ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Cross-Linguistic and Cross-Cultural Effects on Picture Description: A Main Concept Analysis in Korean and English Speakers With and Without Aphasia

Yae Rin Yoo,[a] Jimin Park,[a] and Jee Eun Sung[a]

[a] Department of Communication Disorders, Ewha Womans University, Seoul, South Korea

ARTICLE INFO

Article History:
Received September 15, 2025
Revision received December 2, 2025
Accepted January 27, 2026

Editor-in-Chief: Catherine A. Off
Editor: Brielle C. Stark

https://doi.org/10.1044/2026_AJSLP-25-00422

ABSTRACT

Purpose: This study examined cross-linguistic and cross-cultural influences on picture-based narrative performance in Korean and English speakers with and without aphasia. We evaluated whether pro-drop–adjusted main concept (MC) scoring accounts for linguistic effects, identified picture-driven cultural effects on discourse, and examined associations between aphasia severity and performance.
Method: Narratives were elicited with two AphasiaBank stimuli—*Cat Rescue* and *Refused Umbrella*—from 54 neurotypical adults (27 Korean neurotypical adults [K-NA], 27 English neurotypical adults [E-NA]) and 34 persons with aphasia (17 Korean persons with aphasia [K-PWA], 17 English persons with aphasia [E-PWA]). In addition to the original MC scoring criteria, we adapted the framework by developing a pro-drop adjustment that credited omitted but contextually recoverable arguments. Language and picture effects were examined under both MC criteria, whereas clinical group effects were evaluated using the adjusted criteria. Additional analyses identified MCs distinguishing language groups and examined associations with Aphasia Quotient (AQ).
Results: With original scoring, neurotypicals showed language (K-NA < E-NA) and picture effects (*Cat Rescue* < *Refused Umbrella*). After pro-drop adjustment, group differences were observed only for *Cat Rescue* and persisted when PWA were included. The main storyline MC item of *Cat Rescue* distinguished language groups in both NA and PWA. Picture-specific MC scores differentially predicted AQ: *Cat Rescue* in K-PWA and *Refused Umbrella* in E-PWA.
Conclusions: The findings highlight picture-salient cultural effects in *Cat Rescue*, even after linguistic adjustments. Limited familiarity with the scene likely hindered Korean speakers' ability to generate core narrative elements. Both linguistic and cultural factors should be considered when selecting picture stimuli for discourse assessment across diverse contexts.
Supplemental Material: https://doi.org/10.23641/asha.31974204

Picture description tasks are widely used to elicit connected speech in persons with aphasia (PWA; Goodglass & Kaplan, 1983; Gordon, 2008; Hameister & Nickels, 2018; Schönberger et al., 2014), as well as in individuals with other neurogenic communication disorders such as traumatic brain injury, mild cognitive impairment, and dementia (Forbes-McKay & Venneri, 2005; Fraser et al., 2016; Giles et al., 1996; Mueller et al., 2018; Shimada et al., 1998). As discourse-level assessments, these tasks contribute to clinical diagnosis, assessment, intervention planning, and progress monitoring. Among the various picture stimuli employed as connected speech sampling procedures, several have been incorporated into standardized assessment protocols and are commonly utilized in clinical and research contexts. For instance, the *Cookie Theft* picture is included in the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1972) and the DementiaBank protocol (MacWhinney et al., 2011). The *Picnic Scene* appears in the Western Aphasia

Correspondence to Jee Eun Sung: jeesung@ewha.ac.kr. **Publisher Note:** This article is part of the Special Issue: Select Papers From the 54th Clinical Aphasiology Conference. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Battery–Revised (WAB-R; Kertesz, 2007), while both the *Cat Rescue* (Nicholas & Brookshire, 1993) and *Refused Umbrella* pictures are part of the AphasiaBank protocol (MacWhinney et al., 2011).

However, recent studies have raised concerns that culturally specific content embedded in picture stimuli may limit the ecological validity and generalizability of discourse assessment outcomes. Since most of these pictures were originally developed in Western, monolingual, English-speaking contexts, their applicability to linguistically and culturally diverse populations has been increasingly questioned (Marcotte et al., 2024; Steinberg et al., 2022; Stockbridge et al., 2024). For example, a group of researchers pointed out that the *Cookie Theft* picture contains embedded racial, gendered, cultural, and socioeconomic biases (Poisson et al., 2022; Steinberg et al., 2022). They argued that a scene of suburban mid-century American life—depicting a dress-clad woman gazing from her kitchen window while children attempt to steal cookies from a jar on a high shelf—reflects a specific sociocultural context that may be less effective in eliciting language from individuals with different backgrounds. Although Berube et al. (2019) updated the *Cookie Theft* picture to better reflect contemporary life by reducing potential biases (e.g., depicting a man washing dishes and a woman mowing over flowers) and to incorporate more dynamic, action-based elements (e.g., a dog eating cookies from the floor, a cat chasing birds on the lawn), this modernization still represents a within-language conceptual and cultural adaptation rather than a cross-cultural one. Consequently, those pictures may elicit divergent interpretations and narrative productions depending on an individual's cultural background, underscoring the need to consider the cultural salience—and potential bias—of picture stimuli in discourse assessment.

Despite increasing concerns regarding the cultural salience of picture-based discourse tasks, many standardized assessment protocols still rely on identical picture stimuli across diverse linguistic and cultural contexts (Evans et al., 2022; Santi et al., 2024). For example, the WAB-R has preserved the identical *Picnic Scene* across diverse linguistic adaptations—such as Persian (Nilipour et al., 2014), Mandarin (Wang, 1997), Cantonese (Yiu et al., 1998), Brazilian Portuguese (Neves et al., 2014), Cairene Egyptian Arabic (Jomaa et al., 2025), and Moroccan Arabic (El Ouardi et al., 2023)—without modification to the visual stimulus. Similarly, TalkBank (MacWhinney, 2007) provides large-scale, publicly available databases for the study of communication disorders across multiple languages, including DementiaBank (Becker et al., 1994) and AphasiaBank (MacWhinney et al., 2011). The DementiaBank includes narrative samples elicited with the *Cookie Theft* picture in multiple languages, such as English, German, Greek, Mandarin, Spanish, and Taiwanese. The AphasiaBank provides standardized protocols for the *Cat Rescue* (Nicholas & Brookshire, 1993) and *Refused Umbrella* (MacWhinney et al., 2011) pictures translated into multiple languages, including English, Croatian, French, Italian, Mandarin, Romanian, and Spanish. It also offers corresponding narrative corpora that have been collected and transcribed in a wide range of languages, such as Cantonese, Croatian, English, French, German, Greek, Italian, Japanese, Mandarin, Polish, and Romanian. Although many identical picture stimuli are often used to build cross-linguistic corpora, relatively few studies have directly investigated the picture-driven cultural effects on the discourse performance of clinical populations.

Even within the same language, cultural and ethnic backgrounds have been shown to affect discourse patterns. For instance, Ramkissoon et al. (2013) compared education-matched African American and Caucasian American adults on a picture description task using ethnically diverse photographic stimuli. They found that each group showed a statistically significant preference for photographs depicting models of their own ethnicity. These differences were accompanied by measurable variation in discourse performance, demonstrating that Caucasian American participants produced significantly higher word counts and more content units across picture scenes, whereas African American group often generated fewer content units. Participant reports further suggested that these differences were grounded in culturally shaped expectations about everyday events; African American adults noted that the small two-person picnic depicted in the stimulus set was unrealistic for their community and that they would be more likely to favor large party-style social gatherings. In terms of discourse style, Caucasian Americans adopted topic-centered patterns more often, whereas African American participants produced topic-associating discourse. Together, these findings illustrate that cultural and ethnic backgrounds could influence performance on picture description tasks, even within a single language.

While findings from within-language comparisons (e.g., Ramkissoon et al., 2013) have demonstrated cultural and ethnic influences on discourse performance, these effects are likely to be more salient in cross-linguistic contexts driven by typological differences. In this regard, J. E. Sung et al. (2016) reported distinct discourse patterns between Korean- and English-speaking individuals with and without aphasia using an identical *Cat Rescue* picture from the AphasiaBank. Examining the lexical distribution of verbs and nouns, they found that English-speaking neurotypical participants produced significantly more utterances and nouns than Korean speakers. However, Korean neurotypical speakers generated significantly more verbs per utterance, yielding a significantly lower noun-to-verb ratio than English speakers. These findings likely reflect

typological specificities of the Korean language that Korean is a verb-salient and pro-drop language, where verbs are rarely omitted but noun phrases are more readily dropped. Korean is often described as a discourse-oriented language, where many arguments in a sentence can be inferred from context without being explicitly realized with morphological case marking (i.e., case particles such as –ka/–i for subject or –lul/–ul for object; Y.-J. Kim, 2000; Kwon & Sturt, 2013). As a result, verbs in Korean can function as complete sentences, with arguments omitted when recoverable from the surrounding discourse context. In shared contexts, Korean speakers frequently omit elements positioned before the verb, such as subjects or objects, thereby reinforcing the verb-salient nature of the language (Choi, 2000). However, it remains unclear whether these differences were driven by cross-linguistic factors, rooted in structural and typological properties of the language, or by the cultural salience of the picture, reflecting sociocultural knowledge, expectations, and familiarity with the depicted events.

