ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Cross-Linguistic Perspectives on Aging and Sentence Processing: The Role of Word Order and Semantic Plausibility

Jee Eun Sung,[a] Gayle DeDe,[b] Seunghun J. Lee,[c,d] and Jimin Park[a]

[a] Department of Communication Disorders, Ewha Womans University, Seoul, South Korea [b] Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA [c] College of Liberal Arts, International Christian University, Tokyo, Japan [d] Department of African Languages, University of Venda, Thohoyandou, South Africa

ARTICLE INFO

Article History:
Received September 26, 2025
Revision received January 22, 2026
Accepted April 16, 2026

Editor-in-Chief: Julie A. Washington
Editor: Stephen M. Wilson

https://doi.org/10.1044/2026_JSLHR-25-00765

ABSTRACT

Purpose: This cross-linguistic study investigates how healthy aging affects sentence comprehension across three typologically distinct languages—English (head-initial), Korean, and Japanese (both head-final). It explores how syntactic word order and semantic plausibility interact with aging and whether these effects differ across languages with varying structural characteristics.
Method: A total of 290 adults participated: 103 English speakers (49 younger, 54 older), 98 Korean speakers (50 younger, 48 older), and 89 Japanese speakers (43 younger, 46 older). All participants completed a sentence–picture matching task using auditorily presented stimuli. The sentences were systematically manipulated for semantic plausibility (plausible vs. less plausible) and word order, contrasting instrument-first (instrument–theme) and theme-first (theme–instrument) structures.
Result: Aging effects were observed across all languages, with older adults performing worse than younger adults. Across language groups, participants showed better performance for instrument-first sentences than for theme-first sentences. Semantic plausibility interacted significantly with age across all languages in both accuracy and response times, indicating that older adults were disproportionately affected by reduced plausibility. In contrast, word order effects were largely similar across age groups, although a significant Age Group × Word Order interaction was observed in response times for the Japanese participants only.
Conclusions: This study demonstrates that systematic manipulations of simple sentence structures can effectively expose age-related and cross-linguistic differences in sentence processing. The findings illuminate both universal patterns of cognitive aging and language-specific parsing mechanisms, informing the design of sensitive and typologically flexible tools for assessing cognitive–linguistic decline.

Aging affects multiple cognitive and linguistic domains, including declines in working memory, slower language processing, and changes in both speech production and comprehension. Although much of the existing research has focused on age-related changes in speech production—such as alterations in acoustic properties, slower speech rates, increased pauses, and a higher frequency of filler usage (Beier et al., 2023; Horton et al., 2010)—the topic of aging effects in sentence comprehension has recently yielded a substantial body of research (Murphy, 2025; Reifegerste et al., 2017; Wingfield et al., 2015). Much of this work has concentrated on structurally complex constructions, such as object-relative clauses (Caplan & Waters, 2005) or passive structures (Sung et al., 2020). In contrast, relatively few studies have examined aging effects using paradigms that minimize syntactic and working-memory demands, such as minimal-phrase constructions involving simple agreement relations (e.g., pronoun–verb agreement violations; Poulisse et al., 2019). While minimal-phrase paradigms have been used to investigate

Correspondence to Jee Eun Sung: jeesung@ewha.ac.kr. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

basic combinatory processes in language comprehension, aging effects within such minimal structures remain relatively underexplored.

A key challenge in the previous line of research lies in identifying sentence-processing tasks that are capable of detecting subtle impairments with language-general and culturally agnostic paradigms that can be administered across aging populations whose languages differ substantially in syntactic organization and lexical–semantic structure. The present study addresses these limitations by adopting a sentence template that can be implemented across typologically different languages while maintaining comparable structural configurations. Although the same basic template is used across languages, the consequences of constituent ordering differ depending on the grammatical properties of each language. In particular, the interaction between constituent order and verb position differs between head-initial languages such as English and head-final languages such as Korean and Japanese. This design therefore allows us to examine how aging affects sentence processing under comparable structural conditions while also revealing how language-specific grammatical properties modulate these effects.

In the present study, we focus on sentence structures that impose relatively low syntactic computational demands. The stimuli are short monoclausal active sentences that do not involve syntactic embedding, long-distance dependencies, or movement of core arguments. Instead, the manipulation involves the linear ordering of two thematic elements (theme and instrument) within a single clause. This design differs from many previous studies of aging and sentence comprehension, which often rely on syntactically complex constructions such as passives, object-relative clauses, or sentences involving multiple levels of embedding. Such constructions are known to impose substantial working memory demands during parsing (Caplan & Waters, 2005; Gibson, 1998), making it difficult to determine whether observed aging effects reflect sensitivity to syntactic complexity per se or broader age-related changes in processing resources. By contrast, the present design allows us to examine how aging affects sentence comprehension when syntactic structure itself remains relatively simple while systematically manipulating semantic plausibility and constituent ordering.

To operationalize this paradigm, the sentence structures used in the present study were informed by the Western Aphasia Battery (WAB; Kertesz, 2007, 2022), a widely used clinical assessment tool for neurogenic communication disorders. In the WAB auditory comprehension subtest, a small set of sentences is used to assess the ability to follow serial commands. While these sentence types are effective for identifying the presence of comprehension difficulties, the limited number of items constrains the ability to precisely localize the source of difficulties or to examine graded effects of linguistic manipulations. Motivated by these sentence structures, Sung et al. (2024) developed a more fine-grained assessment to evaluate sentence comprehension in individuals with aphasia by systematically manipulating semantic plausibility and word order in sentences containing instrumental phrases (e.g., "Point to the book with the pencil"). In Korean-speaking individuals with aphasia—a language that permits relatively flexible word order—the authors found reduced accuracy for less plausible sentences compared to plausible ones, with this effect being more pronounced than in healthy controls. Moreover, when semantic plausibility and word order were jointly manipulated, particularly in less plausible sentences with scrambled structures, performance showed the most pronounced decline across both groups.

Importantly, prior findings demonstrate that even simple variations in semantic and syntactic features within active-voice sentences can substantially impact sentence processing, despite the absence of overt syntactic complexity and despite flexibility in word order. Building on this insight, the present study adopts aphasia-informed sentence templates not to equate aphasia and aging mechanisms but to leverage methodologically sensitive, low-complexity sentence structures capable of revealing language-general and language-specific effects of aging. Extending the recent work by Sung et al. (2024), we examine cross-linguistic differences in age-related effects using a simple syntactic structure across languages that vary in their reliance on word order and semantic cues for sentence comprehension. By applying this paradigm to English, Korean, and Japanese, the current study enables a comparative analysis of how aging modulates sentence processing across distinct linguistic systems, thereby offering deeper insight into the mechanisms underlying age-related cognitive and linguistic decline.

Korean and Japanese were selected for this study because they are both head-final languages with relatively flexible word order due to their case-marking systems, whereas English is a head-initial language. The inclusion of these three language groups was motivated by several factors. First, while Sung et al. (2024) focused exclusively on Korean-speaking individuals with aphasia, it remains unclear whether their findings are specific to Korean or if they generalize to other head-final languages, such as Japanese, which shares a similar syntactic structure but differs in phonological representations and writing systems. By including Japanese, we aimed to test the hypothesis that the linguistic manipulations in this study reflect broader head-final linguistic features rather than being unique to Korean. Furthermore, English, as a head-initial language, serves as a critical point of comparison, allowing us to examine how fundamental structural differences influence sentence processing across different language families. By contrasting head-final and head-initial languages, we can gain deeper insights into the interaction between syntactic

structures and semantic plausibility, particularly in the context of aging-related processing declines.

To systematically investigate these cross-linguistic effects, the current study employed the same stimuli used in Sung et al. (2024), which consisted of sequential command sentences such as "Point to the book with the pencil" (a plausible sentence in which the theme precedes the instrument: theme-first order) and "With the book, point to the pencil" (a less plausible sentence in which the instrumental phrase precedes the theme: instrument-first order). These sentence types are particularly well suited for the present investigation because they are relatively short, active-voice constructions that permit controlled manipulation of linear constituent order of the thematic role assignment and semantic plausibility without introducing overt syntactic complexity.

From a semantic perspective, the sentence "With the book, point to the pencil" is expected to be more difficult to process than "Point to the book with the pencil" because the instrumental relation is less plausible, given that books are unlikely instruments for pointing. Reduced semantic plausibility is therefore predicted to increase processing demands across languages. With respect to word order, different expectations arise across languages. In verb-final languages such as Korean and Japanese, an instrumental-first order aligns with commonly attested word order patterns. The instrumental phrase is marked by the instrumental case marker, whereas the theme is marked by the accusative case marker. The noun phrase (NP) bearing the accusative case marker functions as an internal argument positioned adjacent to the clause-final verb. In contrast, a theme-first order places the internal argument further from the verb, potentially increasing integration demands. Thus, for Korean and Japanese, the theme–instrument order may impose additional word order–related processing difficulty beyond the effects of semantic plausibility. For English, however, an instrumental-first word order is less conventional and less frequent ("With the pencil, point to the book") than the theme-first order ("Point to the book with the pencil"; Biber et al., 2021), although both structures are grammatically well formed. Nevertheless, the instrumental-first condition aligns with the isomorphic hypothesis, which proposes that processing is facilitated when the order of linguistic constituents corresponds to the sequential order of the described actions (O'Grady & Lee, 2005). In this case, the instrument-first order may facilitate semantic interpretation by aligning the order of thematic roles with the sequence of the action, despite its lower frequency in English.

Taken together, these considerations suggest that semantic plausibility effects are expected to be robust across languages, whereas word order effects may vary across languages and interact with language-specific structural and discourse conventions. For English, if instrumental-first order enhances semantic accessibility through alignment between theme role order and the sequence of the action, this ordering may partially offset processing difficulty associated with reduced plausibility. However, if the less conventional and less frequent of instrumental-PP (prepositional phrase)–initial ordering outweighs any potential semantic facilitation—particularly when combined with low plausibility—English speakers may exhibit greater processing costs in the instrument-first, less plausible condition than speakers of Korean or Japanese. Such a pattern would indicate a language-specific interaction between semantic plausibility and word order in sentence processing during aging.

