

Quantitative analysis of lexical psycholinguistic properties in aphasia discourse production

Juqiang Chen & Hui Chang

To cite this article: Juqiang Chen & Hui Chang (01 Oct 2025): Quantitative analysis of lexical psycholinguistic properties in aphasia discourse production, International Journal of Speech-Language Pathology, DOI: [10.1080/17549507.2025.2557501](https://doi.org/10.1080/17549507.2025.2557501)

To link to this article: <https://doi.org/10.1080/17549507.2025.2557501>



Published online: 01 Oct 2025.



Submit your article to this journal [↗](#)



Article views: 34



View related articles [↗](#)



View Crossmark data [↗](#)



Quantitative analysis of lexical psycholinguistic properties in aphasia discourse production

JUQIANG CHEN^{1,2} & HUI CHANG^{1,2}

¹School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, P.R. China, ²National Research Centre for Language and Well-being, Shanghai, P.R. China

Abstract

Purpose: This study aims to investigate lexical psycholinguistic properties (i.e. age of acquisition, concreteness, imageability, and familiarity) in English spoken discourse by persons with aphasia. It is hypothesised that persons with aphasia are more likely to use words with lower age of acquisition and higher concreteness, imageability and familiarity compared to the control group of neurologically intact adults, reflecting their lexical retrieval difficulties.

Method: Language samples of picture descriptions, story narratives, and procedural discourse were extracted from AphasiaBank. Words in the samples were cross-referenced with lexical items in large-scale psycholinguistic norms.

Result: Persons with aphasia tend to use words with lower age of acquisition in the story narrative task and higher familiarity in both the story narrative and procedural description task compared to the control group, demonstrating difficulties in retrieving later-acquired, and less familiar lexical items. Story narratives were particularly effective in distinguishing the two groups in terms of lexical psycholinguistic properties.

Conclusion: Distinctive patterns of lexical psycholinguistic properties were found in persons with aphasia discourse production compared with the control groups. These findings hold significant clinical implications for speech-language pathologists, as they underscore the value of integrating psycholinguistic measures into assessment protocols to enhance diagnosis accuracy and inform targeted therapeutic interventions.

History: Received 26 June 2024; Accepted 1 September 2025

Keywords: psycholinguistic norms; age of acquisition; concreteness; imageability; familiarity; aphasia discourse

Introduction

Persons with aphasia (PWA) often face challenges in lexical retrieval across various contexts, as revealed in single-word generation, sentence formulation, and discourse production tasks (Bryant et al., 2016; Heikkola et al., 2022; Marshall et al., 1998; Wilshire et al., 2014), which can impede their engagement in everyday communication. Discourse production offers a valuable data source for evaluating language impairments when multiple linguistic domains interact (Marini et al., 2011) and in a more naturalistic setting (Dipper & Pritchard, 2017) than the production of single words. Therefore, analysing discourse production has gained an increasingly important role in aphasia research (Chen & Chang, 2024a, 2025; Deng et al., 2024; Deng et al., 2024; Dipper et al., 2017, 2021; Pritchard et al., 2017, 2018).

Evaluating lexical properties in aphasia discourse is a critical focus in both clinical settings and aphasia research (for a review, see Bryant et al., 2016).

Linguistic lexical measures

The most frequently used lexical measures in aphasia discourse analysis are the total number of words and the total number and/or proportion of different word classes, such as nouns, verbs, and pronouns. The latter is related to lexical density, which is the ratio of lexical/content words to all words in the discourse. It was found that PWA generally produced a smaller proportion of words (e.g. proportion of narrative words in sentences), a smaller proportion of personal pronouns to the sum of pronouns and nouns, and a smaller proportion of verbs to the sum of verbs and nouns, than the control group (Rochon et al., 2000).

Correspondence: Hui Chang, School of Foreign Languages, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Minhang District, 200240, Shanghai, P. R. China. E-mail: ch9647@sjtu.edu.cn

These lexical measures have been included in several standardised, multi-measure linguistic assessment systems, such as the Quantitative Production Analysis (QPA) (Saffran et al., 1989), the Northwestern Narrative Language Analysis system (NNLA) (Thompson et al., 1995), and the Shewan Spontaneous Language Analysis (SSLA) (Shewan, 1988). Thanks to the development of the Codes for the Human Analysis of Transcripts (CHAT) (MacWhinney et al., 2010), and the Computerised Language Analysis system (CLAN) (designed to automatically analyse these CHAT transcripts; MacWhinney et al., 2011), automated computer programs were developed for linguistic assessment of aphasia discourse (Fromm et al., 2020, 2021; Hsu & Thompson, 2018).

In addition, lexical diversity, the range of vocabulary usage, has also been used in the linguistic analysis of aphasia discourse (Fergadiotis et al., 2013; Fergadiotis & Wright, 2011; Wachal & Spreen, 1973; Wright et al., 2003). It was found that conversation and picture description discourse produced by adults with fluent aphasia had significantly higher lexical diversity than those with nonfluent aphasia (Wright et al., 2003). Similarly, neurologically intact adults demonstrated greater lexical diversity than PWA across all three types of discourse (i.e. single-picture description, sequential picture description, and storytelling) (Fergadiotis & Wright, 2011). It should be noted that for both lexical density and lexical diversity, only word information internal to the text (i.e. frequencies and/or proportions of words and word classes of the speech transcripts) is considered. To further investigate lexical psycholinguistic knowledge, word psycholinguistic norms with information external to the text in question (i.e. human ratings of age of acquisition, concreteness, imageability, and familiarity) should be examined (Crossley & Skalicky, 2019; Salsbury et al., 2011).

Lexical psycholinguistic properties

Lexical psycholinguistic properties, such as age of acquisition, concreteness, familiarity, and imageability, have been shown to correlate in varying degrees with accuracy and speed of lexical processing and retrieval across different populations, such as PWA, older adults, and patients with Parkinson's disease (Aggujaro et al., 2006; Bayram et al., 2021; Brysbaert & Ellis, 2016; Connine et al., 1990; Zhang et al., 2006).

Age of acquisition (AoA) refers to the age at which people learn a word (Brysbaert & Ellis, 2016). It is argued that early-acquired words are located at the centre of a semantic network and are considered to be more typical exemplars of their categories and therefore more accessible and robust relative to late-acquired words. This is supported by the findings that early-acquired words were better retrieved in picture naming by PWA (Bastiaanse et al., 2016), and

were more likely to be generated in semantic verbal fluency tasks (Sailor et al., 2011), and were recognised with higher accuracy by patients with Alzheimer's disease (Cuetos et al., 2010) than later-acquired words.

