Cognitive Mechanisms of the Age of Acquisition Effect: A Three-Level Meta-Analysis

Authors: Zhang Lu, Wang Ziqian, ZHANG Qingfang, Zhang Qingfang

Date: 2025-03-11T00:00:00+00:00

Abstract

Age of Acquisition (AoA) refers to the age at which an individual first understands the meaning of a word. Research has found that words acquired earlier are processed faster and with higher accuracy than words acquired later, demonstrating the AoA effect for vocabulary. Regarding the cognitive mechanisms underlying the AoA effect, researchers have proposed the multi-layer theory, the representational theory, and the arbitrary mapping theory. To systematically investigate the cognitive mechanisms of the AoA effect, this study employed a three-level meta-analysis to quantitatively analyze existing literature examining the AoA effect. The study retrieved and screened publications before May 15, 2024, ultimately including 130 studies (271 effect sizes, N = 10,697). The meta-analytic results revealed that the effect size of the AoA effect was Hedge' s g = 0.47 ($p < 0.001$, 95% CI = [0.41, 0.52]), with its magnitude moderated by task type and word frequency, and primarily dependent on the semantic relatedness of different tasks, but not influenced by target cognitive processes, writing systems, or task phonological relatedness. The findings indicate that AoA primarily affects the semantic processing stage, supporting the representational theory. Words acquired earlier possess richer semantic networks, thereby facilitating the lexical processing.

Full Text

Cognitive Mechanisms Underlying Age of Acquisition Effects: A Three-Level Meta-Analysis

ZHANG Lu†, WANG Ziqian†, ZHANG Qingfang
(Department of Psychology, Renmin University of China, Beijing 100872, China)

Abstract: Age of acquisition (AoA) refers to the age at which an individual first learns and understands the meaning of a word. Research has consistently shown that early-acquired words are processed faster and more accurately than later-acquired words, a phenomenon known as the AoA effect. To explain the cognitive mechanisms underlying this effect, researchers have proposed three main theoretical accounts: the Multiple Loci Account, Representation Theory, and Arbitrary Mapping Hypothesis. To systematically investigate the cognitive mechanisms of the AoA effect, the present study employed a three-level meta-analysis to quantitatively synthesize existing behavioral research on AoA effects. We retrieved and screened studies published before May 15, 2024, ultimately including 130 studies with 271 effect sizes and a total sample of 10,697 participants. The meta-analysis revealed a significant AoA effect with an effect size of Hedge's $g = 0.47$ ($p < 0.001$, 95% CI = [0.41, 0.52]). The magnitude of this effect was moderated by task type and word frequency, and depended primarily on the semantic relevance of different tasks. However, the effect was not significantly influenced by the target cognitive process, writing system, or phonological relevance of tasks. These findings indicate that AoA primarily affects the semantic processing stage, supporting Representation Theory. Early-acquired words appear to possess richer semantic networks, which facilitates lexical processing.

Keywords: picture naming, lexical decision, word naming, age of acquisition effects, three-level meta-analysis

Classification Code: B842

Received: October 7, 2024

This research was supported by the National Natural Science Foundation of China (32171055) and the National Social Science Fund Major Project (24&ZD251).

† ZHANG Lu and WANG Ziqian contributed equally to this work.

Corresponding Author: ZHANG Qingfang, E-mail: qingfang.zhang@ruc.edu.cn

Age of acquisition (AoA) refers to the age at which individuals first encounter and comprehend a word's meaning during language acquisition and development. Researchers have found that early-acquired words are processed faster and more accurately than later-acquired words in both lexical comprehension (Rochford & Williams, 1962a, 1962b) and lexical production (Carroll & White, 1973a), a phenomenon termed the AoA effect. This effect demonstrates cross-linguistic universality, having been observed in English, Spanish, Dutch, Chinese, and other languages (Barry et al., 2006; Cuetos et al., 1999; Ellis & Morrison, 1998; Chen et al., 2004). Previous research has debated the cognitive mechanisms underlying the AoA effect. Therefore, we employed a three-level meta-analysis to quantitatively synthesize existing behavioral studies examining AoA effects, incorporating moderator variables to explore the underlying cognitive mechanisms.

1.1 Cognitive Mechanisms of the AoA Effect

Researchers have primarily debated the specific stage(s) of lexical processing where the AoA effect occurs. The main theoretical perspectives include the

Multiple Loci Account, Representation Theory, and Arbitrary Mapping Hypothesis. More recently, researchers have proposed an Integrated View of AoA based on Representation Theory and Arbitrary Mapping Theory. Tasks examining AoA effects can be divided into two categories based on stimulus materials: word-based tasks, including lexical decision (Johnston & Barry, 2006), word naming (Barry et al., 2006), and semantic categorization (Chen et al., 2007); and picture-based tasks, including picture semantic classification (Catling & Johnston, 2006a), picture perception identification (Holmes & Ellis, 2006), object picture naming (Catling et al., 2008; Laganaro & Perret, 2011; Perret et al., 2014), and action picture naming (Bonin, Boyer, et al., 2004; Shao et al., 2014). The processing pathways for words and pictures differ fundamentally. Word processing involves accessing semantics through orthographic forms, whereas picture processing begins with conceptual activation. In lexical production tasks, this includes conceptual preparation, lexical selection (choosing appropriate words based on conceptual activation, including syntactic and semantic selection), phonological encoding (retrieving segmental and prosodic information), phonetic encoding (programming articulatory movements), and articulation (Levelt et al., 1999).

The Multiple Loci Account posits that AoA effects originate from different levels or stages of orthographic, phonological, and semantic processing. Specifically, AoA does not exclusively affect a single stage such as phonological retrieval in word or picture processing, but rather influences multiple processing stages (Moore et al., 2004). This theory builds upon the earlier Phonological Completeness Hypothesis, which proposed that AoA effects occur only during phonological retrieval in lexical processing. According to this hypothesis, early-acquired words are stored as holistic phonological units, while late-acquired words are stored in a decomposed form including individual segments and prosodic information, resulting in faster retrieval for early-acquired words (Brown & Watson, 1987). Researchers found AoA effects in lexical decision tasks manipulating phonologically related variables, but these effects disappeared when articulatory suppression prevented phonological retrieval (Gerhand & Barry, 1999b). However, other studies found that AoA effects do not depend on holistic or partial phonological retrieval (Zhang et al., 2011), challenging the Phonological Completeness Hypothesis. Subsequent research observed AoA effects in picture reality judgment tasks reflecting perceptual processing stages (Moore et al., 2004), leading Moore and colleagues to propose the Multiple Loci Account. When articulatory suppression was applied during picture reality judgment tasks, a smaller but still significant AoA effect remained (Holmes & Ellis, 2006), providing support for the Multiple Loci Account. Catling and Johnston (2006a) further proposed the Accumulation Hypothesis based on the Multiple Loci Account, suggesting that AoA affects multiple stages of lexical processing, including perceptual, conceptual, semantic, and phonological processing. Research has shown that AoA effects increase as more processing stages are included in experimental tasks. By comparing picture-picture judgment, semantic categorization (object reality judgment), word-picture judgment, and picture naming tasks, researchers found the largest AoA effect in picture naming tasks (Catling & Johnston, 2009), as this task includes all processing stages where AoA effects might occur. Using high temporal resolution EEG measures, researchers have also found that AoA affects multiple processing stages in picture naming, including semantic and phonological encoding (Laganaro & Perret, 2011; Perret et al., 2014; Valente et al., 2014), providing support for both the Multiple Loci Account and Accumulation Hypothesis.

Representation Theory was developed based on the relationship and distinction between AoA and word frequency effects. This theory proposes that the AoA effect comprises two components: one closely related to word frequency and another independent of word frequency (Brysbaert & Ghyselinck, 2006). Studies have found similar magnitudes of word frequency and AoA effects in lexical decision, word naming, and semantic categorization tasks. The frequency-dependent component of the AoA effect may operate at perceptual or phonological word-form processing levels, while the frequency-independent component originates from competition within conceptual and/or semantic representation networks. When a specific concept (e.g., fruit) corresponds to multiple words (apple, durian), early-acquired words serve as typical exemplars of the conceptual category (e.g., apple) and are easier to process compared to late-acquired words (Brysbaert & Ghyselinck, 2006). Additionally, Representation Theory suggests that early-acquired words have stronger lateral inhibition, which reduces the activation levels of competing items, requiring late-acquired words to overcome this inhibition to become activated (Belke et al., 2005). These processes depend on feature differences between early- and late-acquired words in semantic networks, with early-acquired words having richer semantic representations (Steyvers & Tenenbaum, 2005). During lexical retrieval, early-acquired words have higher node connectivity in semantic networks, making these networks more accessible. Researchers have observed typical AoA effects in both semantic association tasks (where participants report the first word that comes to mind when presented with a target word) and semantic categorization tasks (where participants classify words based on semantic information), indicating that AoA effects occur during semantic processing (Brysbaert et al., 2000). Significant AoA effects have also been found in various judgment tasks involving semantic processing, such as animacy judgment (Morrison & Gibbons, 2006) and man-made object judgment tasks (Catling & Johnston, 2006c). AoA effects have been observed even in tasks without phonological retrieval processes (e.g., face recognition, semantic categorization), leading researchers to propose Representation Theory, which suggests that AoA effects arise from organizational principles of lexical semantic representations in the semantic system. The semantic system shaped by early-acquired words forms the foundation for establishing meanings of later-acquired words (Brysbaert et al., 2000; Holmes & Ellis, 2006). Neuroimaging studies have also shown stronger activation for early-acquired words in the precuneus, a brain region associated with semantic memory, supporting the view that early-acquired words have richer semantic connection networks (Fiebach et al., 2003).

The Arbitrary Mapping Hypothesis was influenced by parallel distributed processing perspectives (Seidenberg & McClelland, 1989). According to this hypothesis, AoA effects originate from characteristics of the learning system and are primarily related to connection patterns between representational levels (perceptual, orthographic, semantic, phonological, etc.) rather than being located in a separate system (such as the semantic system) (Anderson & Cottrell, 2001; Ellis & Lambon Ralph, 2000; Monaghan & Ellis, 2010). From a connectionist network perspective, researchers have demonstrated that when individuals incorporate newly learned words into their mental lexicon during vocabulary acquisition, early-acquired words develop neural networks with higher plasticity, facilitating the formation and modification of input and output representations. As the mental lexicon develops, network plasticity decreases. In other words, early-acquired words have a greater influence on lexical connection networks (Ellis & Lambon Ralph, 2000). Reduced plasticity makes it more difficult for late-acquired words to consolidate in the mental lexicon, preventing them from altering previously established network structures and making it harder to form connections between levels (Yum & Law, 2019).

