APHASIOLOGY
2024, VOL. 38, NO. 11, 1766–1785
https://doi.org/10.1080/02687038.2024.2325169

Routledge
Taylor & Francis Group

Core Lexicon analysis of spoken discourse production by Mandarin Chinese speakers with Aphasia

Juqiang Chen [a,b] and Hui Chang [a,b]

[a]School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, P. R. China; [b]National Research Centre for Language and Well-being, P.R. China

ABSTRACT

Background: Spoken discourse production has been increasingly recognised as an important source for assessing language competence of persons with aphasia (PWA). Most linguistic measures rely on transcripts of spoken discourse, so they have seldom been used in clinical practice. Thus, there is a need for standardized and norm-referenced measures that can be easily applied in clinical settings.

Aims: The present study aims to develop core lexicon lists for Mandarin Chinese PWA using AphasiaBank data elicited with different discourse production tasks and test their validity in detecting language deficiency.

Methods & Procedures: We selected seven semispontaneous discourse tasks within three major task types, i.e., picture descriptions, procedural discourse and story narrative, for core lexicon list development and for examining task type effects. Transcripts from 61 control speakers were split into two sets with age – and education matched: one set (n = 43) as a normative sample used to generate the core lexicon lists and the other (n = 18) as a healthy control group served as a baseline for comparison with PWA productions. Transcripts of 18 age- and education- matched PWA were retrieved from the AphasiaBank database. We selected the top 30 words with the widest distribution in the normative sample to generate core lexicon lists. Then for both PWA and control groups, we calculated agreement percentage as core lexicon scores.

Outcomes & Results: Core lexicon scores of PWA were significantly lower than the control group participants for all seven tasks and positively correlated with aphasia severity for two story narrative tasks, suggesting that story narrative tasks were more effective in assessing discourse productions compared with picture and procedural descriptions. In addition, core lexicon scores also correlated with lexical diversity (R scores) and discourse informativeness (Correct Information Units) for all seven tasks.

Conclusions & Implications: Core lexicon scores can effectively distinguish discourse production by Mandarin PWA and healthy controls and they reflect lexical diversity and informativeness of discourse. Our findings, thus, have clinical implications in that these lists could serve as an alternative transcription-less approach in quick Mandarin PWA discourse production assessment.

ARTICLE HISTORY
Received 28 March 2023
Revised 23 February 2024
Accepted 26 February 2024

KEYWORDS
Core lexicon; Mandarin discourse production; persons with aphasia

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

1. Introduction

Spoken discourse production has been increasingly recognised as an important source for assessing language competence of persons with aphasia (PWA). Some linguistic measures, such as type-token ratio (TTR), or mean length of utterance, can be automated, if transcripts of relevant discourse productions are available, to assist PWA discourse analysis. Other types of measures, such as main concepts (Dalton & Richardson, 2015; Dalton & Richardson, 2019; Kong, 2009; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016; Richardson et al., 2021), content units (Yorkston & Beukelman, 1980), Correct Information