From a clinical perspective, these items are of particular interest because they occur in the Sequential Commands subtest of the WAB, a widely used aphasia test battery across multiple languages (Chengappa & Kumar, 2008; Kertesz, 2022; Kim & Na, 2012; Neves et al., 2014). These items are among the most sensitive indicators of comprehension difficulties on the WAB (Lazar et al., 2008; Lwi et al., 2021), but it remains unclear whether individuals with aphasia struggle with them primarily due to syntactic or semantic difficulties. A better understanding of how older adults process these structures, and the respective contributions of syntactic and semantic complexity, will not only shed light on aging-related sentence processing but also inform clinical interpretations of WAB performance across different linguistic backgrounds.

Given these considerations, the current study has two primary objectives: first, to investigate whether aging effects emerge when semantic and syntactic features are systematically manipulated, and second, to examine whether these linguistic manipulations yield different effects between head-final (Korean and Japanese) and head-initial (English) languages, as well as within head-final languages (Korean vs. Japanese). From a theoretical perspective, predictions can be informed by both memory-based parsing accounts and heuristic-based processing models. According to the dependency locality theory (Gibson, 1998, 2000), processing difficulty increases as the distance between syntactically dependent elements grows because greater working memory resources are required to maintain incomplete dependencies during sentence comprehension. In verb-final languages such as Korean and Japanese, the instrument–theme–verb configuration places the theme NP—the internal argument—closer to the clause-final verb, thereby reducing integration cost.

In contrast, the theme–instrument–verb configuration increases the linear distance between the theme NP and the verb, creating a longer dependency and potentially greater integration demands. Under this account, the

instrument-first order may therefore be relatively easier to process than the theme-first order in Korean and Japanese. In contrast, predictions for English may be better explained by good-enough processing accounts (Ferreira & Patson, 2007), which propose that comprehenders often rely on semantic and event structure cues rather than constructing fully detailed syntactic representations, particularly under conditions of increased processing demands such as aging. Although instrument-first ordering (e.g., *With the pencil, point to the book*) is a relatively marked linear configuration in English, it may facilitate interpretation when it aligns with the isomorphic mapping between linguistic structure and the sequence of the described action (O'Grady & Lee, 2005). In such cases, the order of thematic roles in the sentence mirrors the order of the action required in the task, potentially providing a salient semantic cue that supports interpretation without requiring detailed syntactic computation. This type of semantic alignment may be particularly beneficial for older adults, who may rely more heavily on heuristic cues during comprehension.

Based on these considerations, we hypothesized that reduced semantic plausibility would impose greater processing demands for older adults than for younger adults across all languages, reflecting a robust language-general effect of semantic constraints. In contrast, word order effects—operationalized as theme-first versus instrument-first ordering—were expected to vary across languages and interact with language-specific structural and discourse conventions. Specifically, in verb-final languages such as Korean and Japanese, instrument-first order was expected to reduce integration costs by placing the theme NP closer to the verb, whereas theme-first order may introduce greater processing difficulty due to longer dependency distances. For English speakers, however, instrument-first order may engage semantic heuristics such as isomorphic alignment, potentially offsetting the processing costs associated with its marked linear order. Depending on whether semantic alignment or word order markedness predominates, this configuration could either mitigate or exacerbate processing difficulty. By systematically comparing these effects across languages and age groups, the present study aims to clarify how semantic plausibility and word order jointly shape sentence comprehension in aging, thereby contributing to cross-linguistic models of sentence processing and informing the development of sensitive assessment tools for age-related language change.

Method

Participants

A total of 317 individuals participated in this study, comprising 103 English-speaking adults (49 young vs. 54 older), 106 Korean-speaking adults (50 young vs. 56 older), and 108 Japanese-speaking adults (59 young vs. 49 older). English speakers were obtained from Temple University and the surrounding community. They completed the Health Screening Questionnaire (Christensen et al., 1991), confirming no history of neurological or psychiatric conditions, and the Mini-Mental State Examination (MMSE; Folstein et al., 1975), with all scoring above the cutoff of 24. Additionally, hearing was screened using pure-tone audiometry (PTA) across frequencies ranging from 250 to 8000 Hz. Only individuals with hearing thresholds below 25 dB HL were included in the study, based on averaged thresholds across tested frequencies. The final sample was composed of 49 young and 54 older adults, with no significant difference in education between groups ($p = .166$). This portion of the study was approved by the institutional review board for human subjects at Temple University (No. 24858), and all participants provided written informed consent prior to participation.

Korean speakers, recruited from local communities in Seoul, completed the Health Screening Questionnaire (Christensen et al., 1991), the Korean version of MMSE (Kang, 2006), and PTA. Eight older adults were excluded as they fell beyond the age- and education-adjusted normal range (< -2 *SD*s), resulting in a final sample of 50 young and 48 older adults. All participants included in the final analysis passed the health and hearing screenings, reporting no neurological or psychiatric history and demonstrating hearing thresholds below 25 dB HL. The two groups differed significantly in education ($p < .001$). The research was approved by the institutional review board for human subjects at Ewha Womans University (No. 2022-0112), and all participants provided written informed consent before taking part in the study.

Japanese speakers were recruited from the local communities in Tokyo and completed the Health Screening Questionnaire (Christensen et al., 1991). Due to incomplete participation, 15 younger and three older adults were excluded, resulting in a final sample of 43 younger and 46 older participants. All included individuals reported no history of neurological and psychiatric conditions. The Japanese version of the MMSE was not administered, as it could not be obtained without certified personnel. In addition, because data collection for the Japanese group overlapped with the COVID-19 pandemic, PTA was not conducted due to restrictions on in-person testing. Instead, hearing status was evaluated using the Health Screening Questionnaire, which included structured self-report items on hearing. The groups also differed significantly in education ($p < .001$). The Japanese version of the MMSE was not administered, as it could not be purchased without certified personnel. This research was approved by the institutional review board for human subjects at International Christian University (No.

2017-28), and all participants gave written informed consent prior to participation.

In total, 290 participants—103 English, 98 Korean, and 89 Japanese speakers—were included in the final analysis. Demographic information for each language is provided in Table 1.

Experimental Stimuli

The experimental stimuli consisted of 96 sentences, including 64 target sentences and 32 filler sentences. The target sentences varied with respect to plausibility (plausible vs. less plausible) and word order (theme-first vs. instrument-first), resulting in four different conditions each. Word order was manipulated by varying the linear order of two phrases in relative to the verb: a theme phrase (e.g., *the book*) and an instrument phrase (e.g., *the pencil*). In the theme-first condition, the theme precedes the instrument (e.g., *Point to the book with the pencil*), whereas in the instrumental-first condition, the instrument precedes the theme (e.g., *With the book, point to the pencil*). Sentences were classified as plausible when the NPs formed a logical semantic relationship with the verb (e.g., "pencil" as the instrumental argument, "book" as the theme argument, and "point" as the action verb).

The stimuli comprised four different verbs (*point*, *hit*, *push*, *tap*) and 16 distinct nouns (e.g., *hammer*, *box*, *pencil*, *book*), with each verb appearing in four sentences across different alternating NP positions. A subset of stimuli was adapted from Sung et al. (2024). To minimize potential repetition effects on word order processing, target sentences were pseudo-randomized such that no more than three consecutive sentences appeared in the same condition.

Experimental Procedures

Participants underwent cognitive screening tests in their languages. Those who met the inclusion criteria took part in the main task individually in a quiet environment. Experimental stimuli were audio-recorded by a native speaker for each language in a sound-attenuated booth. These recordings were then uploaded into E-prime software (Psychology Software Tools) for use in the experiment. Experimental sentences were presented auditorily through headphones, while two pictures depicting the target sentence and the reversed NP were simultaneously displayed on the screen using the E-prime software (refer to Table 2). Participants were instructed to select the correct picture as quickly as possible by pressing the letter key "q" (left) or "p" (right) after listening to the sentence. Before the main trials began, all participants underwent three practice trials to ensure they understood the instructions.

Data Analysis

Accuracy and response time (RT) were measured for each participant. For accuracy, correct responses were coded as 1 and incorrect responses as 0, with the total sum representing the overall accuracy. A generalized linear mixed-effects model with a logit link function was applied to analyze accuracy data. The analysis was conducted in R (R Core Team, 2022) using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages. The initial model included four fixed effects: age group (young vs. old), plausibility (less plausible vs. plausible), and word order (theme-first vs. instrument-first). Following the recommendations of Barr et al. (2013), a maximal random-effects structure was initially specified,

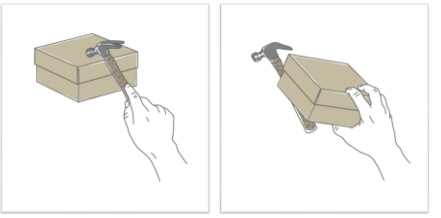
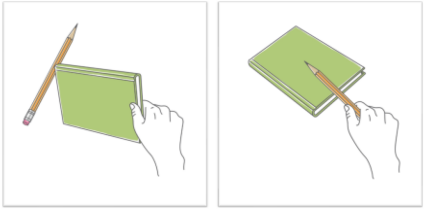
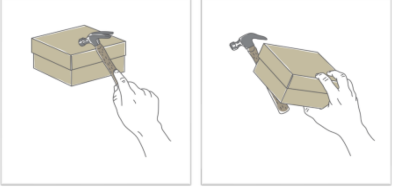
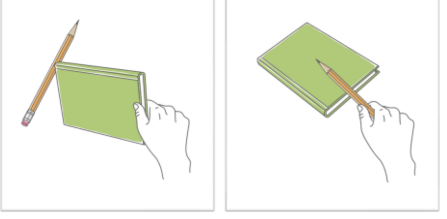
Table 1. Demographic information of participants.

Variable	English		Korean		Japanese	
	Young (N = 49)	Older (N = 54)	Young (N = 50)	Older (N = 48)	Young (N = 43)	Older (N = 46)
Age (*SD*) [Range]	20.76 (2.31) [18–29]	70.35 (34.76) [64–78]	27.76 (3.30) [20–34]	76.10 (9.62) [60–95]	22.62 (6.41) [18–29]	78.22 (6.41) [66–91]
t-test results	$t(89.33) = -81.45$, $p < .001$		$t(57.50) = -32.99$, $p < .001$		$t(55.51) = -56.79$, $p < .001$	
Education (*SD*) [Range]	13.96 (2.00) [12–21]	14.60 (2.66) [11–22]	16.18 (1.27) [12–18]	8.93 (5.56) [0–18]	12.71 (1.60) [12–16]	13.91 (2.26) [9–18]
t-test results	$t(97.78) = -1.40$, $p = .166$		$t(51.72) = 8.82$, $p < .001$		$t(75.93) = -3.42$, $p = .001$	
MMSE (*SD*) [Range]	29.22 (0.92) [27–30]	27.72 (1.66) [24–30]	29.70 (0.58) [28–30]	25.90 (4.50) [8–30]	—	—
t-test results	$t(84.11) = 5.74$, $p < .001$		$t(48.56) = 5.90$, $p < .001$			
Sex (M:F)	12:37	11:43	13:37	12:36	8:36*	16:30

Note. One participant's sex information was not reported. Em dashes indicate data not obtained or not available. MMSE = Mini-Mental State Examination (Folstein et al., 1975; Kang, 2006).

