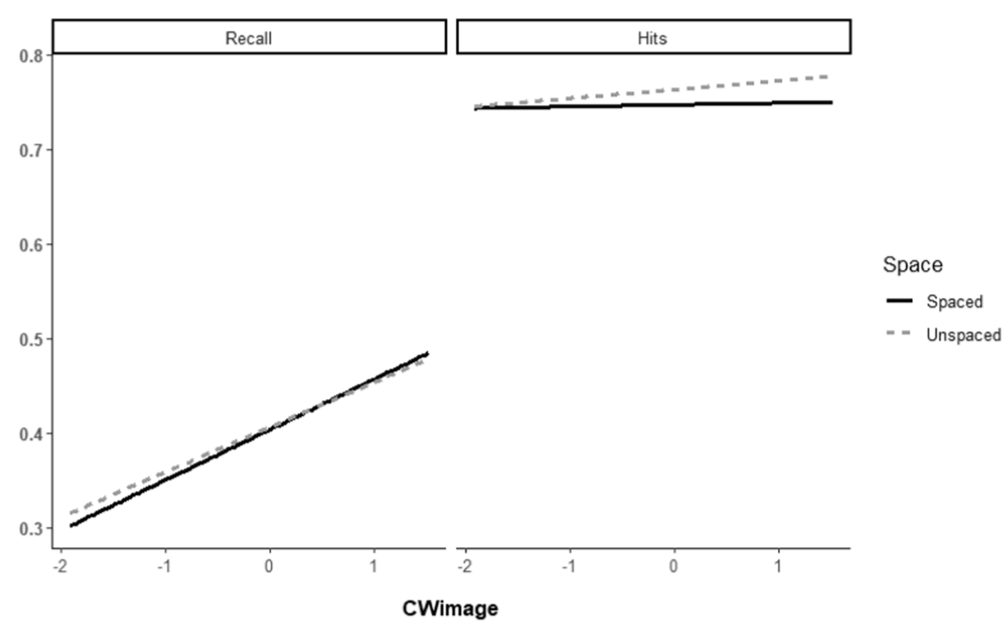

Fig. 3. The interaction between task, Space and Imageability on accuracy.

semantic processing is involved in this task. In addition, the AoA effect of the compound word and modifier decreases from free recall to recognition memory task. This suggests that the AoA effect for recognition memory could occur in the mapping between these representations. However, these findings also demonstrate that the AoA effect within memory cannot be determined solely by the mappings between representations but also by the formation of the semantic representation. Overall, the findings of AoA effects in compound words follow the prediction of the integrated view of AoA (Brysbaert & Ellis 2016; Chang et al., 2019; Chang & Lee, 2020; Elsherif & Catling, 2021; Dirix & Duyck, 2017; Elsherif et al., 2023; Menenti & Burani, 2007).

It is important to note that the direction of the AoA effect differs in recognition memory depending on whether a space is included, such that late-acquired words are recognised more accurately than early-acquired compound words in Experiment 3. One possible explanation for these results is the semantic distinctiveness effect. Cortese et al. (2010) argued that when learning a novel word, early-acquired compound words become the reference point or the hub for which the novel word is accessed. The reverse AoA effect should be observed, as early-acquired compound words have more representations and centres connected to it than late-acquired compound word, thus early-acquired words are not semantically distinct. However, as late-acquired compound words have fewer representations and connections, they become more semantically distinct and are recognised better than early-acquired words (cf. Gullick & Juhasz 2008). The same argument can be applied to low-frequency modifiers in that low-frequency modifiers have fewer representations and connections than their higher-frequency counterparts, thus they become more semantically distinct. However, the semantic distinctiveness account is agnostic to the free recall task. Nevertheless, based on the semantic similarity effect (i.e. items are better recalled from the same semantic categories; e.g. Crowder, 1979; Glanzer & Schwartz, 1971; see also Belke et al., 2005 who observed larger semantic interference in late-acquired, as opposed to early-acquired words), the semantic distinctiveness account would predict an AoA effect such that early-acquired words are recalled better than late-acquired words. Nonetheless, in the current study, we do observe effects of imageability and familiarity in recognition memory, thus the effects observed may be lexical-semantic in nature.