On the other hand, if late-acquired words have more regular morphological-phonological mappings, they can utilize mapping knowledge between orthography and phonology from similar words during lexical processing, which can partially compensate for the processing disadvantage of late-acquired words (Lambon Ralph & Ehsan, 2006; Zevin & Seidenberg, 2002). When mapping relationships are irregular, AoA effects emerge between early- and late-acquired words; when mapping relationships are regular, AoA effects are smaller, attributable to cumulative word frequency effects (Zevin & Seidenberg, 2004). Researchers examining the relationship between AoA and orthographic-phonological mapping consistency in word naming tasks found an interaction between the two factors. Compared to consistent mapping conditions, AoA effects were larger under inconsistent mapping conditions, suggesting that late-acquired words have difficulty utilizing existing network mapping relationships and require greater effort to complete processing (Monaghan & Ellis, 2002a). Similar conclusions were obtained by manipulating AoA and orthographic-semantic mapping relationships in Chinese characters (Chen et al., 2007). ERP studies using visual lexical decision tasks found that when orthographic-phonological mapping was inconsistent, late-acquired words elicited larger LPC amplitudes than early-acquired words, affecting late stages of lexical processing and supporting the mapping hypothesis (Bakhtiar et al., 2016).

Representation Theory and the Multiple Loci Account emphasize AoA effects on lexical processing stages, while the Arbitrary Mapping Hypothesis emphasizes AoA effects on mapping degrees between different representational levels. These three theories are not mutually exclusive (Elsherif & Catling, 2022). Recently, based on existing findings and theoretical perspectives, researchers have proposed an Integrated View of AoA (Brysbaert & Ellis, 2016; Chang & Lee, 2020; Cortese et al., 2020), suggesting that differences in representations between early- and late-acquired words, combined with differences in connection strength between representations, jointly determine lexical processing efficiency. AoA affects both the formation of representations in neural networks and the connection strength between representations (Elsherif & Catling, 2022). Early-acquired words occupy central positions in neural networks with rich semantic representations and conceptual connections, making them more suitable for input and output. Late-acquired words have weaker connections with other words and, due to reduced neural plasticity, are more difficult to modify in established neural networks, requiring greater effort to adapt to their specific input-output relationships. Using word naming and lexical decision tasks, researchers have observed not only AoA effects but also interactions between AoA and orthographic-phonological consistency, indicating that both lexical representations and mapping relationships between representations influence lexical processing, supporting the Integrated View (Chang et al., 2019; Chang & Lee, 2020).

1.2 Factors Influencing AoA Effects

Researchers have observed AoA effects across different languages using various linguistic tasks, with important influencing factors including task type, writing system, and word frequency (Elsherif et al., 2023).

1.2.1 Task Type

Lexical processing can be distinguished into two target cognitive processes: lexical comprehension and lexical production. Lexical comprehension focuses on perceptual processing of input and accessing orthographic, phonological, and semantic information, while lexical production emphasizes selecting appropriate words in the mental lexicon and performing phonological encoding and output. Tasks probing lexical comprehension primarily include word-pseudoword discrimination and semantic categorization, while tasks probing lexical production mainly include picture naming and word naming. In lexical decision tasks, participants discriminate whether presented words are real words based on lexical categories (González-Nosti et al., 2014; Izura & Hernández-Muñoz, 2017; Xu et al., 2021), with filler materials being pseudowords that conform to orthographic and phonological rules but are meaningless (though some studies used non-conforming pseudowords; Gerhand & Barry, 1999b; Ghyselinck, Lewis, Brysbaert, 2004). The primary processing goal involves extracting word meaning, including retrieval of orthographic, semantic, and/or phonological information. In semantic categorization tasks, participants judge whether presented words belong to a specified semantic category (e.g., man-made objects) (Catling & Johnston, 2006c), involving extraction of orthographic and semantic information. Picture naming tasks typically require participants to orally name presented pictures, representing a classic lexical production task that includes lexical-semantic encoding and phonological encoding stages. Word naming tasks require participants to read aloud presented words, primarily involving phonological encoding, phonetic encoding, and articulation processes, with minimal semantic access. Researchers have also used word completion tasks (Gilhooly & Gilhooly, 1979), degraded word recognition tasks (Chen et al., 2010), and phoneme segmentation tasks (Monaghan & Ellis, 2002a) to explore the cognitive mechanisms of AoA effects.

All three theoretical accounts agree that AoA effects exist in both lexical comprehension and production processes, though the magnitude differs due to varying processing demands. According to the Multiple Loci Account, picture naming includes all processing stages from perception to articulation, producing larger AoA effects than lexical decision, semantic categorization, and word naming tasks. The latter two tasks involve additional semantic processing and should therefore produce larger AoA effects than word naming (Moore et al., 2004; Catling & Johnston, 2009). Representation Theory posits that tasks involving semantic processing will show AoA effects, so picture naming, lexical decision, and semantic categorization should all exhibit AoA effects, while word naming, with minimal semantic involvement, should show smaller effects similar to word frequency effects (Brysbaert & Ghyselinck, 2006; Kuperman, 2013). According to the mapping hypothesis, picture naming has the least consistent mapping between semantic and phonological representations across all tasks, producing the largest AoA effect. Compared to word naming, semantic categorization and lexical decision have more inconsistent orthographic-semantic mappings, yielding larger AoA effects. Compared to semantic categorization, the presence of pseudowords in lexical decision tasks emphasizes the orthographic-semantic mapping relationship, resulting in larger AoA effects (Lambon Ralph & Ehsan, 2006; Catling & Elsherif, 2020).

Both lexical comprehension and production may involve semantic and phonological processing. Based on Representation Theory, tasks can be classified according to their semantic or phonological involvement to compare AoA effects. Studies have found significant AoA effects in picture naming but not in word naming, suggesting that task semantic relevance influences AoA effect magnitude (Lambon Ralph & Ehsan, 2006). In contrast, other research found smaller AoA effects in semantic categorization than picture naming, possibly due to differential phonological involvement (Catling & Johnston, 2006c). Tasks with high semantic or phonological relevance produce larger AoA effects compared to tasks with low relevance.

1.2.2 Writing System

Cross-linguistic comparisons reveal consistent AoA effects across languages, with researchers observing the effect in English (Carroll & White, 1973a), Spanish (Izura & Hernández-Muñoz, 2017), German (Kauschke & Von Frankenberg, 2008), Dutch (Ghyselinck, Custers, Brysbaert, 2004), Chinese (Bai et al., 2010; Chen et al., 2007; Zhang et al., 2022), and other languages. Writing systems can be classified into two types: alphabetic systems used in most Indo-European languages (e.g., English, German, French, Spanish, Dutch) and logographic systems such as Chinese. Alphabetic characters directly correspond to sounds, with words marking pronunciation according to spelling rules, resulting in relatively regular orthographic-phonological mapping (Grainger & Holcomb, 2009) but no direct connection to semantics. In contrast, logographic characters typically relate to meaning. Chinese characters are primarily phono-semantic compounds, where semantic radicals reflect semantic categories and phonetic radicals aid semantic extraction (Zhang et al., 2014), making orthographic-semantic mapping more regular in Chinese (Chang & Lee, 2020). Although phonetic radicals in Chinese can partially convey pronunciation, they are not fixed (Chen et al., 2007), and logographic systems remain primarily meaning-based. Therefore, alphabetic systems have more direct orthographic-phonological mapping but potentially more arbitrary orthographic-semantic and phonological-semantic mappings, while logographic systems have more direct orthographic-semantic mapping but potentially more arbitrary orthographic-phonological and phonological-semantic mappings (Weekes, 2011; Yum & Law, 2019; Bai et al., 2007). Comparing two Japanese writing systems with different mapping relationships (logographic Kanji and alphabetic Kana), researchers found larger AoA effects for Kanji than Kana in word naming tasks, suggesting that writing system moderates AoA effect magnitude and supporting the Arbitrary Mapping Hypothesis (Havelka & Tomita, 2006). Zevin and Seidenberg (2002) also argued that the more regular orthographic-phonological mapping in alphabetic systems is insufficient to produce reliable AoA effects. In contrast, Chinese has more irregular orthographic-phonological mapping, and numerous studies have found reliable AoA effects in Chinese word naming tasks (Chen et al., 2004; Chen & Zhu, 2015).

1.2.3 Word Frequency

Word frequency describes how often words appear in a corpus and is one of the earliest linguistic variables studied, influencing lexical recognition and processing speed (Brown & Watson, 1987; Levitt & Healy, 1985). The Cumulative Frequency Hypothesis posits that AoA effects primarily stem from differences in word usage frequency, with early-acquired words showing shorter reaction times in language tasks due to higher cumulative frequency throughout the lifespan (Gerhand & Barry, 1999a). Studies have observed interactions between word frequency and AoA in picture naming (Lambon Ralph & Ehsan, 2006), word naming (Gerhand & Barry, 1998), and lexical decision tasks (Wilson et al., 2013), with AoA effects being smaller for high-frequency words and larger for low-frequency words. From a connectionist perspective, Ellis and Lambon Ralph (2000) simulated lexical acquisition processes through computational modeling, manipulating the order of input-output pattern pairs to simulate AoA effects and training pre-input patterns at predetermined frequencies to examine word frequency' s influence. Results showed that output accuracy was affected by input order, with this effect differing between high- and low-frequency words. Representation Theory suggests that AoA effects partially originate from word frequency effects, with both system state and training frequency influencing network learning outcomes during semantic network construction (Brysbaert &

Ghyselinck, 2006).

In summary, most research on AoA effects has been behavioral, with only a few studies using EEG to examine the temporal course (Bakhtiar et al., 2016; Lou et al., 2019) or fMRI to investigate neural mechanisms (Weekes et al., 2004). Regarding behavioral findings, theoretical perspectives have focused on different aspects of the cognitive mechanisms underlying AoA effects, with early theories primarily explaining the effects from representational or inter-representational mapping perspectives, while recent Integrated Views consider both aspects. Existing research suggests two sources of AoA effects, primarily based on whether differences between early- and late-acquired words across various tasks reach significance.