Beyond the context-based linguistic features of Korean, picture-driven cultural factors may also contribute to the lower number of utterances produced by Korean speakers. As Western culture–centered objects—such as the cookie jar in the *Cookie Theft* picture—may not carry the same familiarity or interpretations across cultures, introducing cultural bias (Poisson et al., 2022; Steinberg et al., 2022), a similar issue might also arise with the *Cat Rescue* picture. The scenario of the picture—firefighters rescuing a cat from a tree— may be less salient for Korean participants, as such situations are not typically regarded as emergencies in Korean cultural contexts. Although fire departments in Korea do respond to animal-related cases, such efforts are usually limited to instances involving threats to human safety or property, such as capturing dangerous or protected animals (W.-J. Kim, 2023). Moreover, H. B. Sung (2021) noted that, compared with the United States, Korea lacks established laws and procedures for animal rescue, even during disaster situations. These cultural differences may lower the familiarity of the *Cat Rescue* picture for Korean speakers, hindering their ability to grasp the intended main storyline. Although previous studies have examined cross-linguistic differences using this stimulus (e.g., J. E. Sung et al., 2016), they relied on a single picture and lacked a control task, making it difficult to disentangle picture-specific cultural effects from linguistic influences.

To address this gap, the present study aims to investigate linguistic and cultural influences on picture-based narrative performance and examine how these factors interact across different populations. Specifically, we compare the discourse performance of Korean- and English-speaking individuals, both with and without aphasia, using two widely adopted pictures from the AphasiaBank protocol: *Cat Rescue* (Nicholas & Brookshire, 1993) and *Refused Umbrella* (MacWhinney et al., 2011). We assumed that *Cat Rescue*, as a single-image stimulus with a culturally specific context, may be less familiar and less easily interpretable for Koreans, where calling firefighters to rescue a pet is not commonly perceived as an emergency (W.-J. Kim, 2023). In contrast, *Refused Umbrella* is assumed to be a less culture-specific situation than the *Cat Rescue* picture, as it presents a sequential set of images that make the storyline easier to follow and depict a socially neutral, everyday event.

To examine the cultural influence of picture salience on discourse production in greater depth, we employed main concept analysis (MCA; Nicholas & Brookshire, 1995)—a method developed to assess lexical–semantic performance in narrative discourse. Main concepts (MCs) are defined as utterances that convey the essential semantic content expected when describing a given scene or narrative (Nicholas & Brookshire, 1995). MCA enables researchers to analyze how participants focus on and express the key informational elements within picture-based discourse tasks (Rivera et al., 2018). According to the *Thinking for Speaking* framework (Slobin, 1987), speakers conceptualize events in ways that are shaped by the grammatical and lexical resources of their native language. In this view, language guides attention to specific aspects of experience that are most readily encoded, resulting in typologically distinct patterns of event conceptualization and linguistic expression. The rationale is also consistent with cognitive linguistic perspectives, highlighting how language typology shapes conceptualization and linguistic encoding of event structure and informational salience across languages (e.g., Croft, 2000; Kuzmina et al., 2019; J. E. Sung et al., 2024; Talmy, 2000). Such typological differences have direct implications for picture-based discourse: Speakers of different languages may consider different components of the same scene as the core elements of an event, influencing which MCs are produced and which are omitted.

Importantly, MC scoring incorporates both the semantic content and syntactic structure of utterances. Each MC item includes a subject and a verb, collectively regarded as the essential information for sentence production, and full credit is awarded only when these core elements are explicitly expressed. However, the original MC scoring system presents challenges when applied to verb-salient, pro-drop languages such as Korean, where subjects are frequently omitted if the referent is clear from the context (Sohn, 2001; J. E. Sung et al., 2016). For instance, in describing the man in the tree in the *Cat Rescue*, a Korean speaker might simply say "*climbed up (ollagatta)*" without explicitly stating the subject. Under the original MC scoring criteria, such responses may be penalized,

reflecting not an actual discourse deficit but linguistic differences in syntactic expression. To address this issue, we developed a pro-drop–adjusted MC scoring method that incorporates language-specific features of Korean, thereby enhancing not only the linguistic validity of the assessment but also adherence to the cross-linguistic equivalence principle in cross-language comparisons (J. E. Sung et al., 2016, 2024). Therefore, examining MCs offers a principled methodological approach for investigating how speakers conceptualize and linguistically encode essential event components, capturing both typologically driven cross-linguistic differences (e.g., pro-drop and verb-salient structures) and cross-cultural differences shaped by socio-pragmatic expectations and experiential familiarity with the depicted scene.

Based on these speculations, the present study aims to investigate how linguistic and cultural factors influence picture-based narrative production across language groups and clinical populations. (a) We first examine whether cross-linguistic differences emerge between English- and Korean-speaking neurotypical adults (NA) in their discourse performance across two picture stimuli—*Cat Rescue* and *Refused Umbrella*—using the original MC scoring method. (b) We then apply the pro-drop–adjusted MC scoring method to investigate whether incorporating language-specific features affects the discourse performance between English- and Korean-speaking NA across the two pictures. (c) Incorporating cross-linguistic features into the analysis, we examine whether picture-driven cultural differences emerge across English- and Korean-speaking adults with and without aphasia, depending on picture type (*Cat Rescue* vs. *Refused Umbrella*), using the pro-drop–adjusted MC scoring method. (d) To further explore the potential source of cross-cultural differences, we examined which specific MCs in the *Cat Rescue* picture differentiated Korean and English speakers, separately within neurotypical and aphasic groups. (e) Lastly, to identify the clinical relevance of narrative performance, we examine whether picture-specific discourse performance (*Cat Rescue* and *Refused Umbrella*) was associated with aphasia severity, as measured by the Aphasia Quotient (AQ), within each language group.

Method

Participants

A total of 88 participants were included in the study, comprising 54 NA (27 English-speaking and 27 Korean-speaking) and 34 PWA (17 English-speaking and 17 Korean-speaking). Monolingual English participants, with and without aphasia, were drawn from the AphasiaBank protocol (MacWhinney et al., 2011), while Korean monolinguals were recruited from local communities in South Korea. All participants with aphasia had experienced a single left hemisphere stroke and were premorbidly right-handed. English PWA were diagnosed with WAB-R (Kertesz, 2007) and Korean PWA with Paradise Korean version of the WAB-R (PK-WAB-R; H. Kim & Na, 2012). The two PWA groups were matched for age, $F(1, 32) = 0.576$, $p = .453$; education, $F(1, 32) = 1.386$, $p = .248$; and aphasia severity, $F(1, 32) = 0.012$, $p = .912$, measured from the AQ from each language version of the WAB.

The English and Korean NA also did not significantly differ in age, $F(1, 52) = 0.003$, $p = .956$, or years of education, $F(1, 52) = 0.045$, $p = .832$. Cognitive screening was conducted using the Mini-Mental State Examination (MMSE; Folstein et al., 1975) for English NA and the Korean version of the MMSE (Kang, 2006) for Korean NA. Although MMSE scores differed significantly between the groups ($p < .001$)—with no English NA scoring below 28 out of 30, while eight Korean NA scored below 28—all Korean scores still fell within the normal range according to age- and education-adjusted norms (> 16th percentile). Demographic information for all groups is summarized in Table 1, and clinical data for PWA groups are presented in Supplemental Materials S1 and S2. This study was approved by the Institutional Review Board (IRB) on Human Subjects at Ewha Womans University (IRB No. 2022-0140), and informed consent was obtained from all Korean participants.

Stimuli

Transcripts from two discourse tasks—a single-picture description (*Cat Rescue*) and a sequential-picture narrative (*Refused Umbrella*)—were obtained from the AphasiaBank database (MacWhinney et al., 2011) for 44 English speakers, comprising 27 English NA and 17 English PWA. For Korean participants, both tasks were administered using translated task instructions from the AphasiaBank, and connected speech samples were collected from 27 Korean NA and 17 Korean PWA, with all responses transcribed for analysis.