Concreteness describes the degree to which a word elicits a tangible mental representation. For instance, the word 'apple' is highly concrete because it can be directly experienced through the senses, while abstract terms like 'love' require explanation using other concepts or words. Concrete words are easier to retrieve and process than abstract words (Lee & Federmeier, 2008; Xiao et al., 2012; Zhang et al., 2006). Several theories have been put forward to explain this phenomenon. According to Paivio's (1990) dual-coding theory, concrete words are encoded through both verbal and imagery systems, whereas *abstract* words are represented exclusively in the verbal system, without associated imagery representations. As a result, abstract words tend to be more cognitively demanding to process compared to concrete words. The contextual availability theory, on the other hand, maintains that contextual information can facilitate word processing. Concrete words benefit from a higher availability of contextual cues than abstract words, which enhances their ease of processing (Schwanenflugel et al., 1988). Furthermore, clinical research has revealed selective impairments in the processing of concrete words compared to abstract words. PWA with multimodal semantic difficulties and lesions in the ventrolateral prefrontal cortex had difficulties in processing abstract words, which could be improved by providing a sentence context for the word (Hoffman et al., 2010). In addition, some studies found that persons with semantic variant primary progressive aphasia (svPPA) often exhibit impairments in processing concrete words compared to abstract words (Catricalà et al., 2014; Joubert et al., 2017). However, as highlighted by Mancano and Papagno (2023), findings on the concreteness effect in svPPA are mixed, with 6 out of 24 studies observing a concreteness effect, 14 studies reporting a reversal of the concreteness effect, and 4 studies finding no difference. This variability underscores the need for further research to clarify the factors (e.g. size and location of atrophy, the duration of the disease, semantic and grammatic category of the word) contributing to these divergent results. Nonetheless, word concreteness may still serve as an indicator of cognitive mechanisms associated with various neurological disorders.

Imageability refers to the ease with which a mental image can be generated in response to a given word. Research has shown that patients with Parkinson's disease demonstrate greater accuracy and faster response times in naming nouns compared to verbs, with the noun-verb discrepancy being more pronounced than in the control group, a difference that can be partially attributed to the imageability of

words (Bayram et al., 2021). Although imageability and concreteness may exhibit overlap for some lexical items, it is essential to recognise that notable distinctions exist between these two psycholinguistic properties. Some concrete objects that are rarely experienced (e.g. ‘armadillo’) are not highly imageable (Bird et al., 2001; Paivio et al., 1968), whereas some abstract words can be highly imageable (e.g. ‘traffic’ is not a very concrete word, but can evoke strong mental images).

Familiarity reflects the extent to which adults recognise and are acquainted with specific words. While familiarity scores often correlate with written word frequency, it is important to note that words with similar frequency, especially low-frequency words, do not necessarily share equivalent levels of familiarity among individuals. Familiarity has been shown to be a more effective predictor of word recognition latencies than frequency (Gernsbacher, 1984). Word frequency is contingent upon the relationship of a target lexical item to a reference corpus, which is usually based more on written than spoken language samples. In contrast, familiarity norms are ratings provided directly by native speakers of a language, reflecting the general distribution of words in different natural communication modes (Stadthagen-Gonzalez & Davis, 2006). Word familiarity serves as a significant predictor of naming accuracy in persons with chronic post-stroke aphasia, whose performance is notably better when responding to familiar words than unfamiliar ones (Alyahya et al., 2018).

Psycholinguistic norms were initially used to select stimuli in lexical decision and naming experiments, but their application has recently expanded to include measuring lexical development in second language learners (Berger et al., 2019; Crossley & Skalicky, 2019; Salsbury et al., 2011), and diachronic semantic change in language use (Hills & Adelman, 2015; Hollis et al., 2017; Li & Hills, 2021; Li & Siew, 2022; Sneffjella et al., 2019).

Lexical retrieval in aphasia discourse production involves multiple stages, including conceptual preparation, lemma selection, and phonological encoding; each of which can be affected differently by different types of brain damage. Psycholinguistic measures can provide a window into how these stages are disrupted. For example, words acquired early in life (low AoA) are often retrieved more easily, suggesting AoA may influence lemma selection or lexical access. Similarly, highly concrete and imageable words are associated with richer semantic representations, which may facilitate conceptual preparation. By examining how persons with aphasia perform on tasks involving words varying along these dimensions, researchers can better identify which stage(s) of lexical retrieval are impaired and tailor therapy accordingly.

Despite numerous studies that have delved into lexical psycholinguistic properties in single-word production or processing tasks, there has been relatively

less exploration regarding the lexical psycholinguistic characteristics during PWA discourse production via psycholinguistic norms except for two studies (Alyahya et al., 2021; Chen & Chang, 2024a). Alyahya et al. (2021) examined nouns and verbs produced in three discourse genres (descriptive, narrative, and procedural discourse) by 46 English participants with chronic post-stroke aphasia and 20 neurotypical adults. The two groups did not differ in word frequency and imageability, although neurotypical participants produced more word tokens and types than PWA. However, this study did not include other psycholinguistic properties such as AoA, concreteness, or familiarity. Chen and Chang (2024a) explored word concreteness, imageability, and familiarity in Mandarin PWA production of picture descriptions, story narratives, and procedural discourse. Relative to the control group, PWA used more concrete and familiar words in both picture descriptions and story narratives, and used more imageable words in picture descriptions than the control group.

Based on previous findings regarding how lexical psycholinguistic properties affect lexical retrieval and processing, it is anticipated that words in discourse produced by PWA would be earlier in AoA, higher in concreteness, imageability, and familiarity than those of the control group.

Effects of discourse elicitation tasks

Various methods have been employed to collect language samples for aphasia discourse analysis. Picture descriptions, story narratives, and procedural discourse are the three most widely used elicitation tasks and have been included in the protocol of AphasiaBank (MacWhinney et al., 2011; MacWhinney & Fromm, 2016). Previous research has found that elicitation methods can impose different cognitive and linguistic challenges on participants, especially for PWA, and consequently result in variations in their discourse production performance (Chen & Chang, 2024b; Deng et al., 2024; Deng et al., 2024; Fergadiotis et al., 2011; Fergadiotis & Wright, 2011). For example, Fergadiotis and Wright (2011) found that the control group displayed higher lexical diversity for story narratives than descriptions of sequential pictures, which in turn showed higher lexical diversity than descriptions of a single picture. However, the PWA group demonstrated lower lexical diversity in description of single pictures than both descriptions of sequential pictures and story narratives, and the latter two had no significant differences between them. In another study with young and ageing participants (Fergadiotis et al., 2011), it was found that both groups displayed higher lexical diversity in story narratives than picture descriptions, which in turn had higher lexical diversity than procedural discourse.

However, there have been few studies that examine variations of lexical psycholinguistic properties as a function of elicitation methods except for one study (Alyahya et al., 2021). In that study, only imageability of words was analysed, and there were no differences among picture descriptions, story narratives, and procedural discourse. It remains unknown how other lexical psycholinguistic properties vary in different discourse types. We hypothesise that picture descriptions are likely to elicit early-acquired, concrete, highly imageable, and familiar words; since this task involves describing tangible and often straightforward visual stimuli, such as observable objects and actions. Story narratives have the potential to elicit a broader range of vocabulary, including abstract terms and words acquired later in life, with varying levels of imageability and familiarity depending on the complexity of the story. Procedural discourse may involve more technical and specialised vocabulary, which may be less familiar and less imageable, but still includes a mix of concrete and abstract words, depending on the task being described.