Traditional analyses have not clarified the relative contributions of various influencing factors to AoA effects or compared their impacts. Therefore, the present study focused on behavioral research measures and employed three-level meta-analysis to systematically and comprehensively investigate the cognitive mechanisms of AoA effects. Traditional meta-analysis summarizes two-level structured data (Pastor & Lazowski, 2018), with result variation originating from two sources: variation in individual data and heterogeneity between studies. Meta-analysis requires the basic assumption of independent effect sizes. When a single study includes multiple effect sizes, these effect sizes are correlated due to factors such as sample and experimental context. Researchers typically select one effect size or average multiple effects, ignoring differences between effect sizes (Lipsey & Wilson, 2001). Based on our review of behavioral research on AoA effects, we found that many articles involve multiple experimental tasks and multiple effect sizes (Catling & Johnston, 2006b; Holmes & Ellis, 2006; Raman, 2018), with effect sizes that are not independent. To address the correlation problem among multiple effect sizes from the same study while maximizing the use of available effect sizes (Cheung, 2014), we adopted three-level meta-analysis, adding a level of variation to traditional meta-analysis: variation among multiple effect sizes from the same sample.

Based on theoretical debates, our meta-analysis systematically analyzed and compared how target cognitive processes, specific task types, and task relevance to semantic or phonological information moderate AoA effects. Second, according to the Arbitrary Mapping Hypothesis, different writing systems have different mapping relationships between representational levels, so we compared AoA effect sizes across writing systems (alphabetic vs. logographic). Additionally, based on linguistic characteristics of words, researchers have found that AoA effects are influenced by word frequency, so we also examined the moderating role of word frequency.

2.1 Literature Search and Inclusion Criteria

Chinese literature was retrieved from CNKI (including journal full-text database, master' s and doctoral dissertation database, and conference paper database), Wanfang Data Knowledge Service Platform, and VIP Chinese Journal Service Platform, using search terms including "lexical acquisition," "age of lexical acquisition," "early/late lexical acquisition," "age of acquisition effect," or "lexical learning age." English literature was obtained from Web of Science (including SCI, SSCI, A&HCI, ESCI, MEDLINE, Preprint, ProQuest), Science Direct, and Scopus databases, using keywords "age of acquisition," "AoA," "age of acquisition effect," or "AoA effect." All literature searches covered publications before May 15, 2024. The literature screening flowchart is shown in Figure 1 [Figure 1: see original paper].

Retrieved studies were screened according to the following inclusion criteria: (1) Published empirical studies on AoA effects, excluding conference papers, dissertations, and purely theoretical or review articles; (2) Exclusion of off-topic articles, such as second language acquisition studies or lexical/picture database construction studies not involving lexical processing tasks; (3) Studies with healthy adult participants, excluding special populations such as children, elderly adults, individuals with language disorders, or Alzheimer' s patients; (4) Experimental materials consisting of nouns, verbs, or adjectives, excluding studies using proper names or emotional words; (5) Only one study included when data were published multiple times; (6) Exclusion of studies not reporting reaction times, such as silent reading, delayed naming, or perceptual identification tasks. Selected studies must report Hedge' s g values for reaction times under different AoA conditions or provide data (reaction time means and standard deviations, F-values, t-values, β-values, R^2) that could be converted to Hedge' s g. A total of 130 studies met these criteria, including 13 Chinese and 117 English studies, yielding 271 effect sizes.

2.2 Literature Evaluation and Coding

To quantitatively synthesize the original studies, we coded the literature following Lipsey and Wilson' s (2001) method and calculated Kappa coefficients for inter-rater reliability using SPSS 26.0. The coded information and reliability coefficients included: (1) Study information (author, publication year); (2) Sample size (Kappa = 0.89); (3) Writing system (alphabetic vs. logographic) (Kappa = 0.84); (4) Target cognitive process (perception vs. production) (Kappa = 0.89); (5) Experimental task (picture naming, lexical decision, word naming, semantic categorization, and other tasks) (Kappa = 0.82); (6) Task semantic relevance (high vs. low) (Kappa = 0.82); (7) Task phonological relevance (high vs. low) (Kappa = 0.82); (8) Word frequency control (controlled, uncontrolled, or independent variable) (Kappa = 0.88); (9) Effect size data (reaction time means and standard deviations, F-values, t-values, β-values, R^2) (Kappa = 0.63).

Coding followed these principles: (1) Each independent sample was coded once for its main effect; if one article contained multiple independent samples, each was coded separately. (2) If the same sample reported multiple significant effect sizes, each effect size was coded separately. (3) If participants included different groups (e.g., young/elderly, healthy adults/individuals with dyslexia), only data from young and healthy adult groups were coded. (4) Alphabetic writing systems included English, French, Spanish, German, Dutch, Polish, Hindi, Icelandic, Turkish, Korean, and Kannada (a Dravidian language from southern India). Researchers consider Korean spelling rules closely related to pronunciation (Han & Kim, 2017; Han et al., 2021), and Kannada a spelling-consistent language (Nishimoto et al., 2005), so they were classified as alphabetic systems. Logographic writing systems primarily included Chinese studies. Due to the complexity of the Japanese writing system, Japanese materials were excluded from moderation analyses of writing system effects. (5) Task semantic relevance was coded based on task type and experimental manipulations. As described in the introduction, word naming tasks involve lower semantic processing depth compared to picture naming, lexical decision, and semantic categorization tasks (though this does not mean word naming involves no semantic processing). Therefore, word naming tasks were generally coded as low semantic relevance, while picture naming, lexical decision, and semantic categorization tasks were coded as high semantic relevance. Some lexical decision tasks used non-orthographic pseudowords; these effect sizes were included in main effect tests and moderation analyses for writing system and target cognitive process but excluded from moderation analyses for task type, task semantic relevance, and task phonological relevance. Some word naming tasks required semantic processing beyond reading aloud, such as reading only real words while ignoring pseudowords, or presenting verb stems and requiring participants to produce present progressive or past tense forms (Morrison et al., 2003); these tasks were excluded from moderation analyses for task type, task semantic relevance, and task phonological relevance. (6) Task phonological relevance was similarly coded based on task type and experimental manipulations, with picture naming and word naming tasks involving articulation coded as high phonological relevance, and lexical decision and semantic categorization tasks coded as low phonological relevance.

Coding was performed independently by two graduate students in psychology, followed by cross-checking. Disagreements were resolved through discussion between the two coders and the corresponding author. The final sample included 130 studies reporting 271 effect sizes from 10,697 participants (see Appendix 1), all published before May 15, 2024. The number of effect sizes per experiment ranged from 1 to 4.

Literature quality was assessed using the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies provided by the National Institutes of Health (NIH). After excluding four quality assessment items clearly inapplicable to AoA effect research, we scored each included study according to the criteria (1 point for meeting the standard, 0 points for not meeting it), with total scores ranging from 0 to 10 (National Institutes of Health, 2018). Higher scores indicated better study quality. Quality assessment results are presented in Appendix 1.

2.3 Meta-Analysis Procedure

We conducted data processing and analysis using R software and the metafor package. The meta-analysis procedure included: (1) testing for publication bias based on obtained effect sizes; (2) conducting heterogeneity tests, main effect tests, and moderation tests using three-level meta-analysis models, with subgroup analysis used to examine categorical moderators.

2.3.1 Effect Size Calculation

We used Hedge' s g as the effect size measure, with 0.10, 0.30, and 0.50 as thresholds for small, medium, and large effects (Cohen, 1992). Multiple regression studies calculated effect sizes from standardized regression coefficients (β), while factorial design studies primarily calculated effect sizes from reaction time means, standard deviations, and sample sizes, with some studies converting F-values to Hedge' s g (Ben-Shachar et al., 2020). We used R software and the metafor package for three-level meta-analysis.

2.3.2 Meta-Analysis Model Selection

Among the 130 included studies, 70 reported two or more effect sizes, with 141 effect sizes coming from the same sample sets, creating dependency among effect sizes (Cheung, 2014) and violating the independence assumption of traditional meta-analysis. Using traditional meta-analysis could lead to overestimated results (Lipsey & Wilson, 2001), whereas three-level meta-analysis can address this issue. Three-level meta-analysis decomposes variance into three sources: sampling error variance (Level 1: sampling variance), variation among effect sizes extracted from the same study (Level 2: within-study variance), and variation among effect sizes extracted from different studies (Level 3: between-study variance) (Cheung, 2014). This approach can extract all effect sizes from original studies, maximizing information retention and improving statistical power. For these reasons, we used three-level meta-analysis models (Guo et al., 2023; Meng et al., 2023; Meng et al., 2024; Zhao et al., 2024; Zhu et al., 2024) for heterogeneity tests, main effect tests, publication bias tests, and moderation tests.

2.3.3 Heterogeneity and Moderation Tests

We assessed overall heterogeneity using Q-tests and further determined heterogeneity distribution through one-tailed log likelihood ratio tests for within-study variance (Level 2) and between-study variance (Level 3). When tests were significant, indicating heterogeneity, we used I^2 values of 25%, 50%, and 75% as boundaries for low, medium, and high heterogeneity (Higgins et al., 2003) and conducted moderation tests to identify sources of heterogeneity (Meng et al., 2024; Assink & Wibbelink, 2016; Zhao et al., 2024). We added moderators as covariates to the three-level meta-analysis model (Meng et al., 2023) to estimate moderation effects. All moderators in this study were categorical variables, including writing system, cognitive process, task type, and word frequency. To ensure representativeness of moderation test results, each level of categorical moderators should contain at least 5 effect sizes (Card, 2016). Therefore, in moderation tests for task type, task semantic relevance, and task phonological relevance, we analyzed only the four main task types: picture naming, lexical decision, word naming, and semantic categorization. Effect sizes from other tasks, such as progressive demasking (Gilhooly & Logie, 1981b), picture verification (Catling & Elsherif, 2020), object reality judgment (Catling & Johnston, 2009), visual duration threshold task (Dent et al., 2007), degraded word recognition (Chen et al., 2010), and word recall (Gilhooly & Gilhooly, 1979), were included only in main effect tests and moderation tests for writing system and target cognitive process.

2.3.4 Publication Bias and Sensitivity Tests

Publication bias occurs when included literature does not comprehensively represent the total research in a field (Rothstein et al., 2005). To test for publication bias, we used funnel plots, Egger' s regression test, and fail-safe N. A funnel plot is a visual scatterplot of transformed effect sizes; in the absence of publication bias, data should be symmetrically distributed and concentrated in the upper middle region, forming a roughly symmetrical funnel shape (Light & Pillemer, 1984). In Egger' s regression test, a non-significant p-value indicates no serious publication bias (Egger et al., 1997). A fail-safe N greater than 5k + 10 (where k is the number of effect sizes) suggests that publication bias can be ignored (Rothstein et al., 2005). If publication bias was detected, we used the trim-and-fill method for bias correction (Duval & Tweedie, 2000). If effect sizes did not change significantly after trim-and-fill, the meta-analysis results were considered minimally affected by publication bias (Duval & Tweedie, 2000).