*$p < .05$.

Table 2. Examples of target sentences.

Plausibility	Language	Word order	Stimuli	Picture example
Less plausible	English	Theme-first	Point to the hammer with the box	
	Korean		망치-를 상자-로 두드리다 Mangchi-*reul* Sangja-*ro* Dudeur*ida* Hammer-$_{ACC}$ box-$_{INS}$ Tap-$_{verb}$	
	Japanese		かなづち-を 箱-で たたく Kanazuchi-o Hako-de Tataku Hammer-$_{ACC}$ Box-$_{INS}$ tap-$_{verb}$	
	English	Instrument-first	With the book, point to the pencil.	
	Korean		책-으로 연필-을 가리키다 Chaek-euro Yeonpil-eul Garikida Book-$_{INS}$ pencil-$_{ACC}$ Point-$_{verb}$	
	Japanese		本-で 鉛筆-を 指す Hon-de Enpitsu-o sasu Book-$_{INS}$ pencil-$_{ACC}$ Point-$_{verb}$	
Plausible	English	Theme-first	Point to the box with the hammer.	
	Korean		상자-를 망치-로 두드리다 Sangja-reul Mangchi-ro Dudeurida Box-$_{ACC}$ Hammer-$_{INS}$ Tap-$_{verb}$	
	Japanese		箱-を かなづち-で たたく Hako-o Kanazuchi-de Tataku Box-$_{ACC}$ Hammer-$_{INS}$ Tap-$_{verb}$	
	English	Instrument-first	With the pencil, point to the book.	
	Korean		연필-로 책-을 가리키다 Yeonpil-ro Chaek-eul Garikida Pencil-$_{INS}$ book-$_{ACC}$ Point-$_{verb}$	
	Japanese		鉛筆-で 本-を 指す Enpitsu-de Hon-o Sasu Pencil-$_{INS}$ book-$_{ACC}$ Point-$_{verb}$	

Note. Korean romanization follows the Revised Romanization system; Japanese romanization follows the Hepburn system. ACC = accusative case marker; INS = instrumental case marker.

including random intercepts and slopes for both participants and items. However, the full model failed to converge. To address convergence issues, accuracy data were analyzed separately for each language group. Categorical predictors were sum-coded as follows: age group (−1 = old, +1 = young), plausibility (−1 = plausible, +1 = less plausible), and word order (−1 = theme-first, +1 = instrument-first). Each model included random intercepts for participants and trials, with the Plausibility × Word Order interaction retained as a random slope by participants. The bobyqa optimizer was applied across all language groups to improve convergence. For the Korean and Japanese data sets, education was included as a covariate to account for age-related group differences.

For RT analysis, measurements were taken from the onset of sentence presentation to the participant's button press. Only correct trials were included. Outliers were identified and excluded using Tukey's method, removing values falling below Q1 − 1.5 × interquartile range or above Q3 + 1.5 × IQR (Tukey, 1977). To meet model assumptions and stabilize variance, raw RT data were log-transformed prior to analysis. A linear mixed-effects model was then fitted using the same fixed effects and an initially maximal random-effects structure. The revealed a significant four-way interaction of language, age group, plausibility, and word order, as well as several lower-order interactions, including Language × Word Order, Language × Plausibility, Age Group × Language, Language × Plausibility × Word Order, Language × Age Group × Plausibility, and Language × Age Group × Word order. To facilitate interpretation and maintain consistency with the accuracy analysis, RT data were analyzed separately for each language group. Categorical variables were sum-coded: age group (−1 young, +1 old), plausibility (−1 plausible, +1 less plausible), and word order (−1 = theme-first, +1 = instrument-first). The models included random intercepts for participants and trials, with word order retained as a random slope by participants, forming a reduced but theoretically motivated random effects structure. The bobyqa optimizer was applied across all language groups to aid convergence. For the Japanese and Korean data sets, education was included as a covariate to account for significant age-related group differences.

Results

Accuracy

Descriptive statistical results for accuracy are presented in Table 3. To analyze accuracy by age group, plausibility, and word order across English, Korean, and Japanese participants, a generalized linear mixed-effects model with a logit link function was computed, with the findings detailed in Table 4.

English Group

The overall pattern of results is illustrated in Figure 1A. A significant main effect of age group was observed ($\beta = 0.337$, $SE = 0.169$, $z = 2.000$, $p = .046$), indicating that young adults ($M = 0.97$, $SD = 0.16$) performed significantly better than older adults ($M = 0.94$, $SD = 0.24$). Moreover, the main effect of word order was significant ($\beta = -0.450$, $SE = 0.153$, $z = -2.94$, $p = .003$), with instrument-first sentences ($M = 0.98$, $SD = 0.15$) showing significantly higher accuracy than theme-first sentences ($M = 0.93$, $SD = 0.25$). In contrast, there was no significant main effect of plausibility ($p = .999$; plausible: $M = 0.96$, $SD = 0.19$; less plausible: $M = 0.95$, $SD = 0.23$).

The two-way interaction between age group and plausibility was significant ($\beta = 0.355$, $SE = 0.110$, $z = 3.23$, $p = .001$). Post hoc comparisons revealed that young adults performed significantly better than older adults on less plausible sentences ($\beta = 1.385$, $SE = 0.416$, $z = 3.330$, $p = .002$), whereas no significant group differences were found for plausible counterparts ($p = 1.000$; see Figure 2A). Additionally, a significant two-way interaction was observed between plausibility and word order ($\beta = -0.241$, $SE = 0.119$, $z = -2.032$, $p = .042$). Follow-up analyses showed that the advantage for instrument-first over theme-first sentences was significant for less plausible sentences ($\beta = -1.38$, $SE = 0.43$, $z = -3.205$, $p = .003$), but not significant for plausible ones ($p = .431$; see Figure 2B). There were no significant effects for the interaction between group and word order ($p = .421$) or for the three-way interaction among age group, plausibility, and word order ($p = .542$).

Korean Group

The overall pattern of results is illustrated in Figure 3A. For the analysis of Korean speakers, education was included as a covariate to account for significant difference in education levels between young and older adults ($p < .001$). A significant main effect of age group was observed ($\beta = 0.919$, $SE = 0.212$, $z = 4.34$, $p < .001$), indicating that young adults ($M = 0.98$, $SD = 0.16$) outperformed older adults ($M = 0.74$, $SD = 0.44$). A significant main effect of word order was also found ($\beta = -0.604$ $SE = 0.123$, $z = -4.913$, $p < .001$), with theme-first sentences ($M = 0.82$, $SD = 0.39$) eliciting lower accuracy than instrument-first sentences ($M = 0.91$, $SD = 0.29$). However, the main effect of plausibility was not significant ($p = .075$).

Regarding interaction effects, the two-way interaction between age group and plausibility was significant ($\beta = 0.513$, $SE = 0.115$, $z = 4.450$, $p < .001$). Post hoc comparisons revealed that young adults performed significantly better than older adults on less plausible sentences ($\beta = 2.864$, $SE = 0.526$, $z = 5.446$, $p < .0001$), while no significant difference was found between groups on plausible

Table 3. Accuracy of sentences by plausibility, word order, and group for each language.

Plausibility	Word order	English		Korean		Japanese	
		Young	Older	Young	Older	Young	Older
Less plausible	Theme-first (*SD*)	0.97 (0.17)	0.89 (0.32)	0.96 (0.19)	0.53 (0.50)	0.97 (0.17)	0.71 (0.45)
	Instrumental-first (*SD*)	0.99 (0.11)	0.96 (0.20)	0.99 (0.11)	0.72 (0.45)	0.98 (0.14)	0.84 (0.36)
Plausible	Theme-first (*SD*)	0.96 (0.20)	0.94 (0.25)	0.96 (0.19)	0.79 (0.41)	0.95 (0.22)	0.83 (0.37)
	Instrumental-first (*SD*)	0.98 (0.15)	0.98 (0.14)	0.99 (0.11)	0.92 (0.26)	0.96 (0.19)	0.93 (0.26)

Note. Mean values are presented.

counterparts ($p = .1230$; see Figure 4A). Moreover, a significant two-way interaction emerged between plausibility and word order ($\beta = -0.224$, $SE = 0.097$, $z = -2.294$, $p = .022$). Follow-up analysis showed that the effect of plausibility was significant in theme-first sentences ($\beta = 2.864$, $SE = 0.526$, $z = 5.446$, $p < .0001$), but not in instrument-first ones ($p = 1.000$; see Figure 4B). In contrast, the interaction between age group and word order was not significant ($p = .709$), nor was the three-way interaction among age group, plausibility, and word order ($p = .254$).

Japanese Group

The overall pattern of results is illustrated in Figure 5A. For Japanese participants, education was included as a covariate due to a significant difference in education levels between young and older adults ($p = .017$). The final model revealed a significant main effect of age group ($\beta = 1.003$, $SE = 0.1402$, $z = 7.154$, $p < .001$), demonstrating that younger adults ($M = 0.97$, $SD = 0.18$) performed significantly better compared to older adults ($M = 0.84$, $SD = 0.38$). Furthermore, the main effect of word order was significant ($\beta = -0.318$, $SE = 0.099$, $z = -3.221$, $p = .001$), indicating lower accuracy for theme-first ($M = 0.87$, $SD = 0.34$) compared to instrument-first sentences ($M = 0.93$, $SD = 0.26$). In contrast, the main effect of plausibility was not significant ($p = .648$).

The two-way interaction between age group and plausibility was significant ($\beta = 0.386$, $SE = 0.089$, $z = 4.350$, $p < .001$). Post hoc analyses revealed that young adults performed significantly better on less plausible compared to plausible sentences ($\beta = 0.864$, $SE = 0.310$, $z = 2.786$, $p = .011$), whereas older adults showed the reversed pattern, performing better on plausible than the less plausible counterparts ($\beta = -0.678$, $SE = 0.223$, $z = -3.042$, $p = .005$; see Figure 6A). In contrast, the two-way interactions between age group and word order ($p = .056$) and between plausibility and word order ($p = .202$), as well as the three-way interaction among the age group, plausibility, and word order ($p = .733$), were not statistically significant.