The present study aims to examine: a) whether there are differences in the spoken language samples between PWA and control participants in terms of four lexical psycholinguistic properties (i.e. AoA, concreteness, imageability, and familiarity), b) whether different discourse elicitation tasks affect these lexical psycholinguistic properties, and c) whether the differences between PWA and control groups vary in magnitude as a function of discourse elicitation methods.

Method

Data

We used language samples from AphasiaBank (MacWhinney et al., 2011; MacWhinney & Fromm, 2016). The protocol contains five sections: Greetings, Picture Description, Story Narrative, Procedural Discourse, and Free Speech. There are four picture description tasks: *Broken window*, *Refused umbrella*, *Cat rescue*, and *Flood*. The Story Narrative section requires participants to tell the *Cinderella* story, and the Procedural Discourse section requires participants to describe the procedures for making a peanut butter and jelly sandwich. Details regarding the administration of these tasks and the transcripts can

be found on the AphasiaBank website (<https://aphasia.talkbank.org/protocol/english/>). All data from AphasiaBank is accessed and used in compliance with its ethical guidelines and participant consent protocols. For assessing lexical psycholinguistic properties and examining task variations, we selected the *Broken window* task from the set of picture description tasks. Unlike single-picture tasks such as *Cat Rescue* or *Flood*, *Broken window* requires describing a sequence of events across multiple images (e.g. a boy kicks a ball, which breaks the window of a living room in which a man is sitting), making it more comparable to story narrative tasks.

Only monolingual English-speaking participants were included in the analysis. PWA who produced insufficient language samples (e.g. a few words in each task) or who were reported to have depression were excluded from the analysis. Additionally, individuals with missing data on key demographic variables, such as age, education, or language background, were also excluded. Control participants were selected based on the same monolingual language criterion and were matched to the PWA group in terms of age and education to ensure comparability across groups. Transcripts of 68 PWAs were retrieved from the AphasiaBank database (see Table I). They were classified by Western Aphasia Battery-Revised (WAB-R) (Kertesz, 2006) as anomic ($n=29$), Broca's ($n=19$), conduction ($n=13$), transcortical motor ($n=2$), transcortical sensory ($n=1$), and Wernicke's ($n=4$) aphasia. Transcripts from 70 healthy control speakers were selected as a baseline for comparison with PWA productions. There were no significant differences between the PWA and control groups in terms of age, $t(115) = -1.71$, $p=0.09$ and years of education, $t(135) = 1.74$, $p=0.08$.

Psycholinguistic norms

Psycholinguistic norms were used to evaluate lexical characteristics in PWA discourse. To maximise the overlap between lexical items in the psycholinguistic norms and the words in our corpus, we selected four large-scale norms (see Table II) from several existing norms: one for each psycholinguistic property, i.e. age of acquisition (Kuperman et al., 2012), concreteness (Brysbaert et al., 2014), imageability (Scott et al., 2019), and familiarity (Scott et al., 2019).

Table I. Demographic information for all individuals.

Group	Sex	<i>n</i>	Age (years)		WAB AQ		Years of education	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	Male	20	51.2	11.7	–	–	15.2	2.1
	Female	50	51.6	11.3	–	–	15.6	2.1
PWA	Male	36	54.8	6.6	71.9	15.6	15.1	2.5
	Female	32	53.5	7.3	76.2	14.4	14.5	1.8

Note. WAB AQ = Western Aphasia Battery Aphasia Quotient, a scale from 0 to 100, with higher scores indicating better language abilities (Kertesz, 2006). All participants are from AphasiaBank (MacWhinney et al., 2011).

Table II. Psycholinguistic norms of AoA, concreteness, imageability, familiarity.

Psycholinguistic properties	Source	Size	Rating scale
AoA	Kuperman et al., 2012	30,121	–
Concreteness	Brysbaert et al., 2014	37,058	1 = abstract 5 = concrete
Imageability	Scott et al., 2019	5,553	1 = difficult to image 7 = easy to image
Familiarity	Scott et al., 2019	5,553	1 = unfamiliar 7 = familiar

Note. AoA = age of acquisition

Table III. Linguistic productivity measures of discourse samples for each task and group.

Tasks	Group	Tokens	Types	TTR	C	R
Story narrative	Control	617	198	0.35	0.83	8.07
	PWA	250	82	0.37	0.81	5.16
Procedural discourse	Control	99	42	0.45	0.82	4.21
	PWA	61	28	0.52	0.82	3.61
Picture description	Control	113	59	0.56	0.87	5.56
	PWA	69	32	0.49	0.82	3.82

Note. TTR = type-token ratios; C = log transformed TTR; R = root transformed TTR.

Data processing

In the present study, we first used the CLAN program (V 05-Mar-2022) to extract transcripts of each task (gem + swindow + n + d1 + t*PAR + t%mor + t%gra + f *.cha) and then to extract clear speech (flo + cr + t*@, no repetitions, repairs or other types of disfluencies as annotated). These processed transcripts were fed into a syntactic dependency parser for lemmatisation via the *UDPipe* package (Straka & Straková, 2017) in R (R Core Team, 2018). The accuracy for lemmatisation was around 96% (Straka & Straková, 2017). Lemmatisation was applied to ensure that different inflected forms of a word (e.g. run, runs, running) were treated as a single lexical item. This step was necessary to enhance the alignment between words in the discourse and psycholinguistic norms. We calculated some basic linguistic productivity measures for each task and group, including types, tokens, type-token ratios (TTR), log transformed TTR (i.e. C), and root transformed TTR (i.e. R) (see Table III). These measures provide information of lexical diversity, which is relevant for characterising differences in language use between groups. In particular, TTR and its variants help capture the range of vocabulary produced.

Each lemmatised word in the transcripts was cross-referenced with a corresponding lexical item in the norms if it exists. Coverage was calculated as the proportion of lemmatised word tokens in the transcripts that had a match in the norms. The coverage percentages of AoA and concreteness norms for each task were high whereas those for imageability and familiarity norms were relatively low due to the different sizes of these norms (see Table IV). However, it should be noted that we did not compare psycholinguistic lexical properties across different norms but only compared the same psycholinguistic property between two groups. In this sense, for all four norms,

the coverage percentages were similar across the two groups. Finally, we calculated the means of each psycholinguistic measure per participant per task.

Result

Given that psycholinguistic measures were obtained from three different studies with different scales, we did not compare the four psycholinguistic properties in a single model. Instead, psycholinguistic properties (i.e. AoA, concreteness, imageability, and familiarity) were each selected as dependent variables and fitted with a separate linear mixed-effects model using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2018). Participant groups (PWA versus healthy controls) and task types (picture description, story narrative, procedural discourse) were used as fixed factors. Four models were built to test all main effects and interactions. For each model, we ran the analysis with participants as random intercepts and used the Kenward-Roger degrees of freedom approximation (Halekoh & Hojsgaard, 2014) to estimate *p*-values for the fixed effects of interest. Here, we report main effects for multilevel factors or interactions using *F*-tests via the *Anova* function from the *car* package (Fox & Weisberg, 2019) in R (see Table V), rather than comparisons with the baseline level using *t*-tests.