To assess the robustness of our meta-analysis results, we used the "one-study-removed" method to identify and exclude potential outliers that might significantly influence results, then re-conducted the three-level meta-analysis to evaluate the impact of anomalous effect sizes and studies. This method involves sequentially removing each included effect size and original study, then re-running the three-level meta-analysis until all effect sizes and original studies have been removed once (Dodell-Feder & Tamir, 2018).

3.1 Heterogeneity Test

The heterogeneity test for AoA effects showed that the Q statistic reached significance, $Q(270) = 1644.16$, $p < 0.001$, indicating heterogeneity among the included effect sizes. Both variation among effect sizes extracted from the same study (Level 2 variance) ($\sigma^2 = 0.05$, $p < 0.001$, $I^2 = 42.84\%$) and variation among effect sizes extracted from different studies (Level 3 variance) ($\sigma^2 = 0.05$, $p < 0.001$, $I^2 = 43.17\%$) were significant. The total I^2 value was 86.01%, meaning that 86.01% of observed variation was due to true differences among effect sizes and 13.99% was due to random error, indicating high heterogeneity between studies and the presence of between-group error interference. This suggests that writing system, cognitive process, task type, and word frequency may moderate AoA effects, necessitating further analysis of relevant moderators.

3.2 Publication Bias and Sensitivity Test

Egger' s regression test results showed a negative precision coefficient ($t = -4.31$, $df = 269$, $p < 0.001$) and an intercept of 0.64 (95% CI = [0.55, 0.73]), indicating significant funnel plot asymmetry. The funnel plot (see Figure 2 [Figure 2: see original paper]) showed that observed effect sizes were concentrated around 0.46 with a standard deviation of approximately 0.15, displaying an asymmetric distribution. To achieve symmetry, we applied the trim-and-fill method, which indicated that 0 (SE = 9.41) effect sizes needed to be imputed on the left side of the funnel plot, suggesting no obvious missing small or non-significant studies. The fail-safe N was 32,491, far exceeding the $5k + 10$ threshold, indicating that the current effect is highly robust. Thus, the main effect results were minimally affected by publication bias.

Using the one-study-removed method, we sequentially excluded each effect size and re-ran the three-level meta-analysis. Results showed that after removing one effect size reported by Karimi and Diaz (2020), the AoA effect reached its smallest magnitude (Hedge's $g = 0.46$, $df = 269$, $p < 0.001$), while after removing one effect size reported in Experiment 2 of Morrison and Gibbons (2006), the effect reached its largest magnitude (Hedge' s $g = 0.47$, $df = 269$, $p < 0.001$). Sensitivity analysis revealed that the main effect remained significant and of medium magnitude before and after removal, indicating that the meta-analysis results were not substantially influenced by outliers.

3.3 Moderation Test

To identify potential sources of effect size variation across studies, we used subgroup analysis to examine various categorical moderators of AoA effects, with each subgroup containing at least 21 effect sizes, exceeding the minimum of 3 per subgroup recommended by Lan et al. (2022).

Subgroup analysis examined whether moderators significantly influenced AoA effects (results shown in Table 1). The moderating effect of target cognitive process was not significant, $F(1, 269) = 0.08$, $p = 0.781$. Task type showed a significant moderating effect, $F(3, 224) = 9.32$, $p < 0.001$. Specifically, effect sizes for picture naming (Hedge' s $g = 0.55$) and lexical decision (Hedge' s $g = 0.59$) were significantly larger than those for word naming (Hedge' s $g = 0.34$) and semantic categorization (Hedge's $g = 0.34$), while picture naming and lexical decision did not differ significantly ($t(131) = -0.57$, $p = 0.569$). Task semantic relevance also showed a significant moderating effect, $F(1, 219) = 14.62$, $p < 0.001$, with larger AoA effects for high semantic relevance (Hedge' s $g = 0.53$) than low semantic relevance (Hedge' s $g = 0.34$). Task phonological relevance did not significantly moderate AoA effects, $F(1, 219) = 1.27$, $p = 0.261$. Writing system showed a marginally non-significant moderating effect, $F(1, 265) = 2.80$, $p = 0.096$, with logographic scripts showing numerically smaller effect sizes than alphabetic scripts (β = -0.11 (SE = 0.07), t = -1.67, p = 0.096, 95% CI = [-0.24, 0.02]). Word frequency control showed a significant moderating effect: compared to uncontrolled conditions (Hedge' s g = 0.37), both controlled word frequency (Hedge' s g = 0.53) and word frequency as an independent variable (Hedge' s g = 0.58) produced significantly larger effect sizes, with no significant difference between controlled and independent variable conditions (t(166) = -0.67, p = 0.503).

Among the 45 studies treating word frequency as an independent variable, 6 reported significant AoA $\times$ word frequency interactions, primarily in lexical decision tasks, with one picture naming and one word naming study each. In studies reporting AoA effects for both high- and low-frequency conditions (all with significant interactions), AoA effects were significantly smaller for high-frequency than low-frequency words (β = 0.80 (SE = 0.33), t(7) = -2.41, p = 0.042, 95% CI = [-1.56, -0.04]).

Table 1
Moderation Tests for Age of Acquisition Effects

Moderator	Hedge' s g (95% CI)	β (95% CI)	Within-Study Variance	Between-Study Variance
Intercept	0.46 (0.38, 0.53)***	-	< 0.001	< 0.001
Target Cognitive Process Task Type	0.47 (0.41, 0.54)***	0.01 (–0.08, 0.10)	-	-
Picture Naming	0.55 (0.47, 0.63)***	0.22 (0.10, 0.33)***	-	-
Lexical Decision	0.59 (0.49, 0.69)***	0.25 (0.13, 0.37)***	-	-
Word Naming	0.34 (0.25, 0.43)***	0.20 (0.10, 0.30)**	-	-
Semantic Categorization	0.34 (0.17, 0.51)***	0.20 (0.10, 0.30)**	-	-

Moderator	Hedge' s g (95% CI)	β (95% CI)	Within-Study Variance	Between-Study Variance
Semantic Relevance				
High	0.53 (0.47, 0.60)***	0.17 (0.07, 0.28)**	-	-
Low	0.34 (0.25, 0.43)***	-	-	-
Phonological Relevance				
High	0.52 (0.44, 0.61)***	-0.06 (-0.16, 0.04)	-	-
Low	0.47 (0.41, 0.54)***	-	-	-
Writing System				
Alphabetic	0.50 (0.45, 0.56)***	-0.11 (-0.24, 0.02)	-	-
Logographic	0.37 (0.29, 0.44)***	-	-	-
Word Frequency Control				
Uncontrolled	0.37 (0.29, 0.44)***	-	-	-
Controlled	0.53 (0.45, 0.61)***	0.22 (0.07, 0.37)**	-	-
Independent Variable	0.58 (0.42, 0.75)***	0.22 (0.07, 0.37)**	-	-

Note: $p < 0.05$, $\boldsymbol{p < 0.01}$, p < 0.001.

4.1 Moderating Roles of Task Type and Semantic Relevance

Meta-analysis results showed that whether the target process was lexical comprehension or production did not significantly affect AoA effect magnitude, while task type significantly influenced effect size. Specifically, all four tasks—picture naming, lexical decision, semantic categorization, and word naming—showed significant AoA effects ($p < 0.001$). Further analysis revealed that effect sizes for picture naming and lexical decision were significantly larger than those for word naming and semantic categorization, with no significant differences between picture naming and lexical decision ($t(131) = 0.57$, $p = 0.569$) or between semantic categorization and word naming ($t(93) = -0.32$, $p = 0.740$). Word naming primarily reflects phonological processing, while semantic categorization mainly involves semantic processing (Balota & Chumbley, 1984). These task-specific effect size patterns indicate that AoA is involved in semantic processing stages and may affect both semantic and phonological processing stages. Additionally, other tasks involving orthographic or perceptual processing (e.g., progressive demasking, degraded word recognition, visual duration threshold tasks) also reported significant AoA effects, suggesting that AoA influences perceptual processing stages as well. These results support the Multiple Loci Account, indicating that AoA effects span perceptual, semantic, and phonological processing stages.

The moderation analysis of task type showed a non-significant effect, a pattern inconsistent with comparisons of specific AoA effect magnitudes. Elsherif et al. (2023) calculated average AoA effect sizes across tasks and found that picture naming (107 ms) produced significantly larger effects than lexical decision (45 ms) and word naming (23 ms), with significant differences between lexical decision and word naming as well. That review used raw AoA effect magnitudes, which can be influenced by differences in experimental materials and sample sizes across tasks. In contrast, our study used Hedge' s g effect sizes, which are comparable across tasks, studies, and populations, controlling for sample size differences and providing more direct comparisons. Our literature inclusion process excluded potentially lower-quality conference papers and retained only studies using nouns and verbs as materials, increasing comparability across studies. Therefore, we believe our meta-analytic index more reliably reflects differences in AoA effects across task types.

Results showed equivalent AoA effect sizes for picture naming and lexical decision, both larger than those for semantic categorization and word naming. One possible explanation from Representation Theory suggests that both lexical decision (reflecting comprehension) and picture naming (reflecting production) involve semantic system processing, producing equivalent AoA effects, while word naming involves minimal semantic processing and primarily phonological processing (Cortese et al., 2018; Elsherif et al., 2020; Kuperman, 2013). Consistent with this explanation, we observed a significant effect of semantic relevance but not phonological relevance. Picture naming and lexical decision showed medium effect sizes, while word naming showed a small effect size, indicating that AoA effects are influenced by both task semantic relevance and phonological relevance (Elsherif et al., 2023).

A second possible explanation comes from the Multiple Loci Account. Both lexical decision and picture naming involve orthographic, phonological, and semantic systems, while semantic categorization primarily involves orthographic and semantic systems, and word naming involves orthographic and phonological systems. In terms of processing stages, lexical decision and picture naming include more processing stages than semantic categorization and word naming, producing larger AoA effects, with no significant difference between the latter two tasks. Nevertheless, researchers have argued that picture naming includes conceptual-level processing not present in lexical decision (Catling & Johnston, 2009), which should produce larger AoA effects in picture naming. Our meta-analytic results contradict this prediction, suggesting that other factors simultaneously influence AoA effect magnitudes.