MCA

Original MC Scoring Method

The MC scoring approach follows the composite framework developed by Richardson and Dalton (2016, 2020), as outlined in the AphasiaBank MC protocol (https://talkbank.org/aphasia/discourse/MainConcepts/index.html). Each picture consisted of 10 MCs, representing essential narrative elements typically comprising a subject and a main verb (Nicholas & Brookshire, 1995; Richardson & Dalton, 2016, 2020).

Table 1. Demographic information for English- and Korean-speaking neurotypicals and persons with aphasia.

Variable	Neurotypical		PWA		Statistical results
	English (n = 27)	Korean (n = 27)	English (n = 17)	Korean (n = 17)	
Age (years)					$F(3, 84) = 0.244$, $p = .865$
M (*SD*)	50.04 (18.02)	50.33 (21.10)	54.36 (13.22)	50.71 (14.79)	
Range	21–77.3	20–77	26–75.7	2–71	
Education (years)					$F(3, 84) = 1.186$, $p = .320$
M (*SD*)	14.00 (1.49)	13.89 (2.26)	13.91 (2.05)	12.77 (3.46)	
Range	12–16	9–16	11–17	6–16	
MMSE					$F(1, 49) = 13.486$, $p < .001$***
M (*SD*)	29.29 (0.86)	28.15 (1.29)			
Range	28–30	25–30			
WAB-R AQ					$F(1, 34) = 0.008$, $p = .931$
M (*SD*)			74.29 (13.88)	74.80 (13.03)	
Range			42.7–96.0	52.4–94.3	

Note. Age and education were compared across all four groups using one-way analysis of variance. MMSE comparisons were conducted between language groups within the neurotypical participants, whereas AQ comparisons were made between language groups within the PWA group. Although MMSE differences reached significance, all Korean participants scored within the normal range based on age- and education-adjusted criteria (i.e., > 16th percentile). PWA = persons with aphasia; MMSE = Mini-Mental State Examination; WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient.

***$p < .001$.

Evaluation of each MC utilized a five-level coding system, incorporating both accuracy and completeness of a sentence. A score of 3 was assigned when all essential elements were accurate and complete (AC). A score of 2 was used for responses that were either accurate but incomplete (AI) or inaccurate but complete (IC). A score of 1 was given when responses were both inaccurate and incomplete (II), and a score of 0 was assigned when all essential elements were absent. MC scores were calculated using the following formula: MC score = (3 × AC) + (2 × AI) + (2 × IC) + (1 × II). Scores across all 10 MCs were summed for each task, yielding a maximum score of 30 per picture.

Pro-Drop–Adjusted MC Scoring Criteria

We adapted the original MC scoring framework by incorporating a subject or object omission adjustment to account for the pro-drop nature of Korean. Since Korean permits the omission of arguments when referents are contextually inferable (Sohn, 2001; J. E. Sung et al., 2024), applying the original scoring criteria could disadvantage Korean speakers, even when their constructions are grammatically complete. To address this, we implemented a pro-drop–adjusted scoring method that did not deduct points for omitted elements if the intended referent was clear from context. This ensured a linguistically appropriate and equitable evaluation for Korean speakers.

For example, consider MC-2 and MC-3 (see AphasiaBank MC protocol link provided above) from the *Refused Umbrella* picture: "The mother says you need to take the umbrella" (MC-2) and "The boy refuses to take the umbrella" (MC-3). In Korean, it is grammatically correct to say, "The mother says you need to take the umbrella, but refuses to take the umbrella," omitting the subject "he" before "refuses," as the referent ("the boy") can be retrieved from the context. According to the original MC scoring criteria, this response receives 3 points for MC-2, as all essential elements are present and accurate. However, for MC-3, it would be scored as 2 points due to the omission of the subject, which constitutes an essential element. In contrast, under the pro-drop–adjusted criteria, the same response would receive 3 points, recognizing that subject omission in this context is grammatically appropriate in Korean and does not reflect incompleteness. A detailed scoring example is provided in Supplemental Material S3.

Reliability

To ensure accuracy and consistency, the English samples were obtained from AphasiaBank, where transcripts are generated using the CLAN system following AphasiaBank's standardized transcription and utterance-segmentation conventions. For the Korean samples, speech was first transcribed using ClovaNote (https://clovanote.naver.com/), after which two Korean speech-language pathologists certified by the Ministry of Health and Welfare manually reviewed and corrected the transcripts based on the original audio recordings. The inter- and intrarater reliability for the Korean transcriptions was 100%.

A separate scoring template, which was not used in the main analyses, was created specifically for the reliability check. Speech samples included in this reliability template were obtained from three Korean-speaking neurotypical participants recruited for the study but not included in the main analyses and three English-speaking neurotypical participants from AphasiaBank.

Using this template, the inter- and intrarater reliability of MC scoring was evaluated using percent agreement. For interrater reliability, two trained raters independently scored each of the 10 MC items for all six participants (6 participants × 10 MC items = 60 scoring decisions). Agreement was calculated based on item-by-item concordance between raters; the raters agreed on 58 of the 60 MC items, yielding an interrater agreement of 97%.

Intrarater reliability was assessed by having each rater rescore the same samples after a 1-week interval. For each rater, agreement was calculated across the same 60 item-level decisions by comparing scores at Time 1 and Time 2. Rater A showed agreement on 59 of 60 items (98.3%), and Rater B showed perfect agreement (100%), resulting in a mean intrarater agreement of 99%.

Statistical Analysis

To examine original MC performance in neurotypical participants, a two-way mixed analysis of variance (ANOVA) was conducted with the language group (English vs. Korean) and picture type (*Cat Rescue* vs. *Refused Umbrella*). Building on this, to evaluate the impact of the pro-drop–adjusted MC scoring method, an additional two-way mixed ANOVA was conducted with the same factors (language group and picture type) in the neurotypical group.

To further extend the analysis to clinical populations, a three-way mixed ANOVA was performed, incorporating the clinical group (NA vs. PWA), alongside language group and picture type. The Shapiro–Wilk test indicated a significant deviation from normality for the *Refused Umbrella* MC score in the Korean NA group ($p = .010$), while all other conditions conformed to the assumption ($p > .05$). Given that ANOVA is generally robust to moderate violations of normality—particularly when group sizes are sufficiently large ($N > 25$) and the data are free from extreme outliers (Lantz, 2013)—the analysis proceeded with ANOVA as planned.

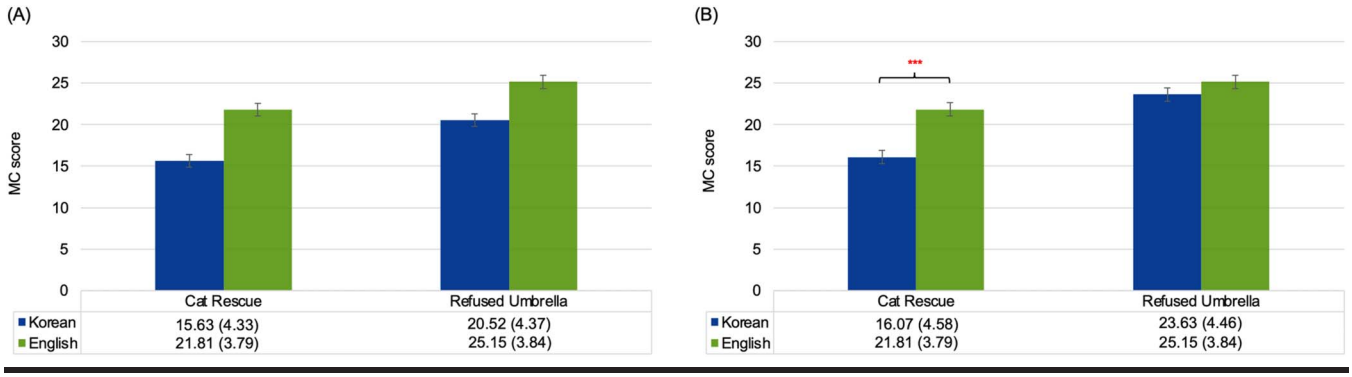
To explore which MC items could distinguish the two language groups, stepwise discriminant analyses were conducted separately for neurotypical participants (English NA vs. Korean NA) and for PWA (English PWA vs. Korean PWA), using individual MC items as predictors in the *Cat Rescue* picture. The stepwise procedure was adopted to screen candidate MC items, as it was not known which of the 10 MC items might contribute to language-group differences, and the sample size was limited. To examine the relationship between aphasia severity and overall MC scores, Pearson's correlation coefficients were conducted for each language group within PWA. Additionally, stepwise multiple linear regression analyses were conducted within the PWA, separately by language group and picture type, with overall MC scores entered as predictors and aphasia severity as the dependent variable.