There were significant main effects of groups and task types for AoA. In addition, there was a two-way interaction (group $\times$ task) (see Table V). To further examine the main effects and the interaction, we conducted pairwise multiple *t*-tests with Tukey adjustments. For the main effect of groups, healthy controls generally produced words that were acquired later in life (i.e. higher AoA, $M = 4.77$, 95% CI [4.73, 4.82]) than PWA ($M = 4.61$, 95% CI [4.56, 4.65]), $\beta = 0.17$, $SE = 0.03$, $t(136) = 5.13$, $p < 0.001$. Participants produced words with lower AoA in picture description ($M = 4.42$, 95% CI [4.37, 4.47])

Table IV. Coverage percentage of psycholinguistic norms for each task and group.

Psycholinguistic measures	Tasks	Groups	Coverage percentage
AoA	Picture description	Control	97.5%
		PWA	98.2%
	Procedural discourse	Control	98.9%
		PWA	99.4%
	Story narrative	Control	98.6%
		PWA	98.7%
Concreteness	Picture description	Control	99.8%
		PWA	99.7%
	Procedural discourse	Control	99.7%
		PWA	99.8%
	Story narrative	Control	99.6%
		PWA	99.5%
Imageability and familiarity	Picture description	Control	58.5%
		PWA	59.3%
	Procedural discourse	Control	66.1%
		PWA	65.1%
	Story narrative	Control	53.7%
		PWA	50.0%

Note. AoA = age of acquisition

Table V. Model details of lexical psycholinguistic measures in discourse production.

Psycholinguistic measures	Factors	<i>F</i>	<i>df</i>	<i>p</i>
AoA	Group	26.36	1	136
	Task	114.18	2	272
	Group × Task	11.35	2	272
Concreteness	Group	0.03	1	136
	Task	259.94	2	272
	Group × Task	5.85	2	272
Imageability	Group	1.74	1	136
	Task	142.33	2	272
	Group × Task	1.98	2	272
Familiarity	Group	34.86	1	136
	Task	84.67	2	272
	Group × task	6.02	2	272

Note. AoA = age of acquisition

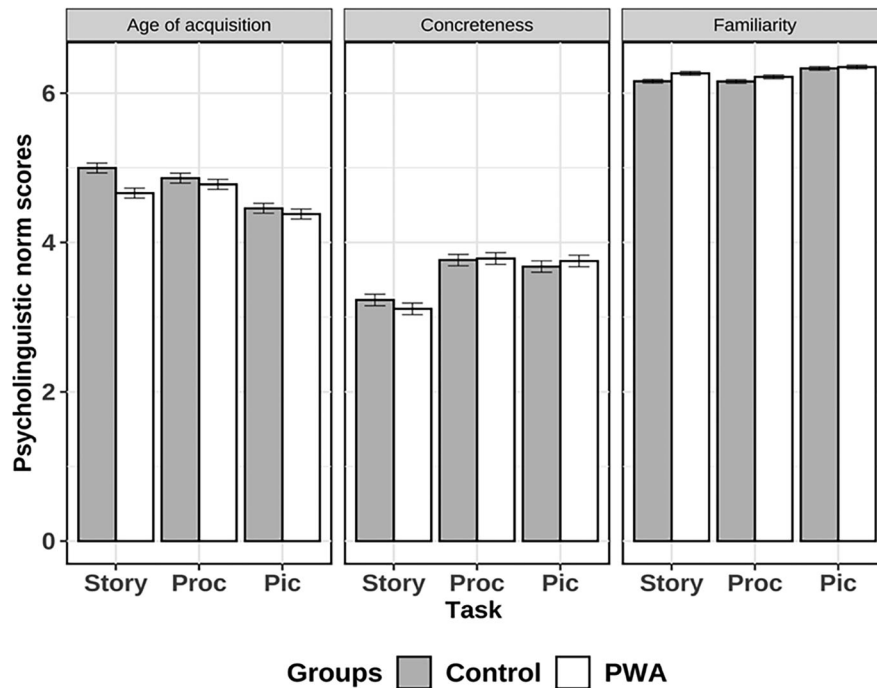


Figure 1. Group × task interactions of AoA, concreteness, and familiarity of words in discourse production.

Table VI. Comparisons of AoA, concreteness, and familiarity between PWA and control groups for each task.

Psycholinguistic measures	Task	β	SE	df	t	p
AoA	Story narrative	0.34	0.05	385.77	6.97	<0.001
	Procedural discourse	0.08	0.05	385.77	1.72	0.518
Concreteness	picture description	0.08	0.05	385.77	1.60	0.602
	Story narrative	0.12	0.06	289	2.12	0.278
	Procedural discourse	-0.02	0.06	289	-0.37	0.999
	picture description	-0.07	0.06	289	-1.34	0.763
Familiarity	Story narrative	-0.11	0.02	407.90	-5.90	<0.001
	procedural discourse	-0.06	0.02	407.90	-3.41	0.009
	picture description	-0.02	0.02	407.90	-1.02	0.910

Note. AoA = age of acquisition.

Table VII. Comparisons among different discourse tasks within each group.

Psycholinguistic measures	Groups	Tasks	β	SE	df	t	p
AoA	Control	Story narrative – procedural discourse	0.14	0.04	272	3.11	0.025
		Story narrative – picture description	0.54	0.04	272	12.37	<0.001
		Procedural discourse – picture description	0.40	0.04	272	9.26	<0.001
	PWA	Story narrative – procedural discourse	-0.12	0.04	272	-2.66	0.088
		Story narrative – picture description	0.28	0.04	272	6.33	<0.001
		Procedural discourse – picture description	0.40	0.04	272	8.99	<0.001
Concreteness	Control	Story narrative – procedural discourse	-0.54	0.04	272	-13.07	<0.001
		Story narrative – picture description	-0.45	0.04	272	-10.93	<0.001
		Procedural discourse – picture description	0.09	0.04	272	2.14	0.270
	PWA	Story narrative – procedural discourse	-0.68	0.04	272	-16.24	<0.001
		Story narrative – picture description	-0.64	0.04	272	-15.43	<0.001
		Procedural discourse – picture description	0.03	0.04	272	0.81	0.965
Familiarity	Control	Story narrative – procedural discourse	0.00	0.02	272	0.15	1.00
		Story narrative – picture description	-0.17	0.02	272	-9.67	<0.001
		Procedural discourse – picture description	-0.18	0.02	272	-9.82	<0.001
	PWA	Story narrative – procedural discourse	0.05	0.02	272	2.63	0.093
		Story narrative – picture description	-0.08	0.02	272	-4.67	<0.001
		Procedural discourse – picture description	-0.13	0.02	272	-7.30	<0.001

Note. AoA = age of acquisition; PWA = persons with aphasia.

compared to the story narrative ($M = 4.83$, 95% CI [4.78, 4.88]), $\beta = 0.41$, $SE = 0.03$, $t(272) = 13.20$, $p < 0.001$) and procedural discourse ($M = 4.82$, 95% CI [4.77, 4.87], $\beta = 0.40$, $SE = 0.03$, $t(272) = 12.90$, $p < 0.001$), with no significant differences between the latter two.