We conducted a meta-analysis of ERP studies on AoA effects[3], which revealed that AoA reliably modulated N400 amplitude (Hedge' s $g = 0.45$, $p < 0.001$, 95% CI = [0.29, 0.62]), with early-acquired words eliciting smaller N400 amplitudes than late-acquired words. This indicates slower processing and more difficult semantic extraction for late-acquired words, demonstrating that semantic processing is involved in AoA effects and supporting Representation Theory. However, whether the N400 component reflects only semantic processing without phonological involvement requires further investigation of the relationship between electrophysiological and behavioral measures to explore brain-behavior relationships.

According to the Arbitrary Mapping Hypothesis, AoA effects stem from changes in lexical network plasticity during vocabulary acquisition and development, with mapping relationships between representational systems influencing AoA effects. Regular mapping relationships can partially compensate for processing disadvantages caused by reduced plasticity. Although lexical decision, semantic categorization, picture naming, and word naming correspond to comprehension and production processes, they share the same orthographic, semantic, and phonological representational systems. Word naming relies on orthographic-phonological mapping relationships at the lexical level, and both alphabetic and logographic systems have orthographic rules that enable phonological access, making orthographic-phonological mapping somewhat regular (Monaghan & Ellis, 2002a; Raman, 2006). Lexical decision and semantic categorization involve orthographic-semantic mapping relationships, where orthography is not directly linked to semantics, making orthographic-semantic mapping less consistent (Chen et al., 2007). Picture naming involves semantic-phonological mapping, where pronunciation is largely arbitrary and not necessarily connected to meaning, making semantic-phonological mapping consistency low. Therefore, due to differences in mapping regularity, picture naming and lexical decision show similar effect sizes, while semantic categorization and word naming show smaller effects than the former two tasks. Compared to the Multiple Loci Account and Representation Theory, the Arbitrary Mapping Hypothesis provides a more consistent explanation for the pattern of AoA effect sizes across tasks observed in this study.

Some researchers hold different views. Elsherif et al. (2023) argued that semantic-phonological mapping in picture naming is less regular and systematic than in lexical decision, so according to the Arbitrary Mapping Hypothesis, AoA effects should be smaller in lexical decision. When analyzing tasks along both semantic and phonological relevance dimensions, results showed significant moderation by semantic relevance but not phonological relevance. This indicates that tasks involving semantic processing produce larger AoA effects than tasks without semantic involvement, but phonological processing does not significantly affect AoA magnitude, suggesting that semantics rather than phonology plays a crucial moderating role. Previous research suggests AoA effects may occur at multiple processing stages, including conceptual, semantic, and phonological processing (Brysbaert & Ellis, 2016; Brysbaert et al., 2017; Elsherif et al., 2023; Hernandez & Li, 2007; Johnston & Barry, 2006; Juhasz, 2005). However, our meta-analytic results on phonological relevance suggest that phonology has limited influence. Although significant AoA effects were found in word naming, indicating that AoA affects phonological processing stages, effect size magnitude was not influenced by phonological processing depth. This finding suggests that investigating the cognitive mechanisms of AoA effects may require focusing on semantic representations or semantic networks to explore semantic-level influences and connections between semantic and other levels.

4.2 Moderating Role of Word Frequency

Meta-analysis results showed that compared to uncontrolled word frequency conditions, studies that matched word frequency between early- and late-acquired words or manipulated word frequency as an independent variable produced larger AoA effects. Additionally, when comparing effect sizes from multiple regression and factorial design studies, we found that multiple regression studies produced significantly smaller AoA effects than factorial design studies ($t(269) = 4.47$, $p < 0.001$). Uncontrolled word frequency studies all used multiple regression analysis, while studies controlling word frequency or treating it as an independent variable used factorial designs. We attribute the difference in AoA effects between these two study types to differences in experimental materials. In multiple regression analysis, researchers simultaneously consider the effects of AoA and word frequency, which are highly correlated (Morrison et al., 1997; Zevin & Seidenberg, 2002). When multiple predictors are highly correlated in regression analysis, the contribution of some variables may be obscured, preventing AoA effects from being fully manifested (Morris, 1981). Chen et al. (2004) used regression analysis in Chinese research to examine the influence of AoA and word frequency on disyllabic word naming, finding that word frequency affected naming latencies but AoA did not affect word naming. In factorial designs, word frequency and AoA are manipulated as independent variables with strict control, and other variables potentially affecting lexical processing must be matched. To independently manipulate these variables, the range of word frequency and AoA values in experimental materials is limited and cannot represent the overall distribution of vocabulary (Brysbaert & Ellis, 2016). In contrast, in regression analysis, word frequency and AoA are continuous variables, and materials used are more representative of actual vocabulary.

Researchers debate the cognitive mechanisms underlying AoA and word frequency effects. The Cumulative Frequency Hypothesis suggests that AoA effects stem from higher cumulative frequency of early-acquired words throughout the lifespan (Balota et al., 2004; Zevin & Seidenberg, 2004), while Representation Theory posits that AoA and word frequency share cognitive mechanisms (Brysbaert & Ellis, 2016; Brysbaert & Ghyselinck, 2006). Other research indicates that AoA and word frequency independently influence lexical processing (Cortese & Khanna, 2007; Gerhand & Barry, 1998; Chen et al., 2004). Our meta-analysis found that 23.8% of factorial design studies treating word frequency as an independent variable reported significant AoA × word frequency interactions, primarily in lexical decision tasks, with one picture naming and one word naming study each. In studies reporting AoA effects for both high- and low-frequency conditions (all with significant interactions), AoA effects were significantly smaller for high-frequency than low-frequency words ($\beta = 0.80$ (SE = 0.33), t(7) = -2.41, p = 0.042, 95% CI = [-1.56, -0.04]).

AoA and word frequency effects may be independent (Cortese & Khanna, 2007) or partially independent, with different cognitive mechanisms. Using a picture-word interference paradigm manipulating semantic and phonological relatedness between distractors and targets, researchers found that AoA interacted only with semantic relatedness, not phonological relatedness, while word frequency interacted with both, suggesting that AoA and word frequency jointly influence lexical selection in spoken word production but only word frequency affects phonological encoding (Zhang et al., 2022). Using high temporal resolution EEG, researchers asked participants to judge whether visually presented words contained specific letters. Results showed an interaction between AoA and word frequency on early selection negativity (SN) amplitude in the left occipitotemporal region between 180-240 ms, indicating joint influence on early lexical processing possibly related to orthographic processing in lexical comprehension. Additionally, in the 190-220 ms window, left anterior negativity (LAN) was independently affected by word frequency and AoA, with larger amplitudes for early-acquired and high-frequency words. Researchers interpreted this pattern as reflecting shared brain mechanisms for the two effects at the electrophysiological level (Adorni et al., 2013). Using computational simulation, researchers manipulated word occurrence ratios and found AoA × word frequency interactions when the high-low frequency ratio was 10:1 but not when it was 3:1 (Ellis & Lambon Ralph, 2000). Based on these computational and ERP findings, we propose that whether AoA and word frequency are independent or interactive depends on the range of word frequency values and relative frequency indices of early- versus late-acquired words.

4.3 AoA Effects Not Influenced by Writing System

Our meta-analysis found no significant effect of writing system on AoA effect magnitude ($p = 0.102$). According to the mapping hypothesis, the regularity of mapping between orthographic, semantic, and phonological systems in lexical processing influences AoA effects, with more regular mapping producing smaller effects. Logographic scripts have more regular orthographic-semantic mapping, while alphabetic scripts have more regular orthographic-phonological mapping. Considering that different lexical tasks involve different mapping relationships, we further compared AoA effect sizes for alphabetic and logographic scripts across the four task types. Results showed marginally larger AoA effects for alphabetic than logographic scripts in picture naming (alphabetic: $g = 0.57$ vs. logographic: $g = 0.31$, $t(77) = -1.75$, $p = 0.085$). Alphabetic scripts also showed numerically larger effects than logographic scripts in lexical decision (0.62 vs. 0.46) and semantic categorization (0.38 vs. 0.37), while alphabetic scripts showed numerically smaller effects than logographic scripts in word naming (0.37 vs. 0.41). Although these differences were not statistically significant, the pattern is generally consistent with the mapping hypothesis. Logographic scripts have stronger connections with semantics, making semantic access easier and reducing AoA effects in lexical decision (primarily orthographic-semantic mapping) and picture naming (primarily semantic-phonological mapping). In contrast, alphabetic scripts have more regular orthographic-phonological mapping, reducing AoA effects in word naming. Bakhtiar et al. (2016) examined vowel omission in Persian, which reduces writing complexity but makes orthographic-phonological mapping more arbitrary. Persian can thus be considered to have two types of scripts with different mapping consistency. Unlike Havelka and Tomita (2006), who found that writing system moderated AoA effects, Bakhtiar et al. found no AoA × vowel omission interaction in reaction times during lexical decision, but did find larger AoA effects for inconsistent mapping words in the LPC component reflecting later processing stages, supporting the Arbitrary Mapping Hypothesis. fMRI research has more directly implicated the left middle frontal gyrus in orthographic-phonological mapping in Chinese (Cao et al., 2009).

Although we found no significant effect of writing system, this does not necessarily contradict the Arbitrary Mapping Hypothesis. Possible explanations include: (1) Writing system may have a small effect on AoA effects. From the perspective of Representation Theory, writing system may affect the type or number of words individuals acquire (Haman et al., 2017) but not how words are connected in semantic networks, making it a non-moderating variable. The Multiple Loci Account focuses on processing stages involved in semantic tasks. Although lexical comprehension and production differ across writing systems (Lukatela & Turvey, 1994; Rubenstein et al., 1971; Peng et al., 1985), the four tasks we examined comprehensively cover perception and production stages, and writing system does not change the number of processing stages involved, so this account also does not predict writing system effects on AoA magnitude. (2) The result may be influenced by the distribution of studies across writing systems. Our meta-analysis included only 47 effect sizes from logographic scripts but 220 from alphabetic scripts, with uneven distribution of task types across writing systems, which may affect detection of moderating effects (Meng et al., 2024). This suggests that more logographic script research is needed to verify the stability of our results. (3) Even within the same writing system, mapping consistency between representational levels may vary substantially. For example, both Italian with high orthographic-phonological consistency and English with low consistency are considered alphabetic scripts (Wilson et al., 2012). To better test the mapping hypothesis, we suggest that future research should design experiments leveraging specific characteristics of writing systems, such as Chen et al.' s (2007) manipulation of consistency from phonetic radicals to whole-character pronunciation (orthographic-phonological predictability) in Chinese phono-semantic compounds to explore AoA effects under different conditions, which supported the Arbitrary Mapping Hypothesis.