It is important to discuss the observed result that the AoA effect was found to be present when a space was included in compound words. In Experiment 4, we found that early-acquired compound words are recognised more accurately than late-acquired word. This poses a challenge for the semantic distinctiveness explanation for the AoA effect. However, it could be explained by the fact that within Experiment 4 we observed a reversed word length effect such that long words were more likely to be recognised than short words. There are two possible, and not mutually exclusive, explanations for this reversed length effect that could link to our AoA findings. One explanation is that long words derived greater benefits from lexical knowledge than short words (e.g. Jefferies et al., 2011), thus additional lexical knowledge is more likely to lead to better recognition, as the representation is degraded. The AoA effect being present may result from additional lexical knowledge beginning to contribute to recognition of specific items, thus early-acquired words can benefit from their connections in order to be recognised more easily than late-acquired words, which do not have as many semantic connections and are thus likely degrading even further. This is shown as we also observe effects of imageability and familiarity in the recognition memory with a space task. The second explanation is that the inclusion of a space within the compound word is more likely to lead two individual words being processed. As a result of these two words being perceived by the reader, they become more orthographically distinct (Cortese et al., 2020), leading to the reader to recognise the items more quickly.

In addition, we examined the interaction between the effect of lexical-semantic variables and insertion of a space and task type. We observed two noted findings: First, when comparing lexical-semantic properties, frequency, familiarity, imageability and AoA of the compound word was larger in recall tasks than recognition memory tasks. This is in line with previous studies (e.g. Ballot et al., 2021; Lau et al., 2018), who have demonstrated that the magnitude of imageability and frequency effects are smaller in recognition memory than free recall, indicating that recall tasks tap more into the lexical-semantic properties than recognition memory, supporting previous findings (e.g. Lau et al., 2018), and the AoA has a common origin with lexical-semantic predictors, as predicted by the representation theory (see review by Elsherif et al., 2023). Second, the spacing between compounds only interacted with one psycholinguistic variable, namely, the AoA of the compound word, such that spaces inserted between lexemes produced a larger slope for the AoA effect. This is further supported by the three-way interaction of imageability, space and task where placing a space between lexemes affected the magnitude of the impact of imageability on recognition tasks only. Furthermore, imageability is more likely to aid in recognising studied items but not when there is a space placed between lexemes. These findings indicate that processing space within memory is independent of lexical or semantic processes and occurs at a perceptual level[6]. One explanation could be that the space affects the stability of lexical representation. According to the mapping theory, compound words when entering the mental lexicon can benefit from the rich resources of the neural network, thus once the compound word is well-established, it can form connections with other compound words and the individual lexemes. However, the space disrupts the representation, making it less stable and more likely to incur a larger processing cost, thus leading to larger AoA effect.

Participants respond more slowly and showed larger *d*' in recognition memory when a space was inserted between lexemes than without a space inserted. This supports research from Juhasz et al. (2005), who used a lexical decision task and found that inserting a space between compound words led to slower reaction times than with no space, as processing is disrupted. Although these tasks are different, they both involve a decision process, not otherwise included in natural reading. However, this explanation contradicts Inhoff et al. (2000), who found that a space led to faster lexical decision latencies than compound words without a space. Both experiments used a decision-based process, thus one explanation could be that Inhoff et al. (2000) used a mixed list such that spaced and unspaced compounds were placed in the same list, whereas, within the current study, we kept spaced and unspaced compounds in separate lists. This difference between the present study and Inhoff et al. may result in participants using different strategies to recognise and discriminate items. Future research should investigate free recall and recognition memory of compound words (i.e. with and without a space) in relation to list composition.