However, these main effects should be interpreted in the context of the task type by group interaction. Only in the story narrative task, healthy controls ($M = 5.00$, 95% CI [4.93, 5.06]) produced words with significantly larger AoA than PWA ($M = 4.66$, 95% CI [4.59, 4.73]) (see Figure 1 and Table VI).

Within the control participant group, words in the story narrative task were significantly larger in AoA than those in the procedural discourse task ($M = 4.86$, 95% CI [4.79, 4.93]), which in turn were larger than those in the picture description task ($M = 4.46$, 95% CI [4.39, 4.52], see Table VII). Within the PWA group, there were no significant differences between story narratives and procedural discourse ($M = 4.78$, 95% CI [4.71, 4.85]) but words produced in both tasks were larger in AoA than those in picture descriptions ($M = 4.38$, 95% CI [4.31, 4.45]).

As for concreteness, there was a main effect of task type, but no main effect of group. In addition, there was a two-way group $\times$ task interaction (see Table V). First, participants produced words with similar concreteness in procedural discourse ($M = 3.77$, 95% CI

[3.72, 3.83]) and picture description ($M = 3.71$, 95% CI [3.66, 3.77]), and both were higher than in story narratives ($M = 3.17$, 95% CI [3.11, 3.22]) (for the story narrative versus procedural discourse comparison, $\beta = -0.61$, $SE = 0.03$, $t[272] = -20.74$, $p < 0.001$; for the story narrative versus picture description comparison, $\beta = -0.54$, $SE = 0.03$, $t[272] = -18.66$, $p < 0.001$).

These findings from the main effects, however, must be understood in relation to the group $\times$ task type interaction. There were no differences between the two groups for any of the three tasks (for statistical details see Table VI). Participants in both control and PWA groups produced words with lower concreteness in narrative storytelling ($M_{Control} = 3.23$, 95% CI [3.15, 3.31]; $M_{PWA} = 3.11$, 95% CI [3.03, 3.19]) than in picture description ($M_{Control} = 3.68$, 95% CI [3.60, 3.75]; $M_{PWA} = 3.75$, 95% CI [3.67, 3.83]) and procedural discourse ($M_{Control} = 3.76$, 95% CI [3.69, 3.84]; $M_{PWA} = 3.78$, 95% CI [3.71, 3.86], for statistical details see Table VII).

For the imageability of words, there was only a main effect for task type (see Table V). The procedural discourse task ($M = 5.68$, 95% CI [5.60, 5.75]) elicited more imageable words than the picture description task ($M = 5.55$, 95% CI [5.47, 5.62], $\beta = -0.13$, $SE = 0.04$, $t[272] = 3.00$, $p < 0.001$), with words produced in the story narrative task being the least imageable ($M = 4.99$, 95% CI [4.91, 5.06], for

Table VIII. Correlations between word psycholinguistic properties and WAB-R subtest scores.

	AQ	Speech	Comp	Rep	Naming	AoA	Conc	Fam
AQ								
Speech	0.89****							
Comp	0.7****	0.5****						
Rep	0.83****	0.64****	0.46****					
Naming	0.79****	0.55****	0.55****	0.53****				
AoA	0.21**	0.19**	0.09	0.17*	0.18*			
Conc	-0.03	-0.15*	0.09	-0.01	0.10	-0.23***		
Fam	-0.09	-0.10	0.01	-0.09	-0.07	-0.52****	0.00	
Imag	0.06	-0.09	0.15*	0.08	0.14*	-0.18*	0.79****	0.06

Note. Word psycholinguistic properties: AoA = age of acquisition; Conc = Concreteness; Fam = familiarity; Imag = imageability; WAB-R = Western Aphasia Battery - Revised. WAB-R subtest scores: AQ = aphasia quotient; Speech = spontaneous speech; Comp = comprehension; Rep = repetition; and Naming.

the story narrative versus procedural discourse comparison, $\beta = -0.69$, $SE = 0.04$, $t[272] = -15.90$, $p < 0.001$; for the story narrative versus picture description comparison, $\beta = -0.56$, $SE = 0.04$, $t[272] = -12.90$, $p < 0.001$).

In terms of familiarity, we found main effects for group, and task type and there was also a group $\times$ task interaction (see Table V). The words in discourse produced by PWA ($M = 6.28$, 95% CI [6.26, 6.29]) have higher familiarity scores than those used by healthy controls ($M = 6.22$, 95% CI [6.20, 6.23], $\beta = -0.06$, $SE = 0.01$, $t[136] = -5.90$, $p < 0.001$). In addition, the picture description task ($M = 6.34$, 95% CI [6.32, 6.36]) elicited words with higher familiarity scores than the story narrative task ($M = 6.21$, 95% CI [6.20, 6.23], $\beta = -0.13$, $SE = 0.01$, $t[272] = -10.11$, $p < 0.001$) and the procedural discourse ($M = 6.19$, 95% CI [6.17, 6.21], $\beta = -0.15$, $SE = 0.01$, $t[272] = -12.09$, $p < 0.001$).

While main effects are presented, their interpretation should be guided by the significant group $\times$ task type interaction. There were no differences between the two groups for the picture description task, but in procedural discourse ($M_{Control} = 6.16$, 95% CI [6.13, 6.18]; $M_{PWA} = 6.22$, 95% CI [6.19, 6.24]) and story narrative tasks ($M_{Control} = 6.16$, 95% CI [6.14, 6.19]; $M_{PWA} = 6.27$, 95% CI [6.24, 6.29]), familiarity scores were higher for PWA than healthy controls (for statistical details see Table VI). Participants in both control and PWA groups produced words with higher familiarity in picture descriptions ($M_{Control} = 6.33$, 95% CI [6.31, 6.36]; $M_{PWA} = 6.35$, 95% CI [6.33, 6.38]) than in story narratives and procedural discourse (for statistical details see Table VII).

We have also conducted a series of correlation analyses examining the relationships between psycholinguistic variables and WAB-R subtest scores using data from AphasiaBank (see Table VIII). Significant correlations were observed between AoA and several WAB measures, including the Aphasia Quotient (AQ), spontaneous speech production, repetition, and naming. In addition, concreteness was significantly correlated with spontaneous speech production, while imageability showed significant correlations with both verbal comprehension and naming.

Discussion

In the present study, we examined four psycholinguistic lexical properties, namely AoA, concreteness, imageability, and familiarity, in PWA discourse production using existing large-scale psycholinguistic norms in English. Our findings revealed distinctive patterns of lexical psycholinguistic properties in discourse production by PWA relative to the control group. In addition, the task types resulted in variations in lexical psycholinguistic properties in discourse production.