4.4 Summary, Limitations, and Future Directions

In summary, our meta-analysis clearly demonstrates the important role of semantic representations in AoA effects and the influence of different mapping relationships across lexical tasks, providing potential support for both Representation Theory and mapping hypotheses. The three theoretical accounts emphasize that AoA effects originate from representational levels or connections between levels in lexical processing. Researchers have integrated these perspectives into an Integrated View, proposing that AoA effects are jointly determined by lexical representations of early- and late-acquired words and connection patterns between representations (Catling & Elsherif, 2020). Early-acquired words have richer semantic representations, occupy central positions in semantic networks, and have greater influence on semantic and corresponding neural network structures. Late-acquired words have weaker connections with other words and, due to reduced network plasticity, have less impact on neural network changes. Therefore, semantic networks are more suitable for input and output of early-acquired words, while processing late-acquired words requires greater effort to adapt to their specific input-output relationships. The Integrated View addresses both semantic representations and connection strength between representations, with both jointly determining AoA effects. Our meta-analytic results emphasize the role of semantic representations, reflecting that AoA-affected semantic networks have greater impact on lexical processing.

Our study has several limitations that warrant future exploration. First, additional factors influencing AoA effects need investigation. We selected task type and writing system as moderators to explore processing stages and cognitive mechanisms of AoA effects, but did not examine variables affecting semantic processing such as typicality (Räling et al., 2015) and concreteness (Song & Li, 2021), or variables affecting phonological processing such as predictability and regularity (Chen et al., 2007). Future research should incorporate more relevant variables for more comprehensive exploration and deeper understanding of how AoA influences lexical processing. Second, our meta-analysis did not clearly support the Arbitrary Mapping Hypothesis, inconsistent with some empirical findings (Monaghan & Ellis, 2002a; Havelka & Tomita, 2006). Previous research has not systematically distinguished mapping relationships between inputs and outputs across tasks. Future studies should manipulate different writing systems based on clear mapping relationships to investigate how mapping consistency influences AoA effects. We classified writing systems based on mapping relationships: logographic scripts have more consistent orthographic-semantic than orthographic-phonological mapping, while alphabetic scripts show the opposite pattern. However, mapping consistency across writing systems is somewhat confounded, as orthographic-phonological mapping in logographic scripts is not necessarily less consistent than in alphabetic scripts, and mapping consistency within the same writing system can be ambiguous. Future research should quantify input-output mapping relationships to more precisely explore how different mapping relationships affect AoA effects. Third, the Integrated View proposes that AoA effects originate from both representational levels and connections between levels. While this theory integrates Representation Theory and the Arbitrary Mapping Hypothesis, its specific claims remain relatively general and have not clarified underlying mechanisms, such as whether contributions from representational levels and inter-level connections are independent (i.e., whether they interact) and their relative contributions (Chang & Lee, 2020; Elsherif et al., 2023; Wang et al., 2023). AoA effects are closely related to vocabulary acquisition and development and have important theoretical implications for understanding brain learning mechanisms. Future research should use computational simulations of AoA effects in language processing with different mapping relationships and neuroimaging methods to reveal how brain structure and function represent mapping relationships, systematically investigating the cognitive and neural mechanisms of AoA effects and clarifying ambiguities in the Integrated View. Fourth, our study focused on four frequently used tasks, excluding other tasks such as picture writing (Bonin et al., 2002), delayed naming (Barry et al., 2001), rapid naming (Ghyselinck, Lewis, Brysbaert, 2004), and eye-tracking studies (Carter & Luke, 2020). As research accumulates, future studies should incorporate more task types and methodologies using current analytical techniques, employing multi-task, multi-modal approaches to further investigate AoA effects.

References (References Used in Meta-Analysis)

*Adorni, R., Manfredi, M., & Proverbio, A. M. (2013). Since when or how often? Dissociating the roles of age of acquisition (AoA) and lexical frequency in early visual word processing. Brain and Language, 124(1), 132−141. https://doi.org/10.1016/j.bandl.2012.11.005

*Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. Behavior Research Methods Instruments & Computers, 36(1), 140–155. https://doi.org/10.3758/bf03195559

*Baek, H., Gordon, P. C., & Choi, W. (2024). Effects of age and word frequency on Korean visual word recognition: Evidence from a web-based large-scale lexical-decision task. Psychology and Aging, 39(3), 231–244. https://doi.org/10.1037/pag0000793

*Bakhtiar, M., Nilipour, R., & Weekes, B. S. (2013). Predictors of timed picture naming in Persian. Behavior Research Methods, 44(3), 862–889. https://doi.org/10.3758/s13428-011-0175-8

*Bakhtiar, M., Su, I. F., Lee, H. K., & Weekes, B. S. (2016). Neural correlates of age of acquisition on visual word recognition in Persian. Journal of Neurolinguistics, 39, 1–9. https://doi.org/10.1016/j.jneuroling.2015.12.001

*Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. Journal of Experimental Psychology: General, 133(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

*Bangalore, S., Robson, H., & Astell, A. J. (2022). Standardizing norms for 180 coloured Snodgrass and Vanderwart pictures in Kannada language. Plos One, 17(4), e0266359. https://doi.org/10.1371/journal.pone.0266359

*Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. Journal of Memory and Language, 44(3), 350–375. https://doi.org/10.1006/jmla.2000.2743

*Barry, C., Johnston, R. A., & Wood, R. F. (2006). Effects of age of acquisition, age, and repetition priming on object naming. Visual Cognition, 13(7–8), 911–927. https://doi.org/10.1080/13506280544000101

*Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 50(3), 560–585. https://doi.org/10.1080/10.1080/783663595

*Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age acquisition in word reading and other tasks: A never ending story? Journal of Memory and Language, 50(4), 456–476. https://doi.org/10.1016/j.jml.2004.02.001

*Bonin, P., Boyer, B., Méot, A., Fayol, M., & Droit, S. (2004). Psycholinguistic norms for action photographs in French and their relationships with spoken and written latencies. Behavior Research Methods, Instruments, and Computers, 36(1), 127–139.

*Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. British Journal of Psychology, 93(1), 89−114. https://doi.org/10.1348/000712602162463

*Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. Behavior Research Methods, Instruments, and Computers, 35(1), 158−167. https://doi.org/10.3758/BF03195507

*Burani, C., Arduino, L. S., & Barca, L. (2007). Frequency, not age of acquisition, affects Italian word naming. European Journal of Cognitive Psychology, 19(6), 828−866. https://doi.org/10.1080/09541440600847946

*Busch, J. L., Haeussler, F. S., Domahs, F., Timmermann, L., Weber, I., & Oehrn, C. R. (2022). German normative data with naming latencies for 283 action pictures and 600 action verbs. Behavior Research Methods, 54(2), 649−662. https://doi.org/10.3758/s13428-021-01647-w

*Catling, J. C., Dent, K., & Williamson, S. (2008). Age of acquisition, not word frequency affects object recognition: Evidence from the effects of visual degradation. Acta Psychologica, 129(1), 130−137. https://doi.org/10.1016/j.actpsy.2008.05.005

*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., & Johnston, R. A. (2006a). Age of acquisition effects on an object-name verification task. British Journal of Psychology, 97(1), 1−18. https://doi.org/10.1348/000712605x53515

*Catling, J. C., & Johnston, R. A. (2006b). Effects of age of acquisition and priming on picture naming. Quarterly Journal of Experimental Psychology, 59(8), 1443−1453. https://doi.org/10.1080/17470210500214291

*Catling, J. C., & Johnston, R. A. (2006c). The effects of age of acquisition on an object classification task. Visual Cognition, 13(7−8), 968−980. https://doi.org/10.1080/13506280544000138

*Catling, J. C., & Johnston, R. A. (2009). The varying effects of age of acquisition. Quarterly Journal of Experimental Psychology, 62(1), 50−62. https://doi.org/10.1080/17470210701814352

*Chalard, M., & Bonin, P. (2006). Age-of-acquisition effects in picture naming: Are they structural and/or semantic in nature? Visual Cognition, 13(7−8), 864−883. https://doi.org/10.1080/13506280544000084

*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, 104025. https://doi.org/10.1016/j.jml.2019.05.003

*Chang, Y. N., & Lee, C. Y. (2020). Age of acquisition effects on traditional Chinese character naming and lexical decision. Psychonomic Bulletin & Review, 27(6), 1317−1324. https://doi.org/10.3758/s13423-020-01787-8

*Chen, B. G., Dent, K., You, W., & Wu, G. (2009). Age of acquisition affects early orthographic processing during Chinese character recognition. Acta Psychologica, 130(3), 196−203. https://doi.org/10.1016/j.actpsy.2008.12.004

*Chen, B. G., Zhou, H. X., Dunlap, S., & Perfetti, C. A. (2007). Age of acquisition effects in reading Chinese: Evidence in favour of the arbitrary mapping hypothesis. British Journal of Psychology, 98(3), 499−516. https://doi.org/10.1348/000712606x165484

*Colombo, L., & Burani, C. (2002). The influence of age of acquisition, root frequency, and context availability in processing nouns and verbs. Brain and Language, 81(1−3), 398−411. https://doi.org/10.1006/brln.2001.2533

*Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. Quarterly Journal of Experimental Psychology, 60(8), 1072−1082. https://doi.org/10.1080/17470210701315467

*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: Human Experimental Psychology, 73(11), 2036−2044. https://doi.org/10.1177/1747021820940302

*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

*Cuetos, F., & Alija, M. (2003). Normative data and naming times for action pictures. Behavior Research Methods, Instruments, and Computers, 35(1), 168−177. https://doi.org/10.3758/BF03195508

*Cuetos, F., & Barbon, A. (2006). Word naming in Spanish. European Journal of Cognitive Psychology, 18(3), 415−436. https://doi.org/10.1080/13594320500165896

*Cuetos, F., Barbón, A., Urrutia, M., & Domínguez, A. (2009). Determining the time course of lexical frequency and age of acquisition using ERP. Clinical Neurophysiology, 120(2), 285−294. https://doi.org/10.1016/j.clinph.2008.11.003

*Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. Behavior Research Methods, Instruments, and Computers, 31(4), 650−658. https://doi.org/10.3758/BF03200741

*Davies, R., Wilson, M., Cuetos, F., & Burani, C. (2014). Reading in Spanish and Italian: Effects of age of acquisition in transparent orthographies? Quarterly Journal of Experimental Psychology, 67(9), 1808−1825. https://doi.org/10.1080/17470218.2013.872155

*De Deyne, S., & Storms, G. (2007). Age-of-acquisition differences in young and older adults affect latencies in lexical decision and semantic categorization. Acta Psychologica, 124(3), 274—295. https://doi.org/10.1016/j.actpsy.2006.03.007