Free recall and recognition memory with and without a space inserted between lexemes is theoretically important not only for AoA effects but also to assess lexemic properties of memory. The results of these tasks established that the influence of lexical-semantic variables differ between tasks, specifically that that these effects were greater in free recall than recognition memory. The directionality of these effects is shown to indicate that lexical-semantic access occurs later in the processing of the compound word. One explanation is that compound words are naturally long words, thus the relationship between spelling and sound are more arbitrary, allowing lexical-semantics processing to be more involved (Cortese et al., 2013). Second, free recall requires processes that entail the reconstruction of degraded word representations, and effortful search to locate lexical items. In addition, the degraded word is likely to be arbitrary and, in both cases, this may necessitate greater semantic processing (Lau et al., 2018). Recognition depends on

[6] With regard to lexeme-AoA properties, we note that there is a larger slope for the AoA effect in free recall than recognition memory. In addition, the modifier AoA is larger in spaced than unspaced compounds, indicating that lexemic processing was involved.

low-level processes (such as perceptual, orthographic and phonological processing) to recognise if the item is studied, and is similar to word naming, thus semantic processing is likely to take a smaller role. Nevertheless, in the current study, recognition and recall of compound words with a space only interacted with AoA, not the other lexical-semantic properties. The explanation could be that inserting a space between lexemes encourages access to lexemic and low-level properties of the word, and the difference between spaced and unspaced compounds is at the perceptual level to the participants. The interaction between task and AoA was very large, while the interaction between space/no space and AoA was small but statistically reliable. It could be that larger differences for the AoA and space/no space may be observed when using stimuli that are arbitrary in nature such as using pictures or tangrams[7], leading to larger AoA effects. A large-scale and systematic comparison of free recall and recognition memory, with and without a space inserted between lexemes, is important to test the integrated view of the AoA effect, perhaps using a creative destruction approach (i.e. pre-specifying alternative results by competing hypotheses on a complex set of experimental findings; Delios et al., 2022; Tierney et al., 2020,2021).

One limitation of the current paper was that the interaction between task and psycholinguistic effects could potentially be explained as a functional ceiling effect. That is, the initial level of performance in recognition memory was already very high (i.e. more than 75% of the required hit rates was observed for several participant, as such represents a functional ceiling), thus a recognition test may simply be not as sensitive as recall (i.e. recognition is an insensitive test) and may be over-estimating the variables importance. As the number of items presented at the study or test phase are likely to increase the chance of revealing a functional effect, it would best in future studies to present more than 19 study items and the 38 test items to ensure that the difficulty of the task can allow us to assess which psycholinguistic property contributes to the recognition memory task, and hence provide a more sensitive measure of recognition memory. Importantly, the recognition test memory was of sufficient sensitivity. The current study observed that our psycholinguistic effects, specifically AoA, was obtained using signal detection theory, that provides us a measure of sensitivity and response bias. Hits do not provide us with a reliable measure of memory. However, to make it easier to interpret the findings between tasks, it is necessary to use the hit measures, as theis measure was used in both tasks. This means that it is better to use sensitivity and response bias within recognition memory tasks but hit rates between memory tasks. Although this limitation is important and would be recommended to review this issues in future research, they are by no means critical to the central message of this research.

The four experiments from the current study highlight that the AoA effect does influence memory. Previous research has demonstrated that AoA affects recognition but does not affect serial or free recall in monomorphemic words (e.g. Macmillan et al., 2021). This pattern of results does not occur in compound word, as AoA contributes strongly to free recall with and without a space and recognition memory with and without a space. This highlights the generality of the AoA effect not only in lexical processing but also more broadly in the mechanisms relating to memory. In addition, these findings support the integrated account of the AoA effect, as we suggest the current study demonstrates that the AoA effect within memory occurs as a result of semantic distinctiveness and the mapping between representations. Put simply, the results from this study add to the growing evidence indicating that the AoA effect contributes to the gradual development of perceptual and semantic representations and the connections between these representations within a novel cognitive domain.

[7] Tangram pictures are abstract pictures formed of seven or more flat polygons described as tans to form shapes (e.g. Fasquel et al., 2022).

CRediT authorship contribution statement

Mahmoud M. Elsherif: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Conceptualization, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Jonathan C. Catling:** Conceptualization, Supervision, Methodology, Investigation, Data curation, Project administration, Resources, Software, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and materials for all experiments are available at https://osf.io/qu9hc/.

Acknowledgement

None. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2023.104449.