Impacts of aphasia on word psycholinguistic properties in discourse production

PWA discourse showed distinctive patterns in terms of AoA and familiarity when compared to the control group. First, the words used by PWA were, on average, acquired earlier in life than those used by the control group. Note there are also important task-induced variations. Only in story narratives, did the PWA group exhibit words with lower AoA compared to the control group. Our findings provide corroborating evidence for the hypothesis that words acquired early in language development occupy a central position within the semantic network as prototypical exemplars of their respective semantic categories, making them more readily accessible and resilient compared to words acquired later in life. Prior research on single-word production also demonstrates that early-acquired words exhibit greater preservation among PWA in picture naming (Bastiaanse et al., 2016), as well as in verbal fluency among patients with Alzheimer's disease (Sailor et al., 2011). The novelty of our research lies in demonstrating the AoA effect in discourse production with high ecological validity.

Second, compared with the control group, PWA tend to generate words with higher familiarity in procedural discourse and story narrative tasks, suggesting that these words are easier for PWA to retrieve and process in the two tasks. Previous naming research has also found that familiarity appeared as one of the significant predictors of naming accuracy in persons with chronic post-stroke aphasia (Alyahya et al., 2018). When a word is highly familiar, it likely benefits from more robust

neural connections, making it more resilient to brain damage. Therefore, PWA may prefer familiar words because these words are more readily accessible in their impaired lexicon.

The tendency to use early-acquired and familiar words could also reflect a compensatory strategy that PWA employ to maintain communication despite linguistic challenges. On the other hand, as prior research in semantic treatment (e.g. Kiran, 2008; Kiran & Thompson, 2003) has shown, targeting atypical items of a category (e.g. hammock versus chair in the furniture category) can generalise to improved access to more typical items. These atypical items are often characterised by lower familiarity and later age of acquisition, suggesting that incorporating such stimuli in therapy may support broader lexical-semantic activation and recovery. Thus, therapies designed to enhance word retrieval in PWA might benefit from incorporating a balanced approach. Speech-language pathologists (SLPs) could tailor interventions by initially leveraging early-acquired and familiar words to establish communicative success and progressively incorporating later-acquired, less familiar, and less typical vocabulary to maximise generalisation and recovery.

However, we did not find significant differences between the two groups in terms of word concreteness and imageability in English discourse production. For word concreteness, a previous Mandarin study (Chen & Chang, 2024a) also found that Mandarin PWA produced single-character and multi-character words of similar concreteness relative to the control group in the *Broken window* picture description task. However, Mandarin PWA produced more concrete multi-character words than the control group in the story narratives of *Crying wolf* and *Tortoise and hare*. We speculate that the differences between this study and the previous one could be attributed to differences in psycholinguistic norms used, the specific story used in eliciting narratives or differences in language systems, which need to be further investigated in future cross-language studies.

For imageability, a similar study on English discourse production failed to find differences between PWA and the control group in all three different tasks (Alyahya et al., 2021). But the Mandarin study (Chen & Chang, 2024a) found differences in word imageability between the two groups for the *Broken window* picture description but not for the story narrative and procedural description tasks. The comparisons of findings across different languages (Mandarin and English) and specific tasks suggest that differences in psycholinguistic properties between PWA and controls may be influenced by both the psycholinguistic norms used in a given language and the specific tasks employed to elicit discourse. This highlights the need for variety in tasks for assessment of discourse in PWA.

Task type effects

Story narratives generally elicited words with higher AoA and lower concreteness, imageability, and familiarity relative to picture descriptions and comparable AoA and familiarity and lower concreteness and imageability, relative to procedural discourse. Previous research also found that propositional density was greater in narrative discourse than in procedural discourse and picture descriptions. It was argued that story narratives are the most sensitive elicitation method for evaluating the depth of vocabulary and content richness (Stark, 2019). Story narratives require participants to maintain the flow of the story by utilising lexical resources that express temporal, spatial, and causal relationships. Such lexical resources involve words acquired later in life and with low concreteness and imageability.

Additionally, the differences between PWA and the control group were significant for AoA only in story narratives and for familiarity in both story narratives and procedural discourse. Thus, narrative discourse may be the most effective means of evaluating impairments or changes in lexical psycholinguistic properties. We attributed the discrepancy between the two groups in story narratives but not in picture description to the fact that the cognitive load in the story narrative task is higher than picture descriptions; because it requires participants to organise thoughts logically over time and consequently elicits more complex language structures that include words acquired later in life and with low familiarity. We speculate that procedural discourse tasks require participants to explain the steps involved in performing a task, including sequencing, problem-solving, and recalling specific actions in order. It thus imposes higher cognitive demands than the picture description task. But, the memory load for procedural description is lower than that for the story narrative because the task involves what participants do in their daily life. For PWA, the story narrative and procedure description tasks with little visual information support (c.f. picture descriptions) are cognitively challenging and force them to simplify their production to shift more cognitive resources to memory, planning, and organisation required by these tasks. This simplification could account for the differences in word AoA and familiarity between the two groups.

One clinical implication here is that, given that narrative discourse, relative to picture descriptions, imposes a higher memory load (both long-term and working memory) as well as efforts relevant to executive function, such as planning and organisation, it may pose challenges (floor effects) in validity or effectiveness for participants with concurrent impairments in these cognitive domains. Thus, researchers and SLPs should be cautious when choosing different discourse tasks for eliciting spoken discourse from persons with aphasia.

There are some limitations to the present study. Although we selected large-scale norms, these norms did not cover all the lexical items in the discourse production. Future research could utilise state-of-the-art semantic similarity algorithms to extend psycholinguistic norms and minimise the impact of coverage issues (Wang & Xu, 2023). In addition, we used only monolingual English discourse data in our present research. This restricts its generalisability to multilingual populations and other language systems. Future research should incorporate discourse produced by multilingual participants across different languages to better capture variations of psycholinguistic properties in aphasia discourse production.

Conclusion

In conclusion, PWA used words with lower AoA in the story narrative task and higher familiarity in both the story narrative task and the procedural description task than the control group, indicating their difficulties in retrieving later-acquired, less familiar lexical items. Story narratives are more sensitive in eliciting lexical psycholinguistic differences between the PWA and the control group and thus can better detect impairments in this respect compared to picture descriptions. Psycholinguistic norms provided us with important information about the lexical difficulties in PWA and therefore should be incorporated into aphasia discourse assessment frameworks.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Humanities and Social Sciences Foundation of the Ministry of Education, PRC (23YJC740002) and National Language Commission (ZDI145-136).