Response Time

Statistical descriptions of RT are presented in Table 5. A linear mixed-effects model was employed, with raw RT data log-transformed, to examine the effects of age group, plausibility, and word order within each language group (see Table 6).

English Group

English speakers showed a significant main effect of age group ($\beta = 0.150$, $SE = 0.021$, $df = 100.50$, $t = 6.96$, $p < .001$), with older adults exhibiting significantly longer RT ($M = 3{,}599.85$, $SD = 1{,}340.49$) than the young adults ($M = 2{,}636.75$, $SD = 911.38$; see Figure 1B). There was also a significant main effect of plausibility ($\beta = 0.023$, $SE = 0.005$, $df = 1747.000$, $t = 4.40$, $p < .001$), indicating that less plausible sentences elicited longer RT ($M = 3{,}204.79$, $SD = 1{,}296.86$) than the plausible counterparts ($M = 3{,}067.22$, $SD = 1{,}197.16$). The main effect of word order was also significant ($\beta = 0.031$, $SE = 0.006$, $df = 380.60$, $t = 5.32$, $p < .001$) because the theme-first sentences showed longer RT ($M = 3{,}199.95$, $SD = 1{,}319.38$) compared to the instrument-first sentences ($M = 3{,}072.43$, $SD = 1{,}172.29$).

A significant two-way interaction was found between age group and plausibility ($\beta = 0.013$, $SE = 0.004$, $df = 4691.00$, $t = 3.66$, $p < .001$). Post hoc analyses revealed that the plausibility effect was significant among older adults ($p < .001$), but not among young adults ($p = .259$; see Figure 2C). In contrast, no significant two-way interactions were observed between age group and word order ($p = .361$) or between plausibility and word order ($p = .685$). The three-way interaction among age group, plausibility, and word order was also not significant ($p = .291$).

Korean Group

For Korean speakers, education was included as a covariate since there was a significant difference in education levels between young and older adults ($p < .001$). The final model revealed a significant main effect for age

Table 4. Summary of the generalized linear mixed effects models for each language.

Predictors	English				Korean				Japanese			
	β	*SE*	*z*	*p* value	β	*SE*	*z*	*p* value	β	*SE*	*z*	*p* value
(Intercept)	4.405	0.191	23.089	< .001***	1.382	0.446	3.095	.002**	1.081	0.860	1.256	.209
Education (covariate)	—				0.156	0.034	4.590	< .001***	0.154	0.065	2.392	.017**
Group	0.337	0.169	1.996	.046*	0.919	0.212	4.337	< .001***	1.003	0.140	7.154	< .001***
Plausibility	0.000	0.122	0.001	.999	−0.218	0.122	−1.780	.075	0.046	0.102	0.456	.648
Word order	−0.450	0.153	−2.942	.003***	−0.604	0.123	−4.913	< .001***	−0.318	0.099	−3.221	< .01**
Group × Plausibility	0.355	0.110	3.231	< .001***	0.513	0.115	4.450	< .001***	0.386	0.089	4.350	< .001***
Group × Word Order	0.113	0.141	0.805	.421	0.043	0.116	0.373	.709	0.163	0.085	1.914	.056
Plausibility × Word Order	−0.241	0.119	−2.032	.042	−0.223	0.097	−2.294	.022*	−0.107	0.084	−1.277	.202
Group × Plausibility × Word Order	−0.066	0.107	−0.610	.542	−0.101	0.088	−1.140	.254	−0.023	0.068	−0.341	.733

Note. R model equation: Accuracy ~ Age Group × Plausibility × Word Order + (1 + Plausibility × Word Order | Participant) + (1 | Trial). Degrees of freedom were estimated using Laplace approximation. β = fixed-effect estimate; *SE* = standard error.

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

Figure 1. Effects of age group, word order, and plausibility in English speakers. (A) Accuracy. (B) Response time. Error bars represent ±1 standard error.

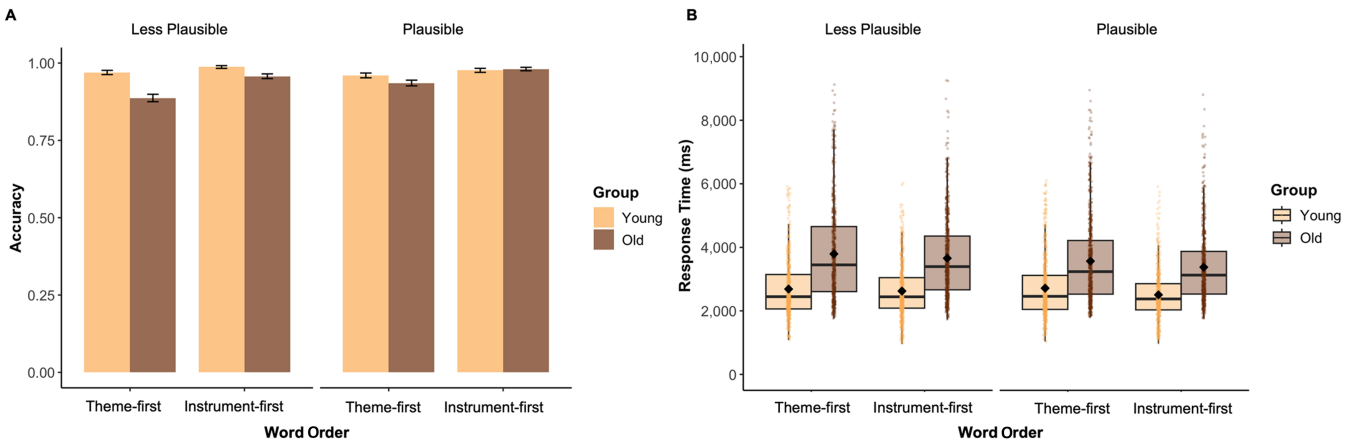


Figure 2. Significant interaction effects in English speakers. (A) Accuracy interaction between age group and plausibility. (B) Accuracy interaction between plausibility and word order. (C) Response time interaction between age group and plausibility. Diamonds indicate means, horizontal lines indicate medians, and dots represent individual data points. $^{**}p < .01$, $^{***}p < .001$.

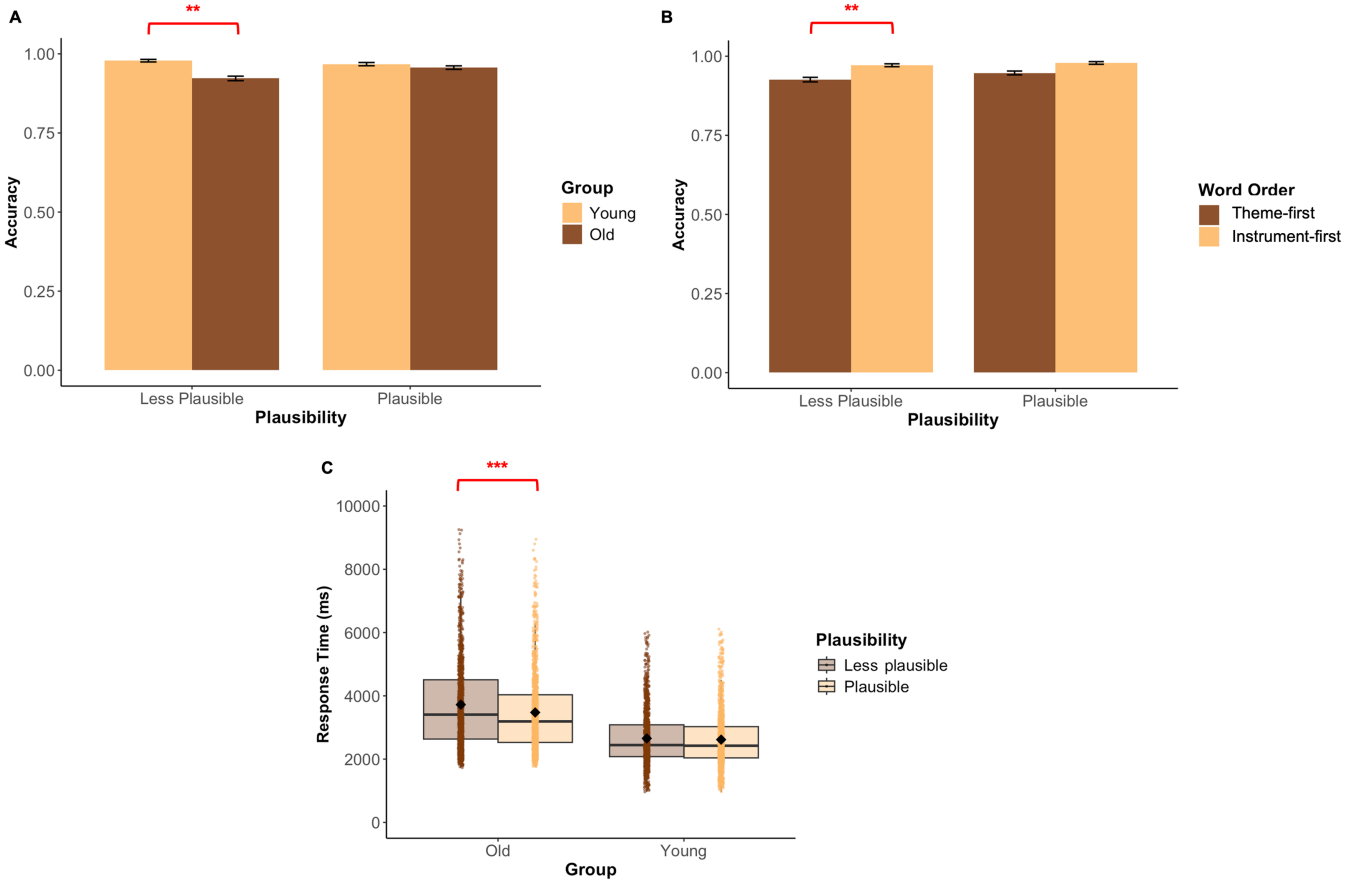


Figure 3. Effects of age group, word order, and plausibility in Korean speakers. (A) Accuracy. (B) Response time. Error bars represent ±1 standard error.