References

Almond, N. M., & Morrison, C. M. (2014). Episodic intertrial learning of younger and older participants: Effects of age of acquisition. *Aging, Neuropsychology, and Cognition, 21*(5), 606–632. https://doi.org/10.1080/13825585.2013.849653

Arsal, G., Suss, J., Ward, P., Ta, V., Ringer, R., & Eccles, D. W. (2021). The modified imitation game: A method for measuring interactional expertise. *Frontiers in Psychology, 12*, Article 730985. https://doi.org/10.3389/fpsyg.2021.730985

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Ballot, C., Mathey, S., & Robert, C. (2021). Word imageability and orthographic neighbourhood effects on memory: A study in free recall and recognition. *Memory, 29*(6), 829–834. https://doi.org/10.1080/09658211.2021.1921216

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*(4), 639–647. https://doi.org/10.3758/BF03200465

Barton, K., & Barton, M. K. (2015). *Package 'mumin'. Version, 1*, 18. <https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf>.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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*(2), B45–B54. https://doi.org/10.1016/j.cognition.2004.11.006

Benczes, R. (2005). Metaphor-and metonymy-based compounds in English: A cognitive linguistic approach. *Acta Linguistica Hungarica, 52*(2–3), 173–198. https://doi.org/10.1556/aling.52.2005.2-3.3

Benczes, R. (2014). Repetitions which are not repetitions: The non-redundant nature of tautological compounds. *English Language and Linguistics, 18*(3), 431–447. https://doi.org/10.1017/S1360674314000112

Benczes, R. (2015). Are exocentric compounds really exocentric? *SKASE Journal of Theoretical Linguistics, 12*(3), 54–73.

Bonin, P., Laroche, B., & Méot, A. (2022). Psycholinguistic norms for a set of 506 French compound words. *Behavior Research Methods, 54*(1), 393–413. https://doi.org/10.3758/s13428-021-01601-w

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language, 125*, Article 104334. https://doi.org/10.1016/j.jml.2022.104334

Brooks, T. L., & de Garcia, D. Cid (2015). Evidence for morphological composition in compound words using MEG. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00215

Brysbaert, M. (2017). Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings "corrected" for frequency. *Quarterly Journal of Experimental Psychology, 70*(7), 1129-1139. <>.

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology, 30*(11), 1240–1263. <https://doi.org/10.1080/02687038.2015.1106439>.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(2018), 1–20. https://doi.org/10.5334/joc.10

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica, 104*(2), 215–226. <https://doi.org/10.1016/S0001-6918(00)00021-4>.

Catling, J. C., & Elsherif, M. M. (2020). The hunt for the age of acquisition effect: It's in the links! *Acta Psychologica, 209*, 103138. <https://doi.org/10.1016/j.actpsy.2020.103138>.

Catling, J. C., Pymont, C., Johnston, R. A., Elsherif, M. M., Clark, R., & Kendall, E. (2021). Age of acquisition effects in recognition without identification tasks. *Memory, 1–13*. https://doi.org/10.1080/09658211.2021.1931695

Chang, Y.-N., & Lee, C.-Y. (2020). Age of acquisition effects on traditional Chinese character naming and lexical decision. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01787-8

Chang, Y.-N., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language, 108*, Article 104025. https://doi.org/10.1016/j.jml.2019.05.003

Coltheart, V., & Winograd, E. (1986). Word imagery but not age of acquisition affects episodic memory. *Memory & Cognition, 14*, 174–179. https://doi.org/10.3758/BF03198377

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384–387. https://doi.org/10.3758/BF03195585

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods, 40*(3), 791–794. https://doi.org/10.3758/BRM.40.3.791

Cortese, M. J., Khanna, M. M., & Hacker, S. D. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*, 595–609. https://doi.org/10.1080/09658211.2010.493892

Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology, 68*, 1489–1501. https://doi.org/10.1080/17470218.2014.945096

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology, 66*(5), 946–972. https://doi.org/10.1080/17470218.2012.722660

Cortese, M. J., Toppi, S., Khanna, M. M., & Santo, J. B. (2020). AoA effects in reading aloud and lexical decision: Locating the (semantic) locus in terms of the number of backward semantic associations. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820940302