Results

Original MC Analyses in Neurotypicals

We conducted a two-way mixed ANOVA to examine how original MC scores varied by language group (English vs. Korean) and picture type (*Cat Rescue* vs. *Refused Umbrella*) in neurotypical participants (see Figure 1A). The results revealed a significant main effect of language group,

Figure 1. Language group differences across picture types in neurotypical groups. (A) Original main concept scoring method; (B) Pro-drop–adjusted main concept scoring method. Error bars represent standard errors, and table values are reported as means with standard deviations in parentheses. MC = main concept.



$F(1, 52) = 46.07$, $p < .001$, partial $\eta^2 = .470$, with the Korean NA ($M = 18.07$, $SD = 4.97$) showing lower MC scores than the English NA ($M = 23.48$, $SD = 4.14$). There was also a significant main effect of the picture type, $F(1, 52) = 27.87$, $p < .001$, partial $\eta^2 = .349$, with participants scoring lower on the *Cat Rescue* ($M = 18.72$, $SD = 5.10$) than the *Refused Umbrella* ($M = 22.83$, $SD = 4.70$). However, the two-way interaction between language group and picture type was not statistically significant, $F(1, 52) = 0.10$, $p = .32$, partial $\eta^2 = .019$.

Adjusted MC Analyses in Neurotypicals

To assess the impact of incorporating the pro-drop–adjusted MC scoring method, a two-way mixed-design ANOVA was conducted to examine language group differences across picture types among neurotypical individuals. Significant main effects were found for both language group, $F(1, 58) = 16.64$, $p < .001$, partial $\eta^2 = .296$, and picture type, $F(1, 58) = 48.47$, $p < .001$, partial $\eta^2 = .451$, consistent with findings from the original MC scoring analysis. Specifically, Korean NA ($M = 19.86$, $SD = 5.88$) scored significantly lower than English NA ($M = 23.48$, $SD = 4.14$). Additionally, performance on the *Cat Rescue* ($M = 18.94$, $SD = 5.07$) was significantly worse compared to that on the *Refused Umbrella* ($M = 24.39$, $SD = 4.19$) across language groups.

Moreover, a significant interaction effect between language group and picture type was observed, $F(1, 58) = 6.83$, $p = .01$, partial $\eta^2 = .110$. To further explore this interaction, post hoc one-way ANOVAs were conducted separately for each picture. The results revealed that Korean participants ($M = 16.07$, $SD = 4.58$) performed significantly worse than English participants ($M = 21.81$, $SD = 3.79$) on the *Cat Rescue* picture, $F(1, 53) = 25.15$, $p < .001$, partial $\eta^2 = .326$. However, no significant group difference was observed on the *Refused Umbrella*, $F(1, 53) = 1.80$, $p = .19$, partial $\eta^2 = .033$ (see Figure 1B).

Adjusted MC Analyses Between Neurotypical and PWA Groups

A three-way mixed-design ANOVA was conducted using the pro-drop–adjusted MC scoring method, incorporating the clinical group (NA vs. PWA), language group, and picture type. Results revealed significant main effects for all factors. First, a significant language group effect was observed, $F(1, 84) = 9.86$, $p < .01$, partial $\eta^2 = .105$, with Korean speakers ($M = 16.17$, $SD = 7.69$) performing significantly worse compared to the English counterparts ($M = 19.13$, $SD = 7.30$). Second, a robust effect of the clinical group emerged, $F(1, 84) = 150.50$, $p < .001$, partial $\eta^2 = .626$, indicating that PWA ($M = 11.26$, $SD = 6.18$) performed significantly worse than the NA ($M = 21.67$, $SD = 5.38$). Third, the picture type effect was also significant, $F(1, 84) = 46.48$, $p < .001$, partial $\eta^2 = .356$, with lower scores on *Cat Rescue* ($M = 15.35$, $SD = 6.97$) compared to the *Refused Umbrella* ($M = 19.94$, $SD = 7.60$) across all groups.

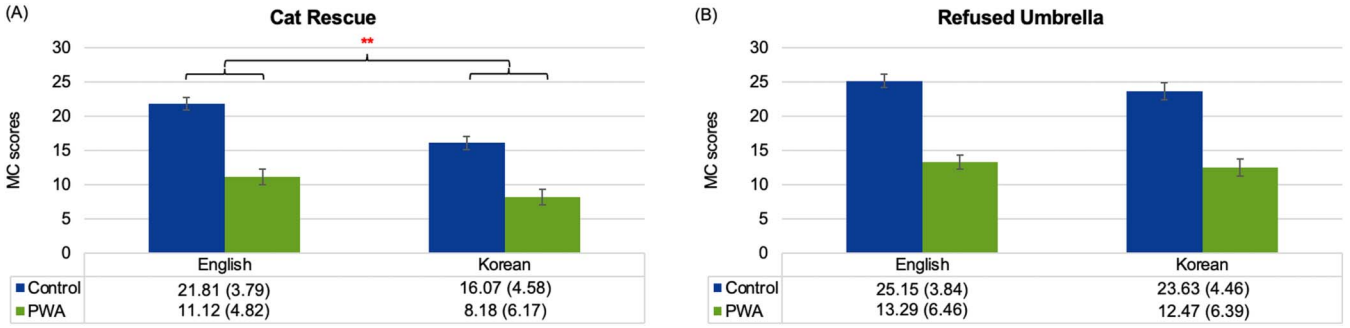
Importantly, a significant interaction between language group and picture type was found, $F(1, 84) = 6.20$, $p = .02$, partial $\eta^2 = .069$, suggesting that the differences in MC performance varied depending on the picture. To further examine the source of the interaction, post hoc one-way ANOVAs were conducted separately to examine language group differences for each picture. A significant effect of language group was found for the *Cat Rescue* picture, $F(1, 86) = 10.97$, $p = .001$, partial $\eta^2 = .113$, with Korean speakers ($M = 16.07$, $SD = 4.58$) producing significantly lower MC scores than English speakers ($M = 21.81$, $SD = 3.79$). In contrast, no significant language group difference was observed for the *Refused Umbrella* picture, $F(1, 86) = 0.59$, $p = .44$, partial $\eta^2 = .007$ (see Figure 2). Other interaction effects were not significant, including the interaction between picture type and clinical group, $F(1, 84) = 3.01$, $p = .09$, partial $\eta^2 = .035$; between language and clinical group, $F(1, 84) = 1.00$, $p = .32$, partial $\eta^2 = .012$; and the three-way interaction among language, clinical group, and picture type, $F(1, 84) = 0.68$, $p = .41$, partial $\eta^2 = .008$.

Stepwise Discriminant Analysis Using Adjusted MC Scores in the Cat Rescue Picture

As a significant language group difference emerged in the *Cat Rescue* picture, stepwise discriminant analyses were conducted separately for the neurotypical and PWA groups to explore which MC items contributed to this difference. In each analysis, 10 MCs were entered as predictor variables, and language group served as the dependent variable (English vs. Korean for the neurotypical group; English vs. Korean for the aphasic group). For the NA, four MCs significantly differentiated English NA from Korean NA: MC-8 ("Someone called the fire department."), MC-5 ("The man tries to rescue the cat."), MC-2 ("The cat was in the tree because the dog chased it."), and MC-6 ("The ladder fell down."). This model accurately classified 92.6% of English NA and 81.5% of Korean NA, Wilks' $\Lambda = .423$, $\chi^2(4) = 42.99$, $p < .001$, indicating strong discriminative validity (see Table 2).

For the PWA group, only MC-5 ("The man tries to rescue the cat.") significantly differentiated English PWA from Korean PWA. The model correctly classified 47.1% of English PWA and 94.1% of Korean PWA participants, Wilks' $\Lambda = .788$, $\chi^2(1) = 7.52$, $p = .006$, suggesting moderate discriminative ability (see Table 3).

Figure 2. Clinical group differences across language groups for each picture (from the analyses of Language Group × Clinical Group × Picture Type). (A) *Cat Rescue*; (B) *Refused Umbrella*. Error bars represent standard errors, and table values are reported as means with standard deviations in parentheses. MC = main concept; PWA = persons with aphasia.



Relationship Between the Adjusted MC Scores and Aphasia Severity

To investigate the relationship between overall MC scores and aphasia severity for each picture, Pearson's correlation coefficients and stepwise linear regression analyses were conducted separately for each PWA language group, using AQ scores. For the Korean PWA group, significant positive correlations were found between AQ and MC scores for both *Cat Rescue* (r = .748, p < .001) and *Refused Umbrella* (r = .552, p = .022). Furthermore, English PWA also showed significant positive correlations for both tasks (*Cat Rescue*: r = .559, p = .020; *Refused Umbrella*: r = .652, p = .005).