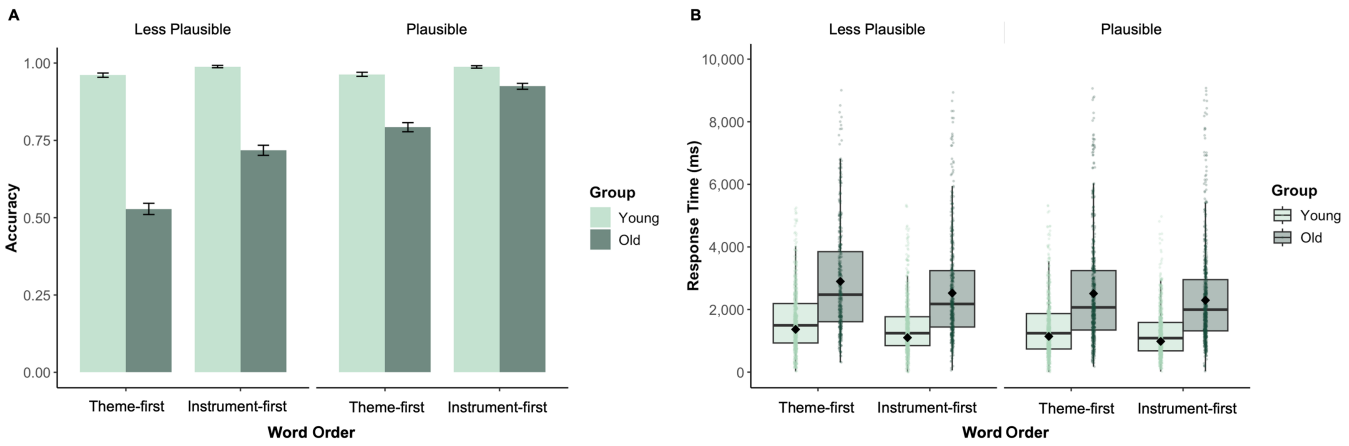


Figure 4. Significant interaction effects in Korean speakers. (A) Accuracy interaction between age group and plausibility. (B) Accuracy interaction between plausibility and word order. (C) Response time interaction between age group and plausibility. Diamonds indicate means, horizontal lines indicate medians, and dots represent individual data points. *p < .05, ***p < .001.

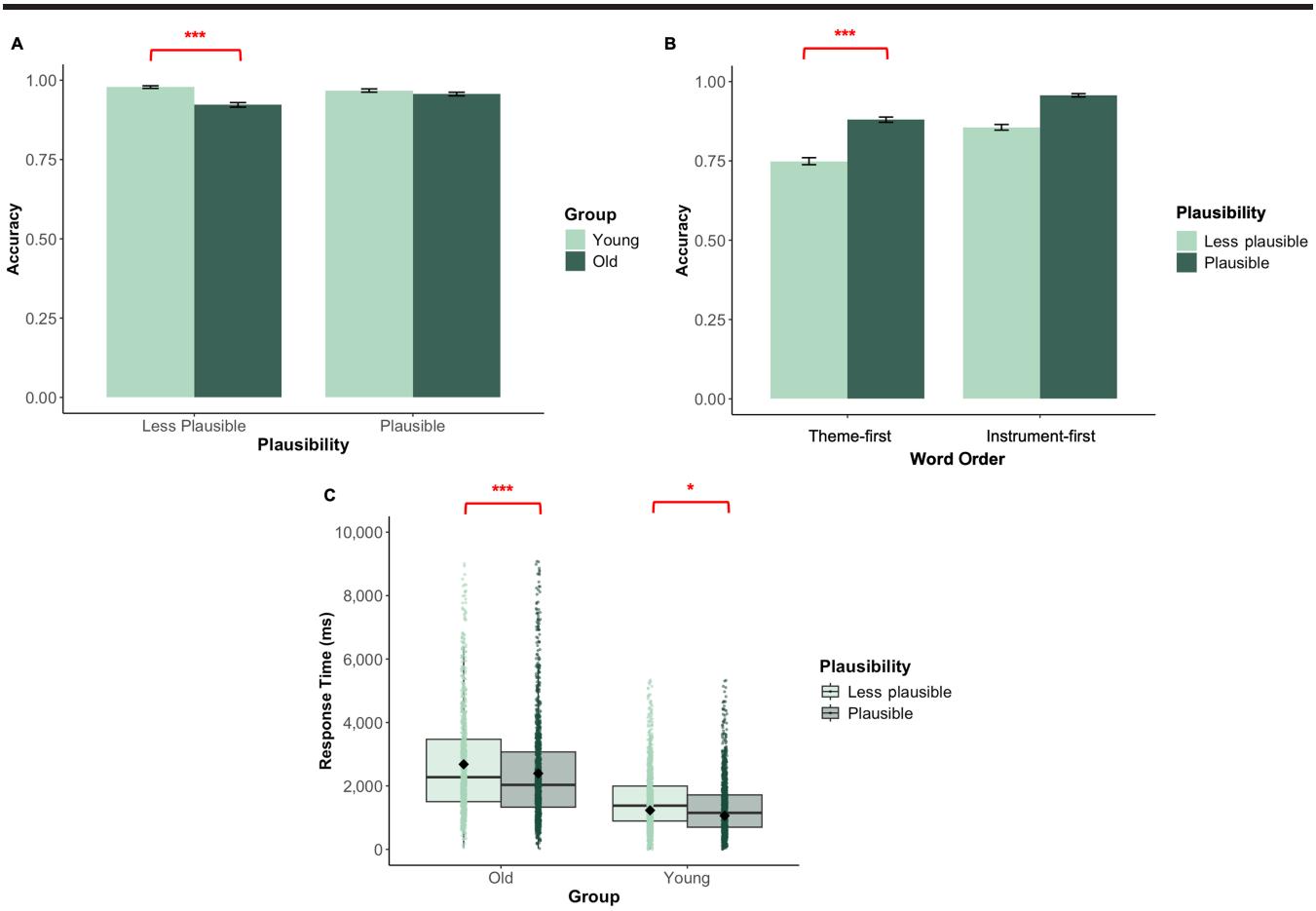


Figure 5. Effects of age group, word order, and plausibility in Japanese speakers. (A) Accuracy. (B) Response time. Error bars represent ±1 standard error.

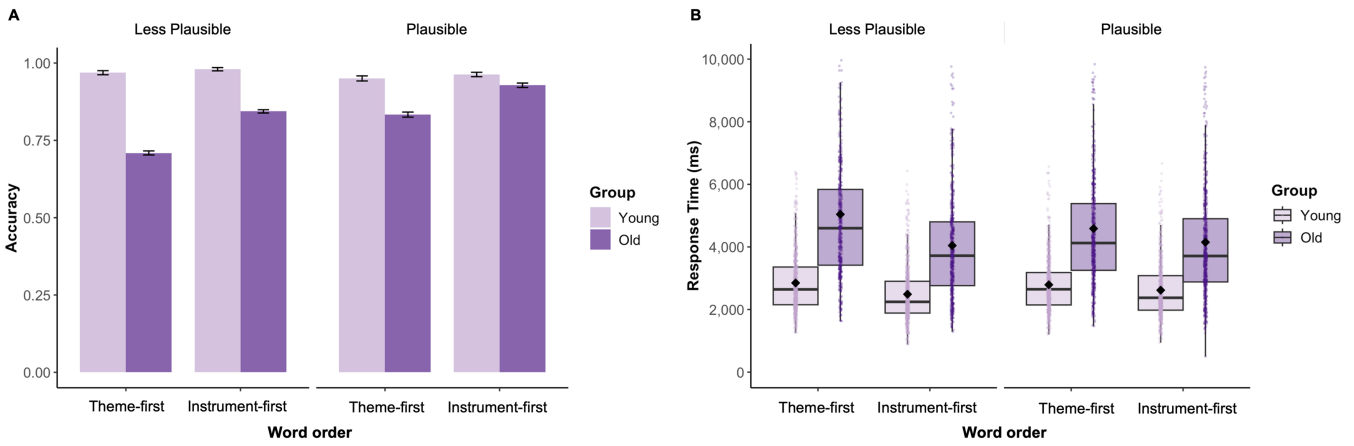


Figure 6. Significant interaction effects in Japanese speakers. (A) Accuracy interaction between age group and plausibility. (B) Response time interaction between age group and plausibility. (C) Response time interaction between age group and word order. Diamonds indicate means, horizontal lines indicate medians, and dots represent individual data points. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