Cortese, M. J., Von Nordheim, D., & Khanna, M. M. (2020). Word length negatively predicts recognition memory performance. *Quarterly Journal of Experimental Psychology, 73*(10), 1675–1683. https://doi.org/10.1177/1747021820921133

Cortese, M. J., Yates, M., Schock, J., & Vilks, L. (2018). Examining word processing via a megastudy of conditional reading aloud. *Quarterly Journal of Experimental Psychology, 71*(11), 2295–2313. https://doi.org/10.1177/1747021817741269

Crowder, R. G. (1979). Similarity and order in memory. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 13, pp. 319–353). New York: Academic Press.

Delios, A., Clemente, E. G., Wu, T., Tan, H., Wang, Y., Gordon, M., Viganola, D., Chen, Z., Dreber, A., Johannesson, M., Pfeiffer, T., Uhlmann, E. L., Abd Al-Aziz, A. M., Abraham, A. T., Trojan, J., Adamkovic, M., Agadullina, E., Ahn, J., Akinci, C., & The Generalizability Tests Forecasting Collaboration. (2022). Examining the generalizability of research findings from archival data. *Proceedings of the National Academy of Sciences of the United States of America, 119*(30), e2120377119. <https://doi.org/10.1073/pnas.2120377119>.

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. https://doi.org/10.1037/0278-7393.24.2.284

Dirix, N., & Duyck, W. (2017). An eye movement corpus study of the age-of-acquisition effect. *Psychonomic Bulletin & Review, 24*(6), 1915–1921. https://doi.org/10.3758/s13423-017-1233-8

Dirix, N., & Duyck, W. (2017). The first- and second-language age of acquisition effect in first- and second-language book reading. *Journal of Memory and Language, 97*, 103–120. https://doi.org/10.1016/j.jml.2017.07.012

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. https://doi.org/10.1037/0278-7393.26.5.1103

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. https://doi.org/10.1037/0278-7393.24.2.515

Elsherif, M. M., & Catling, J. C. (2021). Age of acquisition effects on the decomposition of compound words. *Journal of Cognitive Psychology, 1–14*. https://doi.org/10.1080/20445911.2021.2013246

Elsherif, M. M., Catling, J. C., & Frisson, S. (2020). Two words as one: A multi-naming investigation of the age-of-acquisition effect in compound-word processing. *Memory & Cognition, 48*(4), 511–525. https://doi.org/10.3758/s13421-019-00986-6

Elsherif, M. M., Preece, E., & Catling, J. C. (2023). Age-of-acquisition effects: A literature review. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance Online Publication.*. https://doi.org/10.1037/xlm0001215

Elsherif, M. M., Saban, M. I., & Rotshtein, P. (2017). The perceptual saliency of fearful eyes and smiles: A signal detection study. *PloS one, 12*(3), e0173199.

Fasquel, A., Brunellière, A., & Knutsen, D. (2022). A modified procedure for naming 332 pictures and collecting norms: Using tangram pictures in psycholinguistic studies. *Behavior Research Methods, 1–23*. https://doi.org/10.3758/s13428-022-01871-y

Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology, 99*(1), 87–107. https://doi.org/10.1348/000712607X181304

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of acquisition effects in lexical and episodic memory tasks. *Memory & Cognition, 7*, 214–223. https://doi.org/10.3758/BF03197541

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley.

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67*, 106–148. https://doi.org/10.1016/j.jml.2012.04.004

Glanzer, M., & Schwartz, A. (1971). Mnemonic structure in free recall: Differential effects on STS and LTS. *Journal of Verbal Learning & Verbal Behavior, 10*, 194–198. https://doi.org/10.1016/S0022-5371(71)80013-0

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General, 150*(4), 700–709. https://doi.org/10.1037/xge0000920

Gullick, M. M., & Juhasz, B. J. (2008). Age of acquisition's effect on memory for semantically associated word pairs. *Quarterly Journal of Experimental Psychology, 61*, 1177–1185. https://doi.org/10.1080/17470210802013391