To explore which picture better predicts the AQ, a stepwise multiple linear regression was conducted for each group. In Step 1, overall MC scores from the two pictures were entered as predictors with AQ as the outcome variable. In Step 2, age and education were added to adjust for demographic factors and to confirm that the observed associations were not attributable to confounding variables. For the Korean PWA, the analysis identified that MC scores on the *Cat Rescue* significantly predicted AQ scores, $F(1, 15) = 19.10$, $p < .001$, $R^2 = .560$, accounting for 56% of the variance in aphasia severity (see Figures 3A and 3B). Adding age and education in Step 2 did not significantly increase the explained variance (ΔR^2 ps = .343). For the English PWA, MC scores on the *Refused Umbrella* significantly predicted AQ scores, $F(1, 15) = 11.067$, p = .005, R^2 = .425, accounting for 42.5% of the total variance (see Figures 3C and 3D). Adding age and education in Step 2 did not significantly increase the explained variance (ΔR^2 ps = .166).

Discussion

This study investigated the influence of linguistic and cultural factors on picture-based narrative performance in Korean and English speakers with and without aphasia. Given the assumption that cross-linguistic differences (e.g., pro-drop features) may affect overall discourse performance—particularly in MC analyses, which rely on the accurate counting of subjects and pronouns—we

Table 2. Summary of stepwise discriminant analysis for neurotypical language groups on the *Cat Rescue* picture.

MC item	Description	Standardized coefficient	Structure matrix	Rank
1	The little girl was riding her bicycle.		–.036	
2	The cat was in the tree because the dog chased it.	.470	.405	3
3	The dog was barking up the tree.		.023	
4	The man climbed up the tree.		.107	
5	The man tries to rescue the cat.	.580	.504	2
6	The ladder fell down.	.419	.401	4
7	The father is stuck in the tree with the cat.		.214	
8	Someone called the fire department.	.632	.553	1
9	The fire department comes with a ladder.		–.270	
10	The fire department rescues them.		.054	

Note. Overall classification accuracy: 92.6% for English neurotypicals, 81.5% for Korean neurotypicals. Discriminant function: Wilks' Λ = .423, $\chi^2(4) = 42.99$, $p < .001$. MC = main concept.

Table 3. Summary of stepwise discriminant analysis for PWA language groups on the *Cat Rescue* picture.

MC item	Description	Standardized coefficient	Structure matrix	Rank
1	The little girl was riding her bicycle.		–.173	
2	The cat was in the tree because the dog chased it.		.072	
3	The dog was barking up the tree.		.450	
4	The man climbed up the tree.		.138	
5	The man tries to rescue the cat.	1.000	1.000	1
6	The ladder fell down.		.246	
7	The father is stuck in the tree with the cat.		.364	
8	Someone called the fire department.		.031	
9	The fire department comes with a ladder.		–.031	
10	The fire department rescues them.		.192	

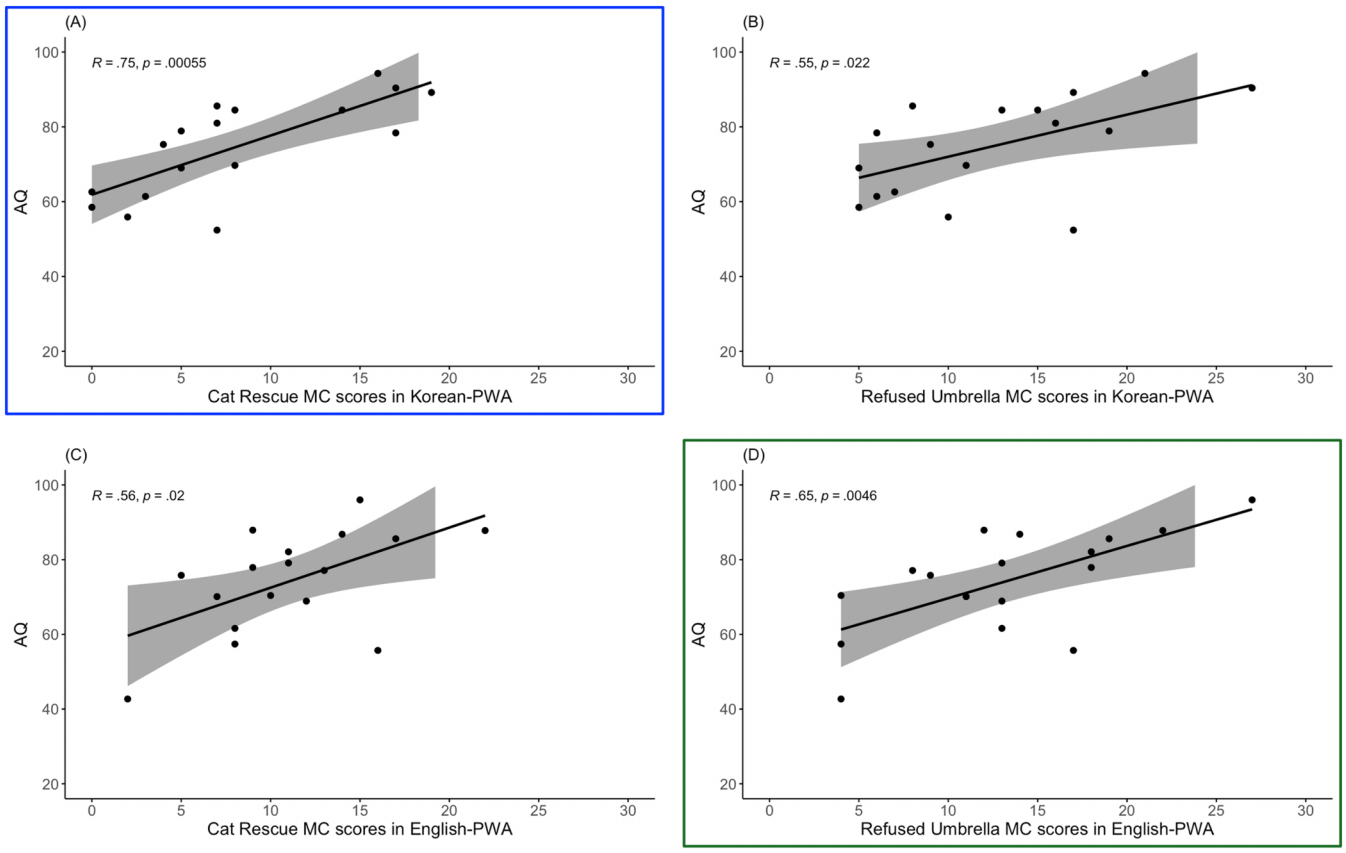
Note. Overall classification accuracy: 47.1% for English neurotypicals, 94.1% for Korean neurotypicals. Discriminant function: Wilks' Λ = .788, $\chi^2(1) = 7.519$, $p = .006$. PWA = persons with aphasia; MC = main concept.

compared picture-specific effects using two scoring methods: (a) the original MC scoring and (b) a pro-drop–adjusted MC scoring. Picture-specific effects were observed under both scoring approaches, in both neurotypical individuals and PWA, even after accounting for language-specific features.

Effects of Picture Stimuli From a Cross-Cultural and Cross-Linguistic Perspective

To examine the influence of linguistic features on discourse performance, we first compared neurotypical

Figure 3. Scatterplots showing the relationship between *Cat Rescue* and *Refused Umbrella* MC scores and AQ scores in Korean- and English-speaking PWA. (A) and (B) depict the Korean-PWA data; (C) and (D) depict the English-PWA data. Blue and green outlines highlight the significant predictors of AQ identified via stepwise linear regression: *Cat Rescue* for Korean-PWA and *Refused Umbrella* for English-PWA. Each plot shows a positive association between MC scores and AQ, indicating that higher MC scores are associated with higher AQ (i.e., lower aphasia severity). AQ = Aphasia Quotient; MC = main concept; PWA = persons with aphasia.



groups under the original MC scoring. Korean speakers scored lower than English speakers, which may reflect the influence of the pro-drop feature of Korean, allowing the omission of contextually recoverable arguments. However, because the original MC scoring credits only explicitly produced arguments, the system may disadvantage speakers of pro-drop languages, even when their utterances are syntactically well formed and pragmatically appropriate. To reduce this limitation and allow more equitable cross-linguistic comparisons, we incorporated Korean pro-drop features into the analyses.