Data availability statement

R codes for processing the data are available at https://osf.io/jysx8/?view_only=951c03787c7e4deaa81efe215fd1452f

References

- Aggujaro, S., Crepaldi, D., Pistarini, C., Taricco, M., & Luzzatti, C. (2006). Neuro-anatomical correlates of impaired retrieval of verbs and nouns: Interaction of grammatical class, imageability and actionality. *Journal of Neurolinguistics*, 19(3), 175–194. <https://doi.org/10.1016/j.jneuroling.2005.07.004>
- Alyahya, R. S. W., Halai, A., Conroy, P., & Lambon-Ralph, M. A. (2018). The relationship between the multidimensionality of aphasia and psycholinguistic features. *Aphasiology*, 32(sup1), 5–6. <https://doi.org/10.1080/02687038.2018.1486374>
- Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2021). Content Word Production during Discourse in Aphasia: Deficits in Word Quantity, Not Lexical-Semantic Complexity. *Journal of Cognitive Neuroscience*, 33(12), 2494–2511. https://doi.org/10.1162/jocn_a_01772
- Bastiaanse, R., Wieling, M., & Wolhuis, N. (2016). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology*, 30(11), 1221–1239. <https://doi.org/10.1080/02687038.2015.1100709>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bayram, E., Yilmaz, R., Qiu, Y., Yalap, O. E., Aydin, O., Ergenc, H. I., & Akbostanci, M. C. (2021). The effect of Subthalamic nucleus deep brain stimulation on verb and noun naming in Turkish-Speaking Parkinson's disease patients. *Brain and Language*, 212, 104865. <https://doi.org/10.1016/j.bandl.2020.104865>
- Berger, C. M., Crossley, S. A., & Kyle, K. (2019). Using Native-Speaker Psycholinguistic Norms to Predict Lexical Proficiency and Development in Second-Language Production. *Applied Linguistics*, 40(1), 22–42. <https://doi.org/10.1093/applin/amx005>
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc.*, 33(1), 73–79. <https://doi.org/10.3758/BF03195349>
- Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, 30(7), 489–518. <https://doi.org/10.3109/02699206.2016.1145740>
- Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology*, 30(11), 1240–1263. <https://doi.org/10.1080/02687038.2015.1106439>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Catricalà, E., Della Rosa, P. A., Plebani, V., Vigliocco, G., & Cappa, S. F. (2014). Abstract and concrete categories? Evidences from neurodegenerative diseases. *Neuropsychologia*, 64, 271–281. <https://doi.org/10.1016/j.neuropsychologia.2014.09.041>
- Chen, J., & Chang, H. (2024a). Assessing Lexical Psycholinguistic Properties in Mandarin Discourse Production by Patients with Aphasia. In M. Dong, J.-F. Hong, J. Lin, & P. Jin (Eds.), *Lecture Notes in Computer Science*. (0?; pp. 11–22). Springer Nature. https://doi.org/10.1007/978-981-97-0583-2_2
- Chen, J., & Chang, H. (2024b). Core Lexicon analysis of spoken discourse production by Mandarin Chinese speakers with Aphasia. *Aphasiology*, 38(11), 1766–1785. <https://doi.org/10.1080/02687038.2024.2325169>
- Chen, J., & Chang, H. (2025). Lexical network properties in aphasic discourse production. *Clinical Linguistics & Phonetics*, 1–19. <https://www.tandfonline.com/doi/abs/10.1080/02699206.2025.2507049> <https://doi.org/10.1080/02699206.2025.2507049>
- Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 16(6), 1084–1096. <https://doi.org/10.1037/0278-7393.16.6.1084>
- Crossley, S. A., & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of Salisbury, Crossley & McNamara (2011). *Language Teaching*, 52(3), 385–405. <https://doi.org/10.1017/S026144817000362>
- Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. *Neuropsychologia*, 48(11), 3329–3334. <https://doi.org/10.1016/j.neuropsychologia.2010.07.017>

- 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
- Deng, B.-M., Liang, L.-S., Zhao, J.-X., Zheng, H.-Q., & Hu, X.-Q. (2024). Correct Information Unit Analysis in Different Discourse Tasks Among Persons With Anomic Aphasia Based on Mandarin AphasiaBank. *American Journal of Speech-Language Pathology*, 33(2), 800–813. https://doi.org/10.1044/2023_AJSLP-23-00217
- Dipper, L., Marshall, J., Boyle, M., Botting, N., Hersh, D., Pritchard, M., & Cruice, M. (2021). Treatment for improving discourse in aphasia: A systematic review and synthesis of the evidence base. *Aphasiology*, 35(9), 1125–1167. <https://doi.org/10.1080/02687038.2020.1765305>
- Dipper, L. T., & Pritchard, M. (2017). Discourse: Assessment and Therapy. In F. D. M. Fernanda Dreux M. Fernandes (Ed.), *Advances in Speech-language Pathology*. IntechOpen. <https://doi.org/10.5772/intechopen.69894>
- Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, 25(11), 1414–1430. <https://doi.org/10.1080/02687038.2011.603898>
- Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, 25(10), 1261–1278. <https://doi.org/10.1080/02687038.2011.606974>
- Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring Lexical Diversity in Narrative Discourse of People With Aphasia. *American Journal of Speech-Language Pathology*, 22(2), S397–S408. [https://doi.org/10.1044/1058-0360\(2013\)12-0083](https://doi.org/10.1044/1058-0360(2013)12-0083)
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage, Thousand Oaks CA.
- Fromm, D., Katta, S., Paccione, M., Hecht, S., Greenhouse, J., MacWhinney, B., & Schnur, T. T. (2021). A Comparison of Manual Versus Automated Quantitative Production Analysis of Connected Speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, 64(4), 1271–1282. https://doi.org/10.1044/2020_JSLHR-20-00561
- Fromm, D., MacWhinney, B., & Thompson, C. K. (2020). Automation of the Northwestern Narrative Language Analysis System. *Journal of Speech, Language, and Hearing Research: JSLHR*, 63(6), 1835–1844. https://doi.org/10.1044/2020_JSLHR-19-00267
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology. General*, 113(2), 256–281. <https://doi.org/10.1037//0096-3445.113.2.256>
- Halekoh, U., & Hojsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbkrtest. *Journal of Statistical Software*, 59(9), 1–30.
- Heikkola, L. M., Kuzmina, E., & Jensen, B. U. (2022). Predictors of object naming in aphasia: does cognitive control mediate the effects of psycholinguistic variables?. *Aphasiology*, 36(11), 1275–1292. <https://doi.org/10.1080/02687038.2021.1950607>
- Hills, T. T., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition*, 143, 87–92. <https://doi.org/10.1016/j.cognition.2015.06.009>
- Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2010). Ventrolateral Prefrontal Cortex Plays an Executive Regulation Role in Comprehension of Abstract Words: Convergent Neuropsychological and Repetitive TMS Evidence. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 30(46), 15450–15456. <https://doi.org/10.1523/JNEUROSCI.3783-10.2010>
- Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *Quarterly Journal of Experimental Psychology* (2006), 70(8), 1603–1619. <https://doi.org/10.1080/17470218.2016.1195417>
- Hsu, C.-J., & Thompson, C. K. (2018). Manual Versus Automated Narrative Analysis of Agrammatic Production Patterns: The Northwestern Narrative Language Analysis and Computerized Language Analysis. *Journal of Speech, Language, and Hearing Research: JSLHR*, 61(2), 373–385. https://doi.org/10.1044/2017_JSLHR-L-17-0185
- Joubert, S., Vallet, G. T., Montembeault, M., Boukadi, M., Wilson, M. A., Laforce, R., Jr., Rouleau, I., & Brambati, S. M. (2017). Comprehension of concrete and abstract words in semantic variant primary progressive aphasia and Alzheimer's disease: A behavioral and neuroimaging study. *Brain and Language*, 170, 93–102. <https://doi.org/10.1016/j.bandl.2017.04.004>
- Kertesz, A. (2006). *Western Aphasia Battery-Revised (Revised)*. Pearson.
- Kiran, S. (2008). Typicality of Inanimate Category Exemplars in Aphasia Treatment: Further Evidence for Semantic Complexity. *Journal of Speech, Language, and Hearing Research: JSLHR*, 51(6), 1550–1568. [https://doi.org/10.1044/1092-4388\(2008\)07-0038](https://doi.org/10.1044/1092-4388(2008)07-0038)
- Kiran, S., & Thompson, C. K. (2003). The Role of Semantic Complexity in Treatment of Naming Deficits. *Journal of Speech, Language, and Hearing Research*, 46(3), 608–622. [https://doi.org/10.1044/1092-4388\(2003\)048](https://doi.org/10.1044/1092-4388(2003)048)
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990. <https://doi.org/10.3758/s13428-012-0210-4>
- Lee, C.-L., & Federmeier, K. D. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language*, 104(2), 145–158. <https://doi.org/10.1016/j.bandl.2007.06.002>
- Li, Y., & Siew, C. S. Q. (2022). Diachronic semantic change in language is constrained by how people use and learn language. *Memory & Cognition*, 50(6), 1284–1298. <https://doi.org/10.3758/s13421-022-01331-0>
- Li, Y., & Hills, T. T. (2021). Language patterns of outgroup prejudice. *Cognition*, 215, 104813. <https://doi.org/10.1016/j.cognition.2021.104813>
- MacWhinney, B., & Fromm, D. (2016). AphasiaBank as BigData. *Seminars in Speech and Language*, 37(1), 10–22. <https://doi.org/10.1055/s-0036-1571357>
- 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>
- MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, 24(6-8), 856–868. <https://doi.org/10.1080/02687030903452632>
- Mancano, M., & Papagno, C. (2023). Concrete and Abstract Concepts in Primary Progressive Aphasia and Alzheimer's Disease: A Scoping Review. *Brain Sciences*, 13(5), 765. <https://doi.org/10.3390/brainsci13050765>
- Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11), 1372–1392. <https://doi.org/10.1080/02687038.2011.584690>
- Marshall, J., Pring, T., & Chiat, S. (1998). Verb retrieval and sentence production in aphasia. *Brain and Language*, 63(2), 159–183. <https://doi.org/10.1006/brln.1998.1949>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Clarendon Press.

- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76(1), Suppl:1–25. <https://doi.org/10.1037/h0025327>
- Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders*, 52(6), 689–732. <https://doi.org/10.1111/1460-6984.12318>
- Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2018). Psychometric properties of discourse measures in aphasia: Acceptability, reliability, and validity. *International Journal of Language & Communication Disorders*, 53(6), 1078–1093. <https://doi.org/10.1111/1460-6984.12420>
- R Core Team. (2018). *R: A language and environment for statistical computing* [Manual]. <https://www.R-project.org/>
- Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative Analysis of Aphasic Sentence Production: Further Development and New Data. *Brain and Language*, 72(3), 193–218. <https://doi.org/10.1006/brln.1999.2285>
- Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, 37(3), 440–479. [https://doi.org/10.1016/0093-934X\(89\)90030-8](https://doi.org/10.1016/0093-934X(89)90030-8)
- Sailor, K. M., Zimmerman, M. E., & Sanders, A. E. (2011). Differential impacts of age of acquisition on letter and semantic fluency in Alzheimer's disease patients and healthy older adults. *Quarterly Journal of Experimental Psychology* (2006), 64(12), 2383–2391. <https://doi.org/10.1080/17470218.2011.596660>
- Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, 27(3), 343–360. <https://doi.org/10.1177/0267658310395851>
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5), 499–520. [https://doi.org/10.1016/0749-596X\(88\)90022-8](https://doi.org/10.1016/0749-596X(88)90022-8)
- Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, 51(3), 1258–1270. <https://doi.org/10.3758/s13428-018-1099-3>
- Shewan, C. M. (1988). Expressive language recovery in aphasia using the Shewan Spontaneous Language Analysis (SSLA) system. *Journal of Communication Disorders*, 21(2), 155–169. [https://doi.org/10.1016/0021-9924\(88\)90003-2](https://doi.org/10.1016/0021-9924(88)90003-2)
- Sneffjella, B., Génereux, M., & Kuperman, V. (2019). Historical evolution of concrete and abstract language revisited. *Behavior Research Methods*, 51(4), 1693–1705. <https://doi.org/10.3758/s13428-018-1071-2>
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, 38(4), 598–605. <https://doi.org/10.3758/BF03193891>
- Stark, B. C. (2019). A Comparison of Three Discourse Elicitation Methods in Aphasia and Age-Matched Adults: Implications for Language Assessment and Outcome. *American Journal of Speech-Language Pathology*, 28(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265
- Straka, M., Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 88–99. <https://doi.org/10.18653/v1/K17-3009>
- Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., Schneider, S. L., & Ballard, K. J. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language*, 51(1), 124–129.
- Wachal, R. S., & Spreen, O. (1973). Some Measures of Lexical Diversity in Aphasic and Normal Language Performance. *Language and Speech*, 16(2), 169–181. <https://doi.org/10.1177/002383097301600207>
- Wang, T., & Xu, X. (2023). The good, the bad, and the ambivalent: Extrapolating affective values for 38,000+ Chinese words via a computational model. *Behavior Research Methods*, 56(6), 5386–5405. <https://doi.org/10.3758/s13428-023-02274-3>
- Wilshire, C. E., Lukkien, C. C., & Burmester, B. R. (2014). The sentence production test for aphasia. *Aphasiology*, 28(6), 658–691. <https://doi.org/10.1080/02687038.2014.893555>
- Wright, H., Silverman, S., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, 17(5), 443–452. <https://doi.org/10.1080/02687030344000166>
- Xiao, X., Zhao, D., Zhang, Q., & Guo, C. (2012). Retrieval of concrete words involves more contextual information than abstract words: Multiple components for the concreteness effect. *Brain and Language*, 120(3), 251–258. <https://doi.org/10.1016/j.bandl.2011.09.006>
- Zhang, Q., Guo, C., Ding, J., & Wang, Z. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language*, 96(1), 59–68. <https://doi.org/10.1016/j.bandl.2005.04.004>