Günther, F., Marelli, M., & Bölte, J. (2020). Semantic transparency effects in German compounds: A large dataset and multiple-task investigation. *Behavior Research Methods, 52*(3), 1208–1224. https://doi.org/10.3758/s13428-019-01311-4

Havelka, J., & Tomita, I. (2006). Age of acquisition in naming Japanese words. *Visual Cognition, 13*(7–8), 981–991. https://doi.org/10.1080/13506280544000156

Hendry, L., & Tehan, G. (2005). An item/order trade-off explanation of word length and generation effects. *Memory, 13*(3–4), 364–371. https://doi.org/10.1080/09658210344000341

Henry, R., & Kuperman, V. (2013). Semantic growth of morphological families in English. *Psihologija, 46*(4), 479–495. <https://doi.org/10.2298/PSI1304479H>.

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General, 109*(1), 49–74. https://doi.org/10.1037/0096-3445.109.1.49

Inhoff, A. W., Radach, R., & Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language, 42*(1), 23–50. https://doi.org/10.1006/jmla.1999.2666

Jefferies, E., Frankish, C., & Noble, K. (2011). Strong and long: Effects of word length on phonological binding in verbal short-term memory. *Quarterly Journal of Experimental Psychology, 64*(2), 241–260. https://doi.org/10.1080/17470218.2010.495159

Juhasz, B. J. (2018). Experience with compound words influences their processing: An eye movement investigation with English compound words. *Quarterly Journal of Experimental Psychology, 71*(1), 103–112. https://doi.org/10.1080/17470218.2016.1253756

Juhasz, B. J., Inhoff, A. W., & Rayner, K. (2005). The role of interword spaces in the processing of English compound words. *Language and cognitive processes, 20*(1–2), 291–316. https://doi.org/10.1080/01690960444000133

Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods, 47*(4), 1004–1019. https://doi.org/10.3758/s13428-014-0523-6

Kilecioğlu, E., Raman, I., & Raman, E. (2020). The influence of age of acquisition on recall and recognition in Alzheimer's patients and healthy ageing controls in Turkish. *Applied Neuropsychology: Adult, 29*(4), 1–10. https://doi.org/10.1080/23279095.2020.179666

Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology, 4*(203), 1–10. https://doi.org/10.3389/fpsyg.2013.00203

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. https://doi.org/10.1080/13506280544000110

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology, 71*(10), 2207–2222. https://doi.org/10.1177/1747021817739834

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language, 84*, 50–64. https://doi.org/10.1016/S0093-934X(02)00520-5

Macmillan, M. B., Neath, I., & Surprenant, A. M. (2021). Re-assessing age of acquisition effects in recognition, free recall, and serial recall. *Memory & Cognition, 49*, 939–954. https://doi.org/10.3758/s13421-021-01137-6

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Erlbaum.

Marful, A., Gómez-Ariza, C. J., Barbón, A., & Bajo, T. (2016). Forgetting "Novel" but Not "Dragon": The role of age of acquisition on intentional and incidental forgetting. *PLOS ONE, 11*(5), e0155110. <https://doi.org/10.1371/journal.pone.0155110>.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*(4), 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

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. https://doi.org/10.1080/17470210601100126

Momenian, M. (2022). Do age of acquisition and orthographic transparency have the same effects in different modalities?. In *Proceedings of the workshop on cognitive aspects of the lexicon* (pp. 25–30).

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. https://doi.org/10.1016/j.jml.2010.08.003

Morris, P. E. (1981). Age of acquisition, imagery, recall, and the limitations of multiple-regression analysis. *Memory & Cognition, 9*, 277–282. https://doi.org/10.3758/BF03196961

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Plag, I. (2018). *Word-formation in English*. Cambridge University Press.

Ploetz, D. M., & Yates, M. (2016). Age of acquisition and imageability: A cross-task comparison. *Journal of Research in Reading, 39*(1), 37-49. <https://doi.org/10.1111/1467-9817.12040>.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology Section A, 48*(2), 384–404. https://doi.org/10.1080/14640749508401396

Poon, P. (2020). *Towards a game-based perceptual training tool for improving motorcycle detection.* [Doctoral dissertation, Nottingham Trent University]. <https://www.proquest.com/docview/2572316600?pq-origsite=gscholar&fromopenview=true>.