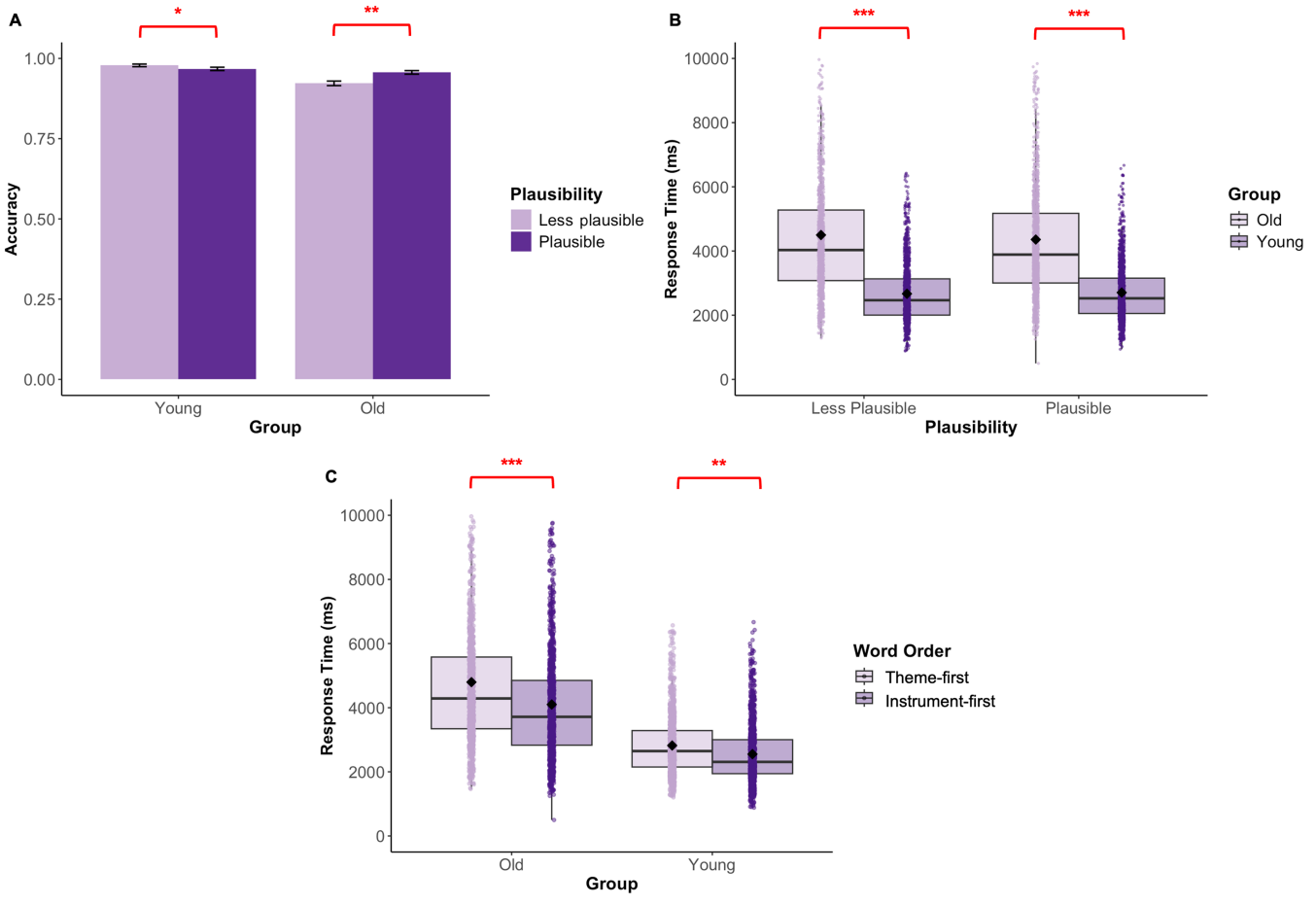


Table 5. Response time of sentences by plausibility, word order, and group for each language.

Plausibility	Word order	English		Korean		Japanese	
		Young	Older	Young	Older	Young	Older
Less plausible	Theme-first (*SD*)	2,687.08 (925.51)	3,793.97 (1,510.47)	4,422.96 (1,194.17)	5,954.25 (1,742.79)	2,849.90 (945.75)	5,041.48 (2,134.20)
	Instrumental-first (*SD*)	2,624.31 (873.08)	3,657.45 (1,282.34)	4,027.77 (1,142.92)	5,446.29 (1,686.37)	2,487.61 (906.29)	4,044.32 (1,815.27)
Plausible	Theme-first (*SD*)	2,714.58 (992.44)	3,566.69 (1,353.30)	4,119.69 (1,123.14)	5,486.68 (1,681.34)	2,791.31 (882.91)	4,586.77 (1,953.18)
	Instrumental-first (*SD*)	2,504.43 (823.71)	3,376.13 (1,162.93)	3,802.62 (988.55)	5,114.43 (1,587.36)	2,617.81 (915.07)	4,150.92 (1,885.59)

Note. Mean values are presented.

Table 6. Summary of the linear mixed-effects models for each language.

Predictors	English					Korean					Japanese				
	β	*SE*	*df*	*t*	*p* value	β	*SE*	*df*	*t*	*p value*	β	*SE*	*df*	*t*	*p value*
(Intercept)	7.980	0.024	135.600	336.673	< .001***	8.710	0.064	102.300	136.347	< .001***	8.579	0.160	86.990	53.750	< .001***
Education (covariate)	—					−0.021	0.005	95.150	−4.411	< .001***	−0.036	0.012	85.080	−3.064	.003**
Group	0.150	0.021	100.500	6.961	< .001***	0.088	0.025	93.720	3.459	< .001***	0.275	0.024	86.160	11.325	< .001***
Plausibility	0.023	0.005	1747.000	4.396	< .001***	0.045	0.012	60.390	3.684	< .001***	0.009	0.016	59.670	0.569	.572
Word order	0.031	0.006	380.600	5.317	< .001***	0.042	0.013	67.160	3.357	< .01**	0.069	0.016	70.580	4.216	< .001***
Group × Plausibility	0.013	0.004	4691.000	3.658	< .001***	0.013	0.003	5028.000	4.824	< .001***	0.021	0.004	4801.000	5.308	< .001***
Group × Word Order	−0.003	0.004	83.000	−0.919	.361	−0.003	0.004	65.150	−0.687	.494	0.014	0.006	73.010	2.234	.029*
Plausibility × Word Order	−0.002	0.005	1587.000	−0.406	.685	0.008	0.012	60.280	0.664	.509	0.025	0.016	59.640	1.619	.111
Group × Plausibility × Word Order	−0.004	0.004	4683.000	−1.056	.291	0.003	0.003	5108.000	0.995	.320	0.006	0.004	4832.000	1.596	.111

Note. Response time data were log-transformed prior to analysis. R model equation: Response Time ~ Age Group × Plausibility × Word Order + (1 + Plausibility × Word Order | Participant) + (1 | Trial). Degrees of freedom were estimated using Satterthwaite's approximation. Em dashes indicate data not obtained or not available. β = fixed-effect estimate; *SE* = standard error.

*$p < .05$. **$p < .01$. ***$p < .001$.

group (β = 0.088, *SE* = 0.025, *df* = 93.7, *t* = 3.46, *p* < .001), with older adults demonstrating significantly longer RT (*M* = 2,515.44, *SD* = 1,671.27) than the young adults (*M* = 1,145.09, *SD* = 1,111.05; see Figure 3B). There was also a significant main effect for plausibility (β = 0.045, *SE* = 0.012, *df* = 60.4, *t* = 3.68, *p* < .001), indicating that less plausible sentences elicited longer RT (*M* = 1,774.09, *SD* = 1,562.67) compared to plausible ones (*M* = 1,660.91, *SD* = 1,496.62). Furthermore, the main effect for word order was statistically significant (β = 0.024, *SE* = 0.015, *df* = 67.2, *t* = 1.54, *p* = .001), with theme-first sentences resulting in longer RT (*M* = 1,804.82, *SD* = 1,563.16) than instrument-first counterparts (*M* = 1,631.47, *SD* = 1,496.62).

A significant two-way interaction was found between age group and plausibility (β = 0.013, *SE* = 0.003, *df* = 50.3, *t* = 4.82, *p* < .001). Post hoc analyses revealed that the difference in RT between less plausible and plausible sentences was significantly greater for older adults (β = 0.116, *SE* = 0.025, *z* = 4.574, *p* < .001) than for young adults (β = 0.064, *SE* = 0.025, *z* = 2.564, *p* = .021; see Figure 4C). However, no significant interactions were found between plausibility and word order (*p* = .509) and age group and word order (*p* = .494) or in the three-way interaction among age group, plausibility, and word order (*p* = .320).

Japanese Group

Japanese speakers' education was included as a covariate, as there was a significant difference in education levels between young and older adults (*p* = .003). The results showed a significant main effect for age group (β = 0.275, *SE* = 0.025, *df* = 86.2, *t* = 11.33, *p* < .001), with older adults showing significantly longer RT (*M* = 4,424.07, *SD* = 1,977.32) than the young adults (*M* = 2,685.85, *SD* = 923.59; see Figure 5B). In addition, the main effect for word order was statistically significant (β = 0.069, *SE* = 0.016, *df* = 70.6, *t* = 4.22, *p* < .001), with theme-first sentences showed longer RT (*M* = 3,704.88, *SD* = 1,818.92) compared to the instrument-first counterparts (*M* = 3,294.87, *SD* = 1,636.17). In contrast, the main effect of plausibility was not significant (*p* = .572).

A significant two-way interaction emerged between age group and plausibility (β = 0.021, *SE* = 0.004, *df* = 48.0, *t* = 5.31, *p* < .001), indicating a greater difference between older and young adults for less plausible sentences (β = 0.591, *SE* = 0.049, *z* = 12.011, *p* < .001) than for plausible sentences (β = 0.507, *SE* = 0.049, *z* = 10.344, *p* < .001; see Figure 6B). Furthermore, the interaction between age group and word order was significant (β = 0.014, *SE* = 0.006, *df* = 73.0, *t* = 2.23, *p* = .029). Follow-up analyses showed that the difference between theme-first and instrument-first sentences was significantly greater for older adults (β = 0.167, *SE* = 0.035, *z* = 4.717, *p* < .001) than for young adults (β = 0.110, *SE* = 0.035, *z* = 3.143, *p* = .003; see Figure 6C). However, the two-way interaction between plausibility and word order (*p* = .110) and a three-way interaction among age group, plausibility, and word order (*p* = .111) were not significant.

Discussion

The current cross-linguistic study investigated how normal aging impacts sentence processing abilities in English, Korean, and Japanese, with a specific focus on the roles of word order and semantic plausibility. Overall, the results revealed clear aging effects on both accuracy and RT across all three languages: Older adults were less accurate and significantly slower than younger adults in processing sentences. The findings suggest that aging effects emerge even when using relatively simple active sentences, provided that linguistic components such as word order and semantic plausibility are systematically manipulated.

The results are somewhat surprising given that previous studies identifying aging differences in sentence processing primarily relied on complex sentences, such as center-embedded structures (Baum, 1993; Caplan et al., 2011; Grossman et al., 2002) or passive constructions (Liu, 2018; Nasiri et al., 2022; Sung et al., 2017, 2020). Even with the simple sentences employed in this study, aging effects consistently emerged across languages, regardless of their typological differences. This indicates that the manipulations of word order and semantic plausibility are sufficiently salient to elicit aging effects across diverse language groups. These findings have significant implications, as they suggest the potential for developing a language-general assessment tool to detect aging effects using the same battery across different languages. By capturing subtle changes in sentence processing that emerge in healthy aging, such a tool could serve as a future screening measure for identifying individuals at increased risk for progression toward mild cognitive impairment and dementia. Importantly, the use of structurally comparable, low-complexity sentence templates makes this approach well suited for application across diverse linguistic and cultural contexts, thereby supporting standardized assessment and cross-population comparisons in both research and clinical settings.