*Dell' Acqua, R., Lotto, L., & Remo, J. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. Behavior Research Methods, Instruments, and Computers, 32(4), 588—615. https://doi.org/10.3758/BF03200532

*den Hollander, J., Jonkers, R., Mariën, P., & Bastiaanse, R. (2019). Identifying the speech production stages in early and late adulthood by using electroencephalography. Frontiers in Human Neuroscience, 13, 298—298. https://doi.org/10.3389/fnhum.2019.00298

*Dent, K., Catling, J. C., & Johnston, R. A. (2007). Age of acquisition affects object recognition: Evidence from visual duration thresholds. Acta Psychologica, 125(3), 301—318. https://doi.org/10.1016/j.actpsy.2006.08.004

*Dewhurst, S. A., & Barry, C. (2006). Dissociating word frequency and age of acquisition: The Klein effect revived (and reversed). Journal of Experimental Psychology-Learning Memory and Cognition, 32(4), 919—924. https://doi.org/10.1037/0278-7393.32.4.919

*Egbert, M. H. Assink, Van Well, S., & Paul P. N. A. Knuijt. (2003). Contrasting effects of age of acquisition in lexical decision and letter detection. The American Journal of Psychology, 116(3), 367—387. https://doi.org/10.2307/1423499

*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/10.1006/jmla.2001.2834

*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, 49(5), 812—847. https://doi.org/10.1037/xlm0001215

*Fargier, R., & Laganaro, M. (2020). Neural dynamics of the production of newly acquired words relative to well-known words. Brain Research, 1727, 146557. https://doi.org/10.1016/j.brainres.2019.146557

*Fiebach, C. J., Friederici, A. D., Muller, K., von Cramon, D. Y., & Hernandez, A. E. (2003). Distinct brain representations for early and late learned words. Neuroimage, 19(4), 1627—1637. https://doi.org/10.1016/s1053-8119(03)00227-1

*Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. Journal of Experimental Psychology-Learning Memory and Cognition, 24(2), 267—283. https://doi.org/10.1037/0278-7393.24.2.267

*Gerhand, S., & Barry, C. (1999a). Age-of-acquisition and frequency effects in speeded word naming. Cognition, 73(2), B27—B36. https://doi.org/https://doi.org/10.1016/S0010-0277(99)00052-9

*Gerhand, S., & Barry, C. (1999b). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. Memory & Cognition, 27(4), 592—602. https://doi.org/10.3758/bf03211553

*Ghyselinck, M., Custers, R., & Brysbaert, M. (2004). The effect of age of acquisition in visual word processing: Further evidence for the semantic hypothesis. Journal of Experimental Psychology. Learning, Memory, and Cognition, 30(2), 550—554. https://doi.org/10.1037/0278-7393.30.2.550

*Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. Acta Psychologica, 115(1), 43—67. https://doi.org/10.1016/j.actpsy.2003.11.002

*Gilhooly, K. J. (1984). Word age-of-acquisition and residence time in lexical memory as factors in word naming. Current Psychological Research & Reviews, 3(2), 24—31. https://doi.org/10.1007/BF02686547

*Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. Memory & Cognition, 7(3), 214—223. https://doi.org/10.3758/BF03197541

*Gilhooly, K. J., & Logie, R. H. (1981a). Word age-of-acquisition, reading latencies and auditory recognition. Current Psychology, 1(3—4), 251—262. https://doi.org/10.1007/BF03186735

*Gilhooly, K. J., & Logie, R. H. (1981b). Word age-of-acquisition and visual recognition thresholds. Current Psychology, 1(3—4), 215—225. https://doi.org/10.1007/BF03186732

*Gilhooly, K. J., & Logie, R. H. (1982). Word age-of-acquisition and lexical decision making. Acta Psychologica, 50(1), 21—34. doi:10.1016/0001-6918(82)90048-8

*González-Nosti, M., Barbón, A., Rodríguez-Ferreiro, J., & Cuetos, F. (2014). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. Behavior Research Methods, 46(2), 517—525. https://doi.org/10.3758/s13428-013-0383-5

*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

*Hernandez, A. E., & Fiebach, C. J. (2006). The brain bases of reading late learned words: Evidence from functional MRI. Visual Cognition, 13(7—8), 1027–1043. https://doi.org/10.1080/13506280544000183

*Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. Visual Cognition, 13(7–8), 884–910. https://doi.org/10.1080/13506280544000093

*Izura, C., & Hernández-Muñoz, N. (2017). The role of semantics in Spanish word recognition: An insight from lexical decision and categorization tasks. Open Linguistics, 3(1), 500–515. https://doi.org/10.1515/opli-2017-0025

*Izura, C., & Playfoot, D. (2012). A normative study of acronyms and acronym naming. Behavior Research Methods, 44(3), 862–889. https://doi.org/10.3758/s13428-011-0175-8

*Johnson, Carla J. & Clark, James M. (1988). Children' s picture naming difficulty and errors: Effects of age of acquisition, uncertainty, and name generality. Applied Psycholinguistics, 9(4), 351–365. doi:10.1017/S0142716400008055

*Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British-English norms and naming times for a set of 539 pictures: The role of age of acquisition. Behavior Research Methods, 42(2), 461–469. https://doi.org/10.3758/BRM.42.2.461

*Karimi, H., & Diaz, M. (2020). When phonological neighborhood density both facilitates and impedes: Age of acquisition and name agreement interact with phonological neighborhood during word production. Memory & Cognition, 48(6), 1061–1072. https://doi.org/10.3758/s13421-020-01042-4

*Khwaileh, T., Body, R., & Herbert, R. (2014). A normative database and determinants of lexical retrieval for 186 Arabic nouns: Effects of psycholinguistic and morpho-syntactic variables on naming latency. Journal of Psycholinguistic Research, 43(6), 749–769. https://doi.org/10.1007/s10936-013-9277-z

*Laganaro, M. (2014). ERP topographic analyses from concept to articulation in word production studies. Frontiers in Psychology, 5, 493–493. https://doi.org/10.3389/fpsyg.2014.00493

*Laganaro, M., & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. Brain Topography, 24(1), 19–29. https://doi.org/10.1007/s10548-010-0162-x

*Laganaro, M., Valente, A., & Perret, C. (2012). Time course of word production in fast and slow speakers: A high density ERP topographic study. Neuroimage, 59(4), 3881–3888. https://doi.org/10.1016/j.neuroimage.2011.10.082

*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

*Laws, K. R., Leeson, V. C., & Gale, T. M. (2002). The effect of 'masking' on picture naming. Cortex, 38(2), 137–147. https://doi.org/10.1016/s0010-9452(08)70646-4

*Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. Plos One, 6(1), e16505. https://doi.org/10.1371/journal.pone.0016505

*Liu, Y., Hao, M., Shu, H., Tan, L. H., & Weekes, B. S. (2008). Age-of-acquisition effects on oral reading in Chinese. Psychonomic Bulletin & Review, 15(2), 344–350. https://doi.org/10.3758/pbr.15.2.344

*Lou, H., Li, C., & Zhang, Q. (2019). Effects of age of acquisition on object and action picture naming: An ERP study. Acta Psychologica Sinica, 51(2), 143–153. https://doi.org/10.3724/sp.J.1041.2019.00143

*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

*Meschyan, G., & Hernandez, A. (2002). Age of acquisition and word frequency: Determinants of object-naming speed and accuracy. Memory & Cognition, 30(2), 262–269. https://doi.org/10.3758/bf03195287

*Monaghan, J., & Ellis, A. W. (2002a). Age of acquisition and the completeness of phonological representations. Reading and Writing, 15(7–8), 759–788.

*Monaghan, J., & Ellis, A. W. (2002b). What exactly interacts with spelling-sound consistency in word naming? Journal of Experimental Psychology-Learning Memory and Cognition, 28(1), 183–206. https://doi.org/10.1037//0278-7393.28.1.183

*Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(1), 116–133. https://doi.org/10.1037/0278-7393.21.1.116

*Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. British Journal of Psychology, 91(2), 167–180. https://doi.org/10.1348/000712600161763

*Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. Memory & Cognition, 20(6), 705–714. https://doi.org/10.3758/BF03202720

*Morrison, C. M., & Gibbons, Z. C. (2006). Lexical determinants of semantic processing speed. Visual Cognition, 13(7–8), 949–967. https://doi.org/10.1080/13506280544000129

*Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. Quarterly Journal of Experimental

Psychology Section A: Human Experimental Psychology, 50(3), 528–559. https://doi.org/10.1080/027249897392017

*Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of acquisition, ageing, and verb production: Normative and experimental data. Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 56(4), 705–730. https://doi.org/10.1080/02724980244000594

*Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. Reading Research Quarterly, 24(3), 262–282. https://doi.org/10.2307/747770

*Navarrete, E., Scaltritti, M., Mulatti, C., & Peressotti, F. (2013). Age-of-acquisition effects in delayed picture-naming tasks. Psychonomic Bulletin and Review, 20(1), 148–153. https://doi.org/10.3758/s13423-012-0310-2

*Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. Behavior Research Methods, 37(3), 398–416. https://doi.org/10.3758/BF03192709

*Perret, C., Bonin, P., & Laganaro, M. (2014). Exploring the multiple-level hypothesis of AoA effects in spoken and written object naming using a topographic ERP analysis. Brain and Language, 135, 20–31. https://doi.org/10.1016/j.brainl.2014.04.006

*Pind, J., & Tryggvadóttir, H. B. (2002). Determinants of picture naming times in Icelandic. Scandinavian Journal of Psychology, 43(3), 221–226. https://doi.org/10.1111/1467-9450.00290

*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

*Räling, R., Holzgrefe-Lang, J., Schröder, A., & Wartenburger, I. (2015). On the influence of typicality and age of acquisition on semantic processing: Diverging evidence from behavioural and ERP responses. Neuropsychologia, 75, 186–200. https://doi.org/10.1016/j.neuropsychologia.2015.05.031

*Räling, R., Hanne, S., Schröder, A., Keßler, C., & Wartenburger, I. (2017). Judging the animacy of words: The influence of typicality and age of acquisition in a semantic decision task. Quarterly Journal of Experimental Psychology, 70(10), 2094–2104. https://doi.org/10.1080/17470218.2016.1223704

*Raman, I. (2006). On the age-of-acquisition effects in word naming and orthographic transparency: Mapping specific or universal? Visual Cognition, 13(7–8), 1044–1053. https://doi.org/10.1080/13506280500153200

*Raman, I. (2011). The role of age of acquisition in picture and word naming in dyslexic adults. British Journal of Psychology, 102, 328–339. https://doi.org/10.1348/000712610x522572

*Raman, I. (2018). The role of context on age of acquisition effect: Strategic control in word naming in Turkish. In Psycholinguistics and cognition in language processing (pp. 19-48). IGI Global.