Even after applying the linguistic adjustment, Korean neurotypical speakers still performed worse than English neurotypical speakers, but only on the *Cat Rescue* picture. These findings suggest that picture-specific factors beyond linguistic features might affect discourse performance. A similar pattern emerged when PWA were included in the analysis. The significant Language Group × Picture interaction showed that Korean speakers, across both neurotypical and aphasic groups, scored lower than English speakers on the *Cat Rescue*, whereas no group difference was found on the *Refused Umbrella*. These picture-specific effects in the *Cat Rescue* may reflect differences in cultural familiarity with the depicted event, as fire department involvement in companion animal rescue is uncommon in Korea and typically limited to cases involving dangerous animals or protected species (W.-J. Kim, 2023). Such unfamiliarity may have disadvantaged Korean speakers in identifying and elaborating the MCs of the story. By contrast, the *Refused Umbrella* picture depicted a relatively universal, everyday situation, enabling both groups to capture the main content without cultural effects. Consistent with our findings, prior research has shown that cultural background plays a critical role in shaping the way individuals perceive and describe visual scenes (Berger & Ponti, 2024; Lee et al., 2024; Marcotte et al., 2024; Pillay et al., 2024; Steinberg et al., 2022; Stockbridge et al., 2024; van Miltenburg et al., 2017). For instance, Berger and Ponti (2024) analyzed a multilingual data set of picture descriptions across 31 languages and identified cross-cultural effects in narrative production. Their results demonstrated that cultural background shapes the way individuals describe visual scenes: Geographically or genetically related languages tended to highlight similar objects in the picture, whereas culturally salient elements—such as clothing, accessories, or landscapes—were emphasized differently across groups.

To further investigate the cultural effect found in the *Cat Rescue*, we conducted MC item-level analyses for both neurotypical participants and PWA to identify which MCs differentiated Korean and English speakers. Among the 10 MCs, MC-5 ("The man tries to rescue the cat")—which conveys the causal foundation and main storyline of the scene—was the primary discriminator of Korean and English speakers in both neurotypical and aphasic groups. Unlike other MCs that describe concrete, visually explicit actions (e.g., MC-3: *The dog is barking*; MC-7: *The man is stuck in the tree*; MC-9: *The fire department comes with a ladder*), MC-5 required integrating multiple visual elements to infer that the cat is in an emergency (e.g., the girl reaching toward the cat, the man climbing the tree, the arrival of the fire department). As this core storyline is not explicitly depicted in the scene, generating MC-5 required additional inferential processing. Such inferencing might become more challenging when the scene itself is culturally unfamiliar, thereby increasing the processing load. Supporting this view, prior studies have demonstrated that unfamiliar visual stimuli hinder language production: Object unfamiliarity reduces naming accuracy in picture-naming tasks (He et al., 2023), and low-familiarity stimuli impose greater cognitive load and reduce maintenance in working memory tasks (Schneider et al., 2024). Consequently, when more cognitive resources are allocated to interpreting an unfamiliar picture, fewer may remain available for higher-level inferencing within a limited time, which could help explain why Korean speakers were less likely to produce MC-5 in the *Cat Rescue* picture.

Exploratory MC item-level findings also suggest that when a picture is culturally unfamiliar, Korean neurotypical speakers may rely more on visually explicit details and produce fewer inference-dependent concepts. Beyond MC-5, several inference-dependent MCs significantly differentiated Korean NA from English NA (e.g., MC-2: *The cat was in the tree because the dog chased it*; MC-6: *The ladder fell down*; MC-8: *Someone called the fire department*). Consistent with this, Korean NA tended to confine their descriptions to concrete, visually explicit elements rather than inference-based content (e.g., "There is a man in a tree…A dog is barking…Firefighters are coming…Why are they coming?"). Taken together, these patterns are hypothesis-generating and suggest that cultural unfamiliarity with the scene may have limited deeper inferencing during the construction of the story.

Picture-salient effects in discourse performance also emerged in predicting aphasia severity, with different regression patterns across language groups. For Korean PWA, MC scores on the *Cat Rescue* picture significantly predicted AQ, suggesting that the added difficulty of interpreting and producing the main contents in a less culturally familiar picture may have been more sensitive to individual differences in language ability. In contrast, for English PWA, MC scores of the *Refused Umbrella* picture significantly predicted AQ, likely reflecting the greater cognitive and linguistic demands of sequential picture tasks, which are known to elicit structurally complex and lexically diverse output (Deng et al., 2024; Schnur &

Wang, 2024). We speculate that, for Korean PWA, cultural unfamiliarity in the *Cat Rescue* functioned as a stronger challenge than the sequential format itself, leading to the observed group-specific patterns. These findings highlight that the clinical implications of using picture stimuli may vary with cultural familiarity, underscoring the need for careful consideration in stimulus selection and interpretation in clinical contexts.

Clinical Implications and Research Directions

A growing body of research also emphasizes that cross-linguistic adaptations of picture-based tasks must begin with conceptual and cultural adaptation of the materials themselves (Jeong et al., 2023; Seçkin & Savaş, 2023). For example, Seçkin and Savaş (2023) compared discourse production in Turkish-speaking adults across three pictures—the *Picnic Scene*, the *Cookie Theft*, and the *Accident Scene* from the Turkish Language Assessment Test. They found that the *Accident Scene* prompted more complex morphosyntactic constructions (e.g., subordinate clauses, derivational and passive suffixes), underscoring the importance of developing culturally adapted picture stimuli tailored to elicit specific linguistic features in Turkish. Similarly, Jeong et al. (2023) introduced a culturally familiar picture stimulus for Korean speakers (*Han River*, Patent No. D2022-0004KR), developed through a multistage process that incorporated linguistic, cultural, and age-specific design considerations. They selected the Han River—an iconic and socially meaningful location in Seoul, South Korea (Baik & Kim, 2012)—as the background and extracted core event components (characters, actions, objects, and context) through extensive image research, then applied principles of simplicity, specificity, and comprehensibility to support effortless recognition for older adults. Building on this culturally adapted picture stimulus, Lee et al. (2024) empirically demonstrated picture-salient cultural effects by comparing discourse performance on two stimuli—the PK-WAB-R *Beach* picture and the *Han River* picture—among healthy younger and older Korean speakers. They examined word-level linguistic variables, including correct information units (CIUs) per minute as an index of informativeness and noun and verb counts per utterance as indices of productivity. Their findings showed that both age groups produced significantly more CIUs when describing the *Han River* picture than the *Beach* picture. Moreover, participants generated a greater number of verb tokens in the *Han River* picture compared with the *Beach* picture, a pattern not observed for noun tokens. Consistent with these approaches, the Collaboration of Aphasia Trialists has also highlighted that developing conceptually and culturally appropriate tools is a critical prerequisite for valid cross-linguistic comparison (Ali et al., 2022; Martínez-Ferreiro et al., 2024).

Beyond cultural adaptation of picture materials, recent efforts have been made to develop culturally neutral picture stimuli for discourse assessment (Marcotte et al., 2024; Stockbridge et al., 2024). For example, Stockbridge et al. (2024) developed a culturally neutral stimulus—a young person falling from a stepladder while painting—as part of the National Institutes of Health Stroke Scale. This stimulus was validated with 101 neurotypical English-speaking individuals from diverse countries (the United States, Germany, Canada, the United Kingdom, Australia, and Zambia). The results showed no significant differences in discourse performance by age, sex, race, ethnicity, or education, demonstrating the feasibility of creating culturally unbiased stimuli for cross-linguistic and cross-cultural language assessment. Such efforts also align with the COSMIN framework (Consensus-based Standards for the Selection of Health Measurement Instruments; Mokkink et al., 2010), which emphasizes that assessment tools should establish adequate construct and cross-cultural validity to ensure psychometric quality in clinical applications. Taken together, our findings highlight the need either to use universally acceptable picture stimuli or to explicitly consider the cultural and linguistic aspects of the materials to minimize cross-conceptual variation and improve measurement validity across populations.

The present findings raise several important considerations for both clinical practice and future research. A key direction for future work is to further clarify how linguistic and cultural factors differentially shape discourse performance, particularly when comparing individuals from diverse linguistic and cultural backgrounds. Given that the present study did not directly measure cultural familiarity and that differences between the *Cat Rescue* and *Refused Umbrella* picture could also reflect other stimulus-related characteristics—such as visual complexity, sequential structure, or cognitive demands—future research would benefit from assessing these influences more explicitly. Incorporating direct measures of cultural familiarity or scene comprehension—for example, through quantitative ratings such as Likert-scale scores—while simultaneously accounting for relevant typological features would help clarify the relative contributions of linguistic, cultural, and stimulus-driven factors. Such approaches would strengthen the validity of cross-cultural discourse assessments and support more culturally sensitive and linguistically appropriate clinical practices.