Semantic plausibility played a strong role in sentence comprehension across languages, with older adults showing disproportionate difficulty for sentences involving less plausible instrumental relations. One possible explanation is that older adults intermittently misparse or incompletely encode sentences under increased processing demands, particularly when semantic expectations conflict with surface structure. Age-related declines in

working memory, processing speed, and inhibitory control may make it more effortful to maintain and integrate syntactic and semantic information in real time, increasing susceptibility to transient parsing failures or perceptual noise (Rozas et al., 2008; Verhaeghen, 2018; Zelinski et al., 2011).

Beyond intermittent misparsing, older adults may also experience greater uncertainty about the reliability of the linguistic input, including the possibility that they misheard or misremembered parts of the sentence. Within this context, reliance on semantic plausibility may reflect an adaptive response to listener-side noise, rather than a purely strategic override of syntactic information. These findings are also compatible with rational inference models of sentence processing (Futrell et al., 2020; Gibson et al., 2016), which posit that listeners integrate linguistic input with prior knowledge while accounting for noise in the communicative system. Under this framework, listeners may infer that an implausible utterance reflects an error—either on the part of the speaker or within their own perceptual system—and adjust their interpretation accordingly. For example, given the sentence "Point to the pencil with the book," older adults may infer that the intended meaning was "Point to the book with the pencil," given real-world knowledge about instrument use. Consistent with prior work on risky or predictive processing, older adults may therefore place greater weight on semantic plausibility when syntactic cues are difficult to maintain or integrate (DeDe, 2014; Milburn et al., 2023; Umanath, 2016). While such top-down processing can be efficient and adaptive in many communicative contexts (DeDe & Flax, 2016; van Boxtel & Lawyer, 2021), it may also increase vulnerability to misinterpretation when syntactic and semantic cues diverge. Thus, the observed plausibility effects likely reflect an interaction between resource limitations, perceptual uncertainty, and experience-based expectations, rather than a single explanatory mechanism.

These findings are also consistent with "good-enough" processing accounts of sentence comprehension (Ferreira & Patson, 2007), which propose that comprehenders often rely on heuristic cues derived from semantic plausibility or event structure rather than constructing fully detailed syntactic representations. Under conditions of increased processing demands—such as aging-related reductions in working memory or increased perceptual noise—listeners may rely more heavily on these heuristic cues to derive an interpretation that is sufficient for the task at hand. In the present paradigm, such heuristic processing may have been supported by the event structure implied by the command sentences, allowing listeners to interpret sentences based on plausible instrument–action relations even when syntactic cues were difficult to maintain or integrate.

Although effects of age and semantic plausibility were comparable across the three language groups, the effects of word order differed across languages. Korean and Japanese, both of which permit relatively flexible word order, showed significant main effects of word order, with better performance for instrument-first sentences than for theme-first sentences. This pattern suggests that, even in languages that allow flexible constituent ordering, sentences with less typical word orders can impose additional processing demands during comprehension. Importantly, the statistical analyses revealed a significant interaction between age group and word order for the Japanese participants (significant in RTs and trending in accuracy), indicating that older Japanese adults were more affected by the word order manipulation than younger adults. In contrast, such interactions were not observed in the English or Korean groups. This pattern is consistent with previous studies reporting age-related sensitivity to word order variation in head-final languages (Arii & Lee, 2023; Arii et al., 2022; Sung et al., 2017, 2024; Tamaoka et al., 2003).

Notably, while previous studies primarily investigated subject–object–verb (SOV) structures, our findings replicate word order effects for sentences containing instrumental phrases, extending prior research. In earlier studies using SOV structures, the scrambling involved syntactic case markers such as nominative and accusative markers. Consistently, prior research suggests that Korean and Japanese speakers prefer nominative-first structures (SOV) over accusative-first structures (object–subject–verb), with this preference being more pronounced in aging populations (Arii et al., 2022; Kuroda, 1992; Sung et al., 2017). In contrast, the current study employed theme–instrument structures, which incorporate both syntactic and semantic marking: The theme is marked by the accusative case marker, while the instrument is expressed through an instrumental case-marked phrase. Despite these structural differences, word order saliency effects emerged consistently, aligning with previous findings based on primarily syntactic case-marked NPs. This suggests that word order effects extend beyond purely syntactic markers and apply broadly, even in languages with flexible word order, with these effects becoming more pronounced with aging. These findings underscore the robustness of word order canonicity effects across varying sentence structures and highlight their relevance in understanding age-related changes in sentence processing, even in languages traditionally considered to have flexible word order.

From a theoretical perspective, these cross-linguistic differences are also compatible with memory-based models of sentence processing, particularly dependency locality theory (Gibson, 1998, 2000), which proposes that processing difficulty increases as the distance between syntactically dependent elements grows. In verb-final languages

such as Korean and Japanese, the instrument–theme–verb configuration places the theme NP—the internal argument—closer to the clause-final verb, thereby reducing integration costs during sentence comprehension. In contrast, the theme–instrument–verb configuration increases the linear distance between the theme NP and the verb, creating a longer dependency that may impose greater working memory demands. Under this account, the observed advantage for instrument-first ordering in Korean and Japanese may reflect reduced integration costs resulting from shorter dependency distances between the theme NP and the verb.

There were also significant word order effects for English speakers. Consistent with the patterns observed for Korean and Japanese speakers, sentences with instrument-first order (e.g., "With the pencil, point to the book") elicited better performance, as reflected in higher accuracy and shorter reaction times, compared to sentences with theme-first order (e.g., "Point to the book with the pencil"). One possibility is that this is an artifact of the task. Because there were only two images, listeners could select the correct response as soon as they knew either the instrument or the indirect object. Consequently, the instrument-first condition may have provided a temporal advantage, as the critical information in the initial phrase becomes available earlier (at the third word) than in the theme-first condition (at the fourth word).

Another potential explanation concerns the mapping between linguistic order and the conceptual sequence of the described action. Although the instrument-first order occurs less frequently in English, it places the instrument earlier in the sentence, creating a direct correspondence between the linguistic input and the sequence of actions required for the task (i.e., identifying the instrument before acting on the theme). This alignment between sentence structure and the action sequence may reduce cognitive load by allowing listeners to map the linguistic input more directly onto the action representation. Such an account is consistent with the isomorphic mapping hypothesis (O'Grady & Lee, 2005), which proposes that sentence comprehension is facilitated when the order of thematic roles in a sentence corresponds to the natural temporal sequence of events in real-world actions. In the present task, the instrument-first configuration preserves this event-based sequencing. For example, in a sentence such as "With the pencil, point to the book," the implied sequence of actions—identifying or grasping the pencil and then pointing to the book—aligns with how the action would naturally be executed. This correspondence between linguistic structure and event sequencing may therefore facilitate comprehension, providing a possible explanation for why English speakers showed better performance in the instrument-first condition.

Within a good-enough processing framework, such isomorphic alignment can be understood as a semantic–event cue that supports comprehension when syntactic analysis is incomplete or resource-limited. In other words, even though instrument-first order represents a relatively marked linear configuration in English, its alignment with the natural sequence of the described action may provide a salient heuristic that facilitates interpretation, particularly for older adults who may rely more strongly on semantic or event-based cues during sentence processing.

By contrast, in Korean and Japanese, the instrument-first structure is canonical and aligns with the natural sequence of events. Because word order and event sequencing converge in these languages, it is difficult to distinguish whether processing is facilitated by syntactic familiarity or the temporal alignment of actions. To better disentangle these contributions, future studies could examine English constructions that vary in mapping, such as "Point with the pen to the book" versus "Point to the book with the pen." Additionally, including a greater number of foil options is essential to rule out heuristic processing strategies and confirm these findings.

Summarizing the findings across all three language groups, when syntactic and semantic cues align (as in Korean and Japanese), participants clearly benefited from this convergence, demonstrating enhanced comprehension and processing efficiency. However, when semantic sequencing and word order canonicity conflict (as in English), participants appeared to prioritize the semantic sequence to interpret and execute sentential commands. This suggests that while syntactic structure plays a crucial role in comprehension, semantic predictability and real-world event mapping can override syntactic expectations, particularly when conflicts arise. Overall, these findings highlight the complex interplay between syntactic processing, event-based semantic mapping, and language-specific parsing strategies, reinforcing the idea that aging and sentence comprehension are influenced not only by structural complexity but also by the extent to which word order aligns with real-world event structures.

Across the three language groups, semantic plausibility and word order differed in how they related to aging effects in sentence comprehension. Robust Age Group × Semantic Plausibility interactions were observed in both accuracy and reaction time, indicating that semantic-level processing demands consistently differentiated younger and older adults. In contrast, word order effects were primarily observed as main effects across languages, influencing performance in both younger and older adults. An exception was observed in the Japanese group, where a significant Age Group × Word Order interaction emerged in RTs (with a marginal trend in accuracy), suggesting that older Japanese adults were more sensitive to the word order manipulation than younger adults.

Overall, these findings suggest that different linguistic manipulations may differentially capture age-related changes in sentence comprehension. One possible explanation is that word order serves distinct functional roles across linguistic typologies, particularly between head-final and head-initial languages, which may limit its usefulness as a uniform indicator of aging effects. By contrast, semantic plausibility appears to function as a more general constraint on sentence interpretation, exerting comparable influences across typologically diverse languages and more reliably revealing age-related differences in the present paradigm.

Taken together, the present findings underscore the importance of distinguishing between language-general and language-specific psycholinguistic variables when examining aging effects. They further highlight the value of developing standardized experimental paradigms based on simple sentence structures that are applicable across typologically diverse languages. Such paradigms are essential for isolating robust linguistic markers of aging and enabling meaningful cross-linguistic comparisons. The development of the current paradigm was inspired by the sequential commands of the WAB, which provide valuable diagnostic information despite their simplicity. However, the WAB contains insufficient items to systematically scrutinize specific psycholinguistic properties such as word order and plausibility. Building on its recent application to individuals with aphasia (Sung et al., 2024), we anticipate that this paradigm will serve as a valuable tool for assessing diverse clinical populations across different languages, thereby facilitating robust cross-linguistic comparisons in both research and clinical settings.

Conclusions

The present study represents the first systematic attempt to develop and apply a common sentence comprehension paradigm across diverse languages with distinct linguistic typologies—English, Korean, and Japanese—to investigate the effects of aging on sentence processing. By focusing on the interplay between word order and semantic plausibility, this research revealed robust aging effects across all three languages, with older adults demonstrating lower accuracy and slower RTs compared to younger adults. These findings are particularly noteworthy because they emerged even with simple active sentences, challenging the prevailing assumption that aging effects are primarily observable in complex syntactic structures, such as center-embedded or passive constructions. The consistent aging effects observed across languages, despite their typological differences, underscore the salience of word order and semantic plausibility manipulations in eliciting age-related processing difficulties. This suggests the potential for developing a language-general assessment tool to detect cognitive aging, offering a standardized approach that can be applied across diverse linguistic and cultural contexts.