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Rabe, M. M. (2018). *Generalized linear mixed modeling of signal detection theory*. [Doctoral thesis, University of Victoria]. <http://dspace.library.uvic.ca/handle/1828/9208>.

Raman, I., Raman, E., İkier, S., Kilecioğlu, E., Uzun Eroğlu, D., & Zeyveli, Ş. (2018). Differential effects of age of acquisition and frequency on memory: Evidence from free recall of pictures and words in Turkish. *Writing Systems Research, 10*, 1–14. https://doi.org/10.1080/17586801.2017.1420727

Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., & Brown, G. D. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22*(6), 695–701. https://doi.org/10.3758/BF03209254

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior, 19*(6), 736–755. https://doi.org/10.1016/S0022-5371(80)90415-6

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user 's guide*. Pittsburgh, PA: Psychology Software Tools.

Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(4), 971–977. https://doi.org/10.3758/s13428-012-0209-x

Schock, J., Cortese, M. J., & Khanna, M. M. (2012). Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods, 44*(2), 374–379. https://doi.org/10.3758/s13428-011-0162-0

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704

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. https://doi.org/10.1207/s15516709cog2901_3

Tehan, G., & Tolan, G. A. (2007). Word length effects in long-term memory. *Journal of Memory and Language, 56*(1), 35–48. https://doi.org/10.1016/j.jml.2006.08.015

Tierney, W., Hardy, J. H., Ebersole, C. R., Leavitt, K., Viganola, D., Clemente, E. G., Gordon, M., Dreber, A., Johannesson, M., Pfeiffer, T., & Uhlmann, E. L. (2020). Creative destruction in science. *Organizational Behavior and Human Decision Processes, 161*, 291–309. https://doi.org/10.1016/j.obhdp.2020.07.002

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language, 92*, 265–280. <https://doi.org/10.1016/j.jml.2016.07.004>.

Tierney, W., Hardy, J., Ebersole, C. R., Viganola, D., Clemente, E. G., Gordon, M., Hoogeveen, S., Haaf, J., Dreber, A., Johannesson, M., Pfeiffer, T., Huang, J. L., Vaughn, L. A., DeMarree, K., Igou, E. R., Chapman, H., Gantman, A., Vanaman, M., Wylie, J.,…Uhlmann, E. L. (2021). A creative destruction approach to replication: Implicit work and sex morality across cultures. *Journal of Experimental Social Psychology, 93*, 1–18. <https://doi.org/10.1016/j.jesp.2020.104060>.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for british english. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

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

Volkovyskaya, E., Raman, I., & Baluch, B. (2017). Age of acquisition (AoA) effect in monolingual Russian and bilingual Russian (L1)-English (L2) speakers in a free recall task. *Writing Systems Research, 9*, 148–163. https://doi.org/10.1080/17586801.2017.1405136

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Williams, E. (1981). Argument structure and morphology. *The Linguistic Review, 1*(1). https://doi.org/10.1515/tlir.1981.1.1.81

Wilson, D. M., Potter, K. W., & Cowell, R. A. (2018). Recognition memory shielded from semantic but not perceptual interference in normal aging. *Neuropsychologia, 119*, 448–463. https://doi.org/10.1016/j.neuropsychologia.2018.07.031

Woods, A. D., Gerasimova, D., Van Dusen, B., Nissen, J., Bainter, S., Uzdavines, A., Davis-Kean, P. E., Halvorson, M., King, K. M., Logan, J. A. R., Xu, M., Vasilev, M. R., Clay, J. M., Moreau, D., Joyal-Desmarais, K., Cruz, R. A., Brown, D. M. Y., Schmidt, K., & Elsherif, M. M. (2023). Best practices for addressing missing data through multiple imputation. *Infant and Child Development, e2407*. https://doi.org/10.1002/icd.2407

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science, 5*(2), 25152459221095823. <https://doi.org/10.1177/25152459221095823>.

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. https://doi.org/10.1006/jmla.2001.2834

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition, 32*(1), 31–38. https://doi.org/10.3758/BF03195818