*Raman, I., Raman, E. İkier S., Kilecioğlu, E., Uzun, E. 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.

*Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., & Brown, G. D. A. (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

*Schwier, V., Boyer, B., Meot, A., Bonin, P., & Laganaro, M. (2004). French normative data and naming times for action pictures. Behavior Research Methods Instruments & Computers, 36(3), 564—576. https://doi.org/10.3758/bf03195603

*Sereno, S. C., & O' Donnell, P. J. (2009). Participant and word gender in age of acquisition effects: The role of gender socialization. Sex Roles, 61(7—8), 510—518. https://doi.org/10.1007/s11199-009-9649-x

*Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. Acta Psychologica, 119(2), 159—187. https://doi.org/10.1016/j.actpsy.2005.01.002

*Shao, Z., Roelofs, A., & Meyer, A. S. (2014). Predicting naming latencies for action pictures: Dutch norms. Behavior Research Methods, 46(1), 274—283. https://doi.org/10.3758/s13428-013-0358-6

*Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. Behavior Research Methods, Instruments, and Computers, 28(4), 516—536. https://doi.org/10.3758/BF03200540

*Song, D., & Li, D. (2021). Psycholinguistic norms for 3,783 two-character words in simplified Chinese. Sage Open, 11(4). https://doi.org/10.1177/21582440211054495

*Spataro, P., Lpngobardi, E., Saraulli, D., & Rossi-Arnaud, C. (2013). Interactive effects of age-of-acquisition and repetition priming in the lexical decision task: A multiple-loci account. Experimental Psychology, 60(4), 235—242. https://doi.org/10.1027/1618-3169/a000192

*Tainturier, M. J., Tamminen, J., & Thierry, G. (2005). Age of acquisition modulates the amplitude of the P300 component in spoken word recognition. Neuroscience Letters, 379(1), 17—22. https://doi.org/10.1016/j.neulet.2004.12.038

*Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. Memory and Cognition, 26(6), 1282—1291. https://doi.org/10.3758/BF03201200

*Urooj, U., Cornelissen, P. L., Simpson, M. I. G., Wheat, K. L., Woods, W., Barca, L., & Ellis, A. W. (2014). Interactions between visual and semantic processing during object recognition revealed by modulatory effects of age of acquisition. NeuroImage, 87, 252–264. https://doi.org/10.1016/j.neuroimage.2013.10.058

*Valente, A., Buerki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: A trial by trial multiple regression analysis from stimulus onset to response. Frontiers in Neuroscience, 8, 390. https://doi.org/10.3389/fnins.2014.00390

*Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. The Quarterly Journal of Experimental Psychology Section A, 48(4), 822–848. https://doi.org/10.1080/14640749508401419

*Wang, J., Jiang, X., & Chen, B. (2023). Second language age of acquisition effects in a word naming task: A regression analysis of ERP data. Journal of Neurolinguistics, 66, 101125. https://doi.org/10.1016/j.jneuroling.2023.101125

*Wang, Z., Zhang, L., & Xuan, B. (2024). Age-related Chinese word recognition across different AoA and parts of speech. Current Psychology, 43(22), 19939–19952. https://doi.org/10.1007/s12144-024-05794-z

*Weekes, B. S., Chan, A., & Tan, L. H. (2008). Effects of age of acquisition on brain activation during Chinese character recognition. Neuropsychologia, 46(7), 2086–2090. https://doi.org/10.1016/j.neuropsychologia.2008.01.020

*Weekes, B. S., Chan, A., Kwok, J. S. W., Hia Tan, L., & Jin, Z. (2004). AoA effects on Chinese language processing: An fMRI study. Brain and Language, 91(1), 33–34. https://doi.org/10.1016/j.bandl.2004.06.020

*Weekes, B. S., Shu, H., Hao, M., Liu, Y., & Tan, L. H. (2007). Predictors of timed picture naming in Chinese. Behavior Research Methods, 39(2), 335–342. https://doi.org/10.3758/bf03193165

*Wilson, M. A., Cuetos, F., Davies, R., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: The role of item imageability. Journal of Experimental Psychology-Learning Memory and Cognition, 39(6), 1842–1859. https://doi.org/10.1037/a0033090

*Wilson, M. A., Ellis, A. W., & Burani, C. (2012). Age-of-acquisition affects word naming in Italian only when stress is irregular. Acta Psychologica, 139(3), 417–424. https://doi.org/10.1016/j.actpsy.2011.12.012

*Wolna, A., Luniewska, M., Haman, E., & Wodniecka, Z. (2023). Polish norms for a set of colored drawings of 168 objects and 146 actions with predictors of naming performance. Behavior Research Methods, 55(5), 2706–2732. https://doi.org/10.3758/s13428-022-01923-3

*Xu, X., Li, J., & Guo, S. (2021). Age of acquisition ratings for 19,716 simplified Chinese words. Behavior Research Methods, 53(2), 558–573. https://doi.org/10.3758/s13428-020-01455-8

*Yamazaki, M., Ellis, A. W., Morrison, C. M., & Ralph, M. A. L. (1997). Two age of acquisition effects in the reading of Japanese Kanji. British Journal of Psychology, 88(3), 407−421.https://doi.org/10.1111/j.2044-8295.1997.tb02648.x

*Young, A. W., Bion, P. J., Ellis, A. W. (1982). Age of reading acquisition does not affect visual hemifield asymmetries for naming imageable nouns. Cortex, 18(3), 477−482.

*Yum, Y. N., & Law, S. P. (2019). Interactions of age of acquisition and lexical frequency effects with phonological regularity: An ERP study. Psychophysiology, 56(10), e13433. https://doi.org/10.1111/psyp.13433

*Zhang, M., Liu, Z., Botezatu, M. R., Dang, Q., Yuan, Q., Han, J., ···Guo, T. (2023). A large-scale database of Chinese characters and words collected from elementary school textbooks. Behavior Research Methods, 56(5), 4732−4757. https://doi.org/10.3758/s13428-023-02214-1

*Zhang, Q., Zhou, Y., & Lou, H. (2022). The dissociation between age of acquisition and word frequency effects in Chinese spoken picture naming. Psychological Research-Psychologische Forschung, 86(6), 1918−1929. https://doi.org/10.1007/s00426-021-01616-0

*Bai, L., & Chen, B. (2011). The effect of age of acquisition of Chinese characters on the processing of word class information. Psychological Science, 34(2), 343−347. https://doi.org/10.16719/j.cnki.1671-6981.2011.02.031

*Bai, X., Wang, L., Lü, Y., & Hu, W. (2007). An ERP study of age of acquisition effects under auditory presentation conditions. (Abstract). In Proceedings of the 3rd National Conference on Brain and Cognitive Sciences and International Symposium on Brain and Cognitive Sciences (p. 24). Kunming, Yunnan.

*Bai, X., Wang, L., Lü, Y., & Hu, W. (2010). Age of acquisition effects in lexical processing: An ERP study. Psychological Exploration, 30(1), 21−26. https://doi.org/10.3969/j.issn.1003-5184.2010.01.005

*Chen, B., Wang, L., Wang, L., & Peng, D. (2004). Effects of age of acquisition and word frequency on lexical recognition. Psychological Science, 27(5), 1060−1064. https://doi.org/10.16719/j.cnki.1671-6981.2004.05.009

*Chen, B., You, W., & Zhou, H. (2007). Age of acquisition effects in Chinese lexical acquisition: Evidence for the semantic hypothesis. Acta Psychologica Sinica, 39(1), 9−17.

*Chen, B., You, W., Zhang, Y., & Liu, W. (2010). Age of acquisition effects in early orthographic processing stages of Chinese character recognition. Psychological Science, 33(3), 726−728.

*Chen, X., & Zhang, J. (2010). A regression analysis of factors affecting Chinese verb and noun recognition. Psychological Science, 33(1), 60−63. doi:10.16719/j.cnki.1671-6981.2010.01.014.

*Chen, Y., & Zhu, L. (2015). Dissociation of age of acquisition effects for spoken and written language in lexical naming. (Abstract). In Abstracts of the 13th

Annual Conference of the Developmental Psychology Professional Committee of the Chinese Psychological Society (p. 72). Changchun, Jilin.

*Hao, M., Liu, Y., Shu, H., & Cheng, C. (2003). The role of age of acquisition in Chinese picture naming. Studies of Psychology and Behavior, 1(4), 268−273.

*Hao, S., Chen, J., & Xue, L. (2014). A preliminary exploration of the mechanism underlying age of acquisition effects under auditory conditions. Psychological Science, 37(2), 277−282. doi:10.16719/j.cnki.1671-6981.2014.02.006.

*Li, C., Zhang, Q., & Huang, R. (2017). Different roles of age of acquisition in object and action picture naming. Psychological Exploration, 37(3), 220−225.

*Lou, H., Li, C., & Zhang, Q. (2019). Different effects of age of acquisition on spoken naming of object and action pictures: An ERP study. Acta Psychologica Sinica, 51(2), 143−153. https://doi.org/10.3724/sp.J.1041.2019.00143

*Wang, L., Wang, Y., & Yan, G. (2010). An eye movement study of age of acquisition effects. Studies of Psychology and Behavior, 8(4), 289−295.

*Zhang, J., Wang, J., & Yin, C. (2014). The role of phonetic and semantic radicals in phonological and semantic extraction of phono-semantic compounds: Evidence from eye movement studies of component perception. Acta Psychologica Sinica, 46(7), 885−900. https://doi.org/10.3724/SP.J.1041.2014.00885

*Zhang, Z., Ding, G., & Chen, B. (2011). Age of acquisition effects in Chinese character acquisition: Testing the phonological completeness hypothesis. Psychological Development and Education, 27(6), 577−583.

*Zhao, Z., Yu, J., Chen, J., Wang, Y., Huang, J., & Chen, C. (2024). Symptom and functional changes in individuals at clinical high risk for psychosis: A systematic review and three-level meta-analysis. Advances in Psychological Science, 33(1), 42−61.

*Zhu, Y., He, B., & Sun, L. (2024). The effect of state power on prosocial behavior: A three-level meta-analysis. Advances in Psychological Science, 32(11), 1786−1799.

---

[3] The ERP meta-analysis included 11 studies with 17 effect sizes from 421 participants. Due to the small number of ERP studies and dispersed amplitude components (including N400, LPC, SN, P2, P3, N540, and studies reporting significant amplitude differences without specifying components), we considered only N400 results reliable according to meta-analytic standards. See Appendix 2 for details.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.