Author Contributions

Jee Eun Sung: Conceptualization, Methodology, Supervision, Writing – original draft. **Gayle DeDe:** Conceptualization, Formal analysis, Writing – review & editing. **Seunghun J. Lee:** Conceptualization, Formal analysis, Writing – review & editing. **Jimin Park:** Formal analysis, Writing – original draft.

Data Availability Statement

Further information can be requested from the corresponding author, subject to institutional and ethical approval.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (RS-2022-NR070151, RS-2024-00461617, NRF-2026S1A5A2A03006789) and by the Ewha Global Excellence Program (EGEP) awarded to Jee Eun Sung, and Kakenhi(C) 26K03907 awarded to Seunghun J. Lee. The authors would like to thank the study participants for their time and contribution to this research.

References

Arii, T., & Lee, S. J. (2023). The impact of information structure on ditransitive alternation in Japanese. *The Bulletin of Yokohama City University, Humanities, 75*(1), 115–131.

Arii, T., Lee, S. J., Yoshida, T., & Sung, J. E. (2022). Effects of age, word order, and sentence types on Japanese sentence comprehension: A replication study of Sung et al. (2017) and Sung (2015) on Korean. *Educational Studies,* (64), 37–49.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Baum, S. R. (1993). Processing of center-embedded and right-branching relative clause sentences by normal elderly individuals. *Applied Psycholinguistics, 14*(1), 75–88. https://doi.org/10.1017/S0142716400010158

Beier, E. J., Chantavarin, S., & Ferreira, F. (2023). Do disfluencies increase with age? Evidence from a sequential corpus study of disfluencies. *Psychology and Aging, 38*(3), 203–218. https://doi.org/10.1037/pag0000741

Biber, D., Johansson, S., Finegan, E., Conrad, S., & Leech, G. N. (2021). Word order and related syntactic choices. In *Grammar*

of spoken and written English. John Benjamins. https://doi.org/10.1075/z.232.c11

Caplan, D., Dede, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging, 26*(2), 439–450. https://doi.org/10.1037/a0021837

Caplan, D., & Waters, G. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory, 13*(3–4), 403–413. https://doi.org/10.1080/09658210344000459

Chengappa, S. K., & Kumar, R. (2008). Normative & clinical data on the Kannada Version of Western Aphasia Battery (WAB-K). *Language in India, 8*(6).

Christensen, K. J., Multhaup, K. S., Nordstrom, S., & Voss, K. (1991). A cognitive battery for dementia: Development and measurement characteristics. *Psychological Assessment, 3*(2), Article 168. https://doi.org/10.1037//1040-3590.3.2.168

DeDe, G. (2014). Sentence comprehension in older adults: Evidence for risky processing strategies. *Experimental Aging Research, 40*(4), 436–454. https://doi.org/10.1080/0361073X.2014.926775

DeDe, G., & Flax, J. K. (2016). Language comprehension in aging. In H. H. Wright (Ed.), *Cognition, language and aging* (pp. 107–133). John Benjamins.

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' approach to language comprehension. *Language and Linguistics Compass, 1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science, 44*(3), Article e12814. https://doi.org/10.1111/cogs.12814

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. P. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the First Mind Articulation Project Symposium* (Vol. 2000, pp. 95–126). MIT Press. https://doi.org/10.7551/mitpress/3654.003.0008

Gibson, E., Sandberg, C., Fedorenko, E., Bergen, L., & Kiran, S. (2016). A rational inference approach to aphasic language comprehension. *Aphasiology, 30*(11), 1341–1360. https://doi.org/10.1080/02687038.2015.1111994

Grossman, M., Cooke, A., DeVita, C., Alsop, D., Detre, J., Chen, W., & Gee, J. (2002). Age-related changes in working memory during sentence comprehension: An fMRI study. *NeuroImage, 15*(2), 302–317. https://doi.org/10.1006/nimg.2001.0971

Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging, 25*(3), 708–713. https://doi.org/10.1037/a0019424

Kang, Y. (2006). A normative study of the Korean-Mini Mental State Examination(K-MMSE) in the elderly. *Korean Journal of Psychology, 25*(2), 1–12. http://kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART001153430

Kertesz, A. (2007). *Western Aphasia Battery–Revised (WAB-R)*. Pearson. https://doi.org/10.1037/t15168-000

Kertesz, A. (2022). The Western Aphasia Battery: A systematic review of research and clinical applications. *Aphasiology, 36*(1), 21–50. https://doi.org/10.1080/02687038.2020.1852002

Kim, H., & Na, D. L. (2012). *Paradise· Korean Version of the Western Aphasia Battery Revised (PK-WAB-R)*. Paradise Welfare Foundation.

Kuroda, S.-Y. (1992). Case marking, canonical sentence patterns, and counter Equi in Japanese (A preliminary survey). In *Japanese syntax and semantics: Collected papers* (pp. 222–239). Springer Netherlands. https://doi.org/10.1007/978-94-011-2789-9_7

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lazar, R. M., Speizer, A. E., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2008). Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery & Psychiatry, 79*(5), 530–534. https://doi.org/10.1136/jnnp.2007.122457

Liu, X. (2018). Effects of age and working memory load on the comprehension of passive sentences. *International Journal of Psychological Studies, 10*(3), 13–20. https://doi.org/10.5539/ijps.v10n3p13

Lwi, S. J., Herron, T. J., Curran, B. C., Ivanova, M. V., Schendel, K., Dronkers, N. F., & Baldo, J. V. (2021). Auditory comprehension deficits in post-stroke aphasia: Neurologic and demographic correlates of outcome and recovery. *Frontiers in Neurology, 12,* Article 680248. https://doi.org/10.3389/fneur.2021.680248

Milburn, E., Dickey, M. W., Warren, T., & Hayes, R. (2023). Increased reliance on world knowledge during language comprehension in healthy aging: Evidence from verb-argument prediction. *Aging, Neuropsychology, and Cognition, 30*(1), 1–33. https://doi.org/10.1080/13825585.2021.1962791

Murphy, E. (2025). Time slows down in the future: Aging and the brain rhythms of language. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.70019

Nasiri, M., Moayedfar, S., Purmohammad, M., & Ghasisin, L. (2022). Investigating sentence processing and working memory in patients with mild Alzheimer and elderly people. *PLOS ONE, 17*(11), Article e0266552. https://doi.org/10.1371/journal.pone.0266552

Neves, M. d. B., Borsel, J. V., Pereira, M. M. d. B., & Paradela, E. M. P. (2014). Cross-cultural adaptation of the Western Aphasia Battery - Revised screening test to Brazilian Portuguese: A preliminary study. *CoDAS, 26*(1), 38–45. https://doi.org/10.1590/S2317-17822014000100006

O'Grady, W., & Lee, M. (2005). A mapping theory of agrammatic comprehension deficits. *Brain and Language, 92*(1), 91–100. https://doi.org/10.1016/j.bandl.2004.05.009

Poulisse, C., Wheeldon, L., & Segaert, K. (2019). Evidence against preserved syntactic comprehension in healthy aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(12), 2290–2308. https://doi.org/10.1037/xlm0000707

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing.

Reifegerste, J., Hauer, F., & Felser, C. (2017). Agreement processing and attraction errors in aging: Evidence from subject-verb agreement in German. *Aging, Neuropsychology, and Cognition, 24*(6), 672–702. https://doi.org/10.1080/13825585.2016.1251550

Rozas, A. X. P., Juncos-Rabadán, O., & González, M. S. R. (2008). Processing speed, inhibitory control, and working memory: Three important factors to account for age-related cognitive decline. *The International Journal of Aging and Human Development, 66*(2), 115–130. https://doi.org/10.2190/AG.66.2.b

Sung, J. E., Choi, S., Eom, B., Yoo, J. K., & Jeong, J. H. (2020). Syntactic complexity as a linguistic marker to differentiate mild cognitive impairment from normal aging. *Journal of Speech, Language, and Hearing Research, 63*(5), 1416–1429. https://doi.org/10.1044/2020_JSLHR-19-00335

Sung, J. E., DeDe, G., & Park, J. (2024). Interplay of semantic plausibility and word order canonicity in sentence processing of people with aphasia using a verb-final language. *American Journal of Speech-Language Pathology, 33*(6S), 3236–3246. https://doi.org/10.1044/2024_AJSLP-23-00353

Sung, J. E., Yoo, J. K., Lee, S. E., & Eom, B. (2017). Effects of age, working memory, and word order on passive-sentence comprehension: Evidence from a verb-final language. *International Psychogeriatrics, 29*(6), 939–948. https://doi.org/10.1017/S1041610217000047

Tamaoka, K., Sakai, H., Kawahara, J.-i., & Miyaoka, Y. (2003). The effects of phrase-length order and scrambling in the processing of visually presented Japanese sentences. *Journal of Psycholinguistic Research, 32*(4), 431–454. https://doi.org/10.1023/A:1024851729985

Tukey, J. W. (1977). *Exploratory data analysis.* Addison-Wesley.

Umanath, S. (2016). Age differences in suggestibility to contradictions of demonstrated knowledge: The influence of prior knowledge. *Aging, Neuropsychology, and Cognition, 23*(6), 744–767. https://doi.org/10.1080/13825585.2016.1167161

van Boxtel, W., & Lawyer, L. (2021). Sentence comprehension in ageing and Alzheimer's disease. *Language and Linguistics Compass, 15*(6), Article e12430. https://doi.org/10.1111/lnc3.12430

Verhaeghen, P. (2018). Working memory and cognitive aging. In F. C. Worrell (Ed.), *Oxford research encyclopedia of psychology*. Oxford Academic. https://doi.org/10.1093/ACREFORE/9780190236557.013.382

Wingfield, A., Amichetti, N. M., & Lash, A. (2015). Cognitive aging and hearing acuity: Modeling spoken language comprehension. *Frontiers in Psychology, 6,* Article 684. https://doi.org/10.3389/fpsyg.2015.00684

Zelinski, E., Dalton, S., & Hindin, S. (2011). Cognitive changes in healthy older adults. *Generations, 35*(2), 13–20.