In addition, our results suggest that linguistic adaptations to the MC scoring system may be warranted when it is applied to typologically diverse languages. In the current study, incorporating Korean pro-drop features enabled MC scores to more accurately reflect language-specific discourse patterns, indicating that typologically informed scoring adjustments can improve cross-linguistic comparability. Although the present adjustment was tailored to Korean,

the underlying principle of typology-sensitive scoring may be relevant to other pro-drop languages. However, language-specific criteria and empirical validation would be required before applying similar adjustments in other linguistic contexts. This approach aligns with Sahraoui et al. (2025), who argued that without linguistically and culturally informed scoring procedures, cross-linguistic discourse measures risk reduced conceptual validity and limited clinical interpretability. Future research should explore whether additional linguistic or cultural modifications to the MC scoring procedure enhance its clinical utility in multilingual settings.

Furthermore, as the current study utilized preexisting data from AphasiaBank for English speakers, potential differences in elicitation context—such as examiner behavior, testing environment, and adherence to administration protocols—may have influenced the discourse outcomes. These contextual factors can introduce variability that is unrelated to language or cultural differences, potentially confounding cross-linguistic or cross-cultural comparisons. Accordingly, future research should aim to control for, or systematically investigate, the effects of elicitation context to ensure more valid and reliable cross-group analyses. Additionally, given the relatively small PWA sample sizes in the present study, future research with larger and more systematically collected clinical data sets will be needed to further evaluate and replicate these findings.

Building on these considerations, ongoing discussions regarding how best to balance standardized assessment materials with the need for cultural adaptation remain essential in multilingual aphasia evaluation. Continued work in this area will help clarify when standardization supports reliable cross-group comparison and when culturally tailored materials or scoring approaches may be necessary to avoid conceptual bias and improve the ecological validity of discourse assessment.

Conclusions

In summary, this study showed that both linguistic and cultural features influence picture-based narrative performance in Korean and English speakers with and without aphasia. Notably, the findings demonstrated that even widely used, standardized picture stimuli can elicit systematically different narrative patterns across language groups, independent of core linguistic features such as pro-drop. For Korean speakers, the lower cultural familiarity with the *Cat Rescue* picture may have contributed to greater difficulty in interpreting and producing the main storyline, particularly in contexts that appear to require additional inferential processing. Such picture-salient patterns suggest that culturally embedded aspects of visual stimuli could influence discourse production and may complicate cross-linguistic interpretation if not carefully considered. By integrating typology-informed scoring adjustments with picture-level analyses, the present study underscores the importance of jointly considering linguistic structure and picture-salient contextual factors when interpreting discourse performance in multilingual clinical settings. Together, these findings highlight the need for caution in selecting and interpreting picture stimuli in clinical assessment across linguistically and culturally diverse populations and point to future research directions examining how linguistic and cultural factors interact with picture-based stimuli across broader cultural contexts and clinical populations.

Author Contributions

Yae Rin Yoo: Conceptualization, Formal analysis, Writing – original draft. **Jimin Park:** Conceptualization, Formal analysis, Writing – original draft. **Jee Eun Sung:** Conceptualization, Supervision, Writing – review & editing.

Data Availability Statement

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by the National Research Foundation of Korea grants funded by the Ministry of Science and Information and Communication Technology (RS-2022-NR070151, RS-2024-00461617) to Jee Eun Sung.

References

Ali, M., Soroli, E., Jesus, L. M. T., Cruice, M., Isaksen, J., Visch-Brink, E., Grohmann, K. K., Jagoe, C., Kukkonen, T., Varlokosta, S., Hernandez-Sacristan, C., Rosell-Clari, V., Palmer, R., Martinez-Ferreiro, S., Godecke, E., Wallace, S. J., McMenamin, R., Copland, D., Breitenstein, C., . . . Collaboration of Aphasia Trialists (CATs). (2022). An aphasia research agenda—A consensus statement from the collaboration of aphasia trialists. *Aphasiology, 36*(4), 555–574. https://doi.org/10.1080/02687038.2021.1957081

Baik, S. K., & Kim, E. J. (2012). Han River Renaissance Project in semiotics of space. In *Communication of culture* (pp. 959–973). Universidade da Coruña. http://hdl.handle.net/2183/13402

Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis. *Archives of Neurology, 51*(6), 585–594. https://doi.org/10.1001/archneur.1994.00540180063015

Berger, U., & Ponti, E. M. (2024). *Cross-lingual and cross-cultural variation in image descriptions*. arXiv. https://doi.org/10.48550/arXiv.2409.16646

Berube, S., Nonnemacher, J., Demsky, C., Glenn, S., Saxena, S., Wright, A., Tippett, D. C., & Hillis, A. E. (2019). Stealing cookies in the twenty-first century: Measures of spoken narrative in healthy versus speakers with aphasia. *American Journal of Speech-Language Pathology, 28*(1S), 321–329. https://doi.org/10.1044/2018_AJSLP-17-0131

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language, 27*(1), 69–96. https://doi.org/10.1017/S0305000999004018

Croft, W. (2000). *Explaining language change: An evolutionary approach*. Longman.

Deng, B.-M., Gao, J., Liang, L.-S., Zhao, J.-X., Lin, F., Yin, M.-Y., Zheng, H.-Q., & Hu, X.-Q. (2024). Discourse task type–specific linguistic characteristics in anomic aphasia and healthy controls: Evidence from Mandarin–Chinese AphasiaBank. *American Journal of Speech-Language Pathology, 33*(2), 937–951. https://doi.org/10.1044/2023_AJSLP-23-00078

El Ouardi, L., Yeou, M., Rami, Y., & Belahsen, M. F. (2023). The Moroccan Arabic bedside Western Aphasia Battery–Revised: Linguistic and psychometric properties. *Journal of Speech, Language, and Hearing Research, 66*(7), 2376–2389. https://doi.org/10.1044/2023_JSLHR-22-00475

Evans, E., Coley, S. L., Gooding, D. C., Norris, N., Ramsey, C. M., Green-Harris, G., & Mueller K. D. (2022). Preliminary assessment of connected speech and language as marker for cognitive change in late middle-aged Black/African American adults at risk for Alzheimer's disease. *Aphasiology, 36*(8), 982–1005. https://doi.org/10.1080/02687038.2021.1931801

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

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences, 26*(4), 243–254. https://doi.org/10.1007/s10072-005-0467-9

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease, 49*(2), 407–422. https://doi.org/10.3233/JAD-150520

Giles, E., Patterson, K., & Hodges, J. R. (1996). Performance on the Boston Cookie Theft picture description task in patients with early dementia of the Alzheimer's type: Missing information. *Aphasiology, 10*(4), 395–408. https://doi.org/10.1080/02687039608248419

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Lea & Febiger.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Lea & Febiger.

Gordon, J. K. (2008). Measuring the lexical semantics of picture description in aphasia. *Aphasiology, 22*(7–8), 839–852. https://doi.org/10.1080/02687030701820063

Hameister, I., & Nickels, L. (2018). The cat in the tree—Using picture descriptions to inform our understanding of conceptualisation in aphasia. *Language, Cognition and Neuroscience, 33*(10), 1296–1314. https://doi.org/10.1080/23273798.2018.1497801

He, Y., Liao, X., Liang, J., & Boleda, G. (2023). *The impact of familiarity on naming variation: A study on object naming in Mandarin Chinese*. arXiv. https://doi.org/10.48550/arXiv.2311.10181

Jeong, I., Han, J., Choi, J., Choi, Y., Choi, S., & Sung, J. E. (2023). Improving picture description tasks for Korean seniors. *Journal of Digital Contents Society, 24,* 2767–2775. https://doi.org/10.9728/dcs.2023.24.11.2767

Jomaa, I. M. A., Moustafa, R. R., Wahba, W. B., Hamid, E., El Nahas, N., & Roushdy, T. (2025). Developing a culturally adapted Cairene Egyptian Arabic–Western Aphasia Battery—Revised (CEA-WAB-R). *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery, 61*(1), Article 57. https://doi.org/10.1186/s41983-025-00994-1

Kang, Y. (2006). A normative study of the Korean Mini-Mental State Examination (K-MMSE) in the elderly. *The Korean Journal of Psychology, 25*(2), 1–12.

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

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

Kim, W.-J. (2023). Legislative improvement plan to strengthen firefighting field response. *Public Land Law Review, 103,* 223–244. https://doi.org/10.30933/kpllr.2023.103.223.

Kim, Y.-J. (2000). Subject/object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics, 9,* 325–351. https://doi.org/10.1023/A:1008304903779

Kuzmina, E., Goral, M., Norvik, M., & Weekes, B. S. (2019). What influences language impairment in bilingual aphasia? A meta-analytic review. *Frontiers in Psychology, 10,* Article 445. https://doi.org/10.3389/fpsyg.2019.00445

Kwon, N., & Sturt, P. (2013). Null pronominal (pro) resolution in Korean, a discourse-oriented language. *Language and Cognitive Processes, 28,* 377–387. https://doi.org/10.1080/01690965.2011.645314

Lantz, B. (2013). The impact of sample non-normality on ANOVA and alternative methods. *British Journal of Mathematical and Statistical Psychology, 66*(2), 224–244. https://doi.org/10.1111/j.2044-8317.2012.02047.x

Lee, H., Choi, Y., & Sung, J. E. (2024). Age-related changes in connected speech production: Evidence from eye-tracking in the culturally adapted picture description task. *Frontiers in Psychology, 15,* Article 1334788. https://doi.org/10.3389/fpsyg.2024.1334788

MacWhinney, B. (2007). The TalkBank Project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora* (pp. 163–180). Palgrave Macmillan. https://doi.org/10.1057/9780230223936_7

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Marcotte, K., Roy, A., Brisebois, A., Jutras, C., Leonard, C., Rochon, E., & Brambati, S. M. (2024). Reliability of the picture description task of the Western Aphasia Battery–Revised in Laurentian French persons without brain injury. *The Clinical Neuropsychologist, 38*(8), 1980–2008. https://doi.org/10.1080/13854046.2024.2340777

Martínez-Ferreiro, S., Arslan, S., Fyndanis, V., Howard, D., Kraljević, J. K., Škorić, A. M., Munarriz-Ibarrola, A., Norvik, M., Peñaloza, C., Pourquié, M., Simonsen, H. G., Swinburn, K., Varlokosta, S., & Soroli, E. (2024). Guidelines and recommendations for cross-linguistic aphasia assessment: A review of 10 years of comprehensive aphasia test adaptations. *Aphasiology, 40*(2), 125–239. https://doi.org/10.1080/02687038.2024.2343456

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international

Delphi study. *Quality of Life Research, 19*(4), 539–549. https://doi.org/10.1007/s11136-010-9606-8

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology, 40*(9), 917–939. https://doi.org/10.1080/13803395.2018.1446513

Neves, M. de B., Van Borsel, J., Pereira, M. M. de 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

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non–brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research, 38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Nilipour, R., Pourshahbaz, A., & Ghoreyshi, Z. S. (2014). Reliability and validity of bedside version of Persian WAB (P-WAB-1). *Basic and Clinical Neuroscience, 5*(4), 253–258.

Pillay, B. S., Van Der Linde, J., Graham, M. A., & Dada, S. (2024). Image use in aphasia rehabilitation: Practices of speech-language pathologists in South Africa and international contexts. *Pathology, 28*(1), 112–125. https://doi.org/10.1080/17549507.2024.2434229

Poisson, S. N., Block, K. L., & Orjuela, K. D. (2022). Implications of the National Institutes of Health Stroke Scale Cookie Theft picture—A closer look. *JAMA Neurology, 79*(8), 734–735. https://doi.org/10.1001/jamaneurol.2022.1409

Ramkissoon, I., Dagenais, P. A., Evans, K. J., Camp, T. J., & Ferguson, N. N. (2013). Effects of ethnically diverse photographic stimuli on preference and discourse tasks in African American and Caucasian American adults. *Communication Disorders Quarterly, 34*(2), 97–105. https://doi.org/10.1177/1525740111420707

Richardson, J. D., & Dalton, S. G. H. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology, 30*(1), 45–73. https://doi.org/10.1080/02687038.2015.1057891

Richardson, J. D., & Dalton, S. G. H. (2020). Main concepts for two picture description tasks: An addition to Richardson and Dalton, 2016. *Aphasiology, 34*(1), 119–136. https://doi.org/10.1080/02687038.2018.1561417

Rivera, A., Hirst, J., & Edmonds, L. A. (2018). Evaluation of language predictors of main concept production in Spanish/English bilingual discourse using Nicholas and Brookshire stimuli. *American Journal of Speech-Language Pathology, 27*(1), 52–70. https://doi.org/10.1044/2017_AJSLP-15-0186

Sahraoui, H., Martínez-Ferreiro, S., & Soroli, E. (2025). Theoretical and methodological challenges to define a cross-linguistic discourse core outcome set of measures for aphasia. *Aphasiology*. Advance online publication. https://doi.org/10.1080/02687038.2025.2552801

Santi, G. C., Catricala, E., Kwan, S., Wong, A., Ezzes, Z., Wauters, L., Esposito, V., Conca, F., Gibbons, D., Fernandez, E., Santos-Santos, M. A., TaFu, C., Li-Ying, K.-C., Lo, R., Tsoh, J., Chan, L.-T., Garcia, A. M., de Leon, J., Miller, Z., . . . & Tee, B. L. (2024). *Cross-linguistic analysis of speech markers: Insights from English, Chinese, and Italian speakers*. medRxiv. https://www.medrxiv.org/content/10.1101/2024.10.15.24314191v1

Schneider, P., Vergauwe, E., & Camos, V. (2024). The visual familiarity effect on attentional working memory maintenance. *Memory & Cognition, 52*(8), 1882–1899. https://doi.org/10.3758/s13421-024-01548-1

Schnur, T. T., & Wang, S. (2024). Differences in connected speech outcomes across elicitation methods. *Aphasiology, 38*(5), 816–837. https://doi.org/10.1080/02687038.2023.2239509

Schönberger, E., Heim, S., Meffert, E., Pieperhoff, P., da Costa Avelar, P., Huber, W., Binkofski, F., & Grande, M. (2014). The neural correlates of agrammatism: Evidence from aphasic and healthy speakers performing an overt picture description task. *Frontiers in Psychology, 5,* Article 246. https://doi.org/10.3389/fpsyg.2014.00246

Seçkin, M., & Savaş, M. (2023). Picnic, accident or cookies? A systematic approach to guide the selection of the picture definition tasks in linguistic assessment. *Archives of Clinical Neuropsychology, 38*(2), 236–246. https://doi.org/10.1093/arclin/acac109

Shimada, M., Meguro, K., Yamazaki, H., Horikawa, A., Hayasaka, C., Yamaguchi, S., Yamaguchi, K., Katsuyama, N., Nakano, M., & Yamadori, A. (1998). Impaired verbal description ability assessed by the Picture Description Task in Alzheimer's disease. *Archives of Gerontology and Geriatrics, 27*(1), 57–65. https://doi.org/10.1016/S0167-4943(98)00099-5

Slobin, D. I. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (Eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 435–445). Berkeley Linguistics Society. https://doi.org/10.3765/bls.v13i0.1826

Sohn, H.-M. (2001). *The Korean language*. Cambridge University Press.

Steinberg, A., Lyden, P. D., & Davis, A. P. (2022). Bias in stroke evaluation: Rethinking the Cookie Theft picture. *Stroke, 53*(6), 2123–2125. https://doi.org/10.1161/STROKEAHA.121.038515

Stockbridge, M. D., Kelly, L., Newman-Norlund, S., White, B., Bourgeois, M., Rothermel, E., Fridriksson, J., Lyden, P. D., & Hillis, A. E. (2024). New picture stimuli for the NIH Stroke Scale: A validation study. *Stroke, 55*(2), 443–451. https://doi.org/10.1161/STROKEAHA.123.044384

Sung, H. B. (2021). *A legal study on animal protection and animal welfare in disaster management: Focusing on disaster management for animals in the U.S.* [Master's thesis, Kangwon National University].

Sung, J. E., DeDe, G., & Lee, S. E. (2016). Cross-linguistic differences in a picture-description task between Korean- and English-speaking individuals with aphasia. *American Journal of Speech-Language Pathology, 25*(4S), S813–S822. https://doi.org/10.1044/2016_AJSLP-15-0140

Sung, J. E., Scimeca, M., Li, R., & Kiran, S. (2024). Cross-linguistic and multicultural considerations in evaluating bilingual adults with aphasia. *American Journal of Speech-Language Pathology, 33*(6), 2716–2731. https://doi.org/10.1044/2024_AJSLP-23-00496

Talmy, L. (2000). *Toward a cognitive semantics* (Vols. 1–2). MIT Press.

van Miltenburg, E., Elliott, D., & Vossen, P. (2017). Cross-linguistic differences and similarities in image descriptions. In *Proceedings of the 10th International Conference on Natural Language Generation* (pp. 21–30). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1707.01736

Wang, Y. (1997). Introduction of the Western Aphasia Battery. *Chinese Journal of Rehabilitation Theory and Practice, 3*(2), 87–89.

Yiu, E., Worrall, L., & Baglioni, T. (1998). Classification of aphasic Chinese speakers: Cluster and discriminant function analyses. *Aphasiology, 12*(1), 37–48. https://doi.org/10.1080/02687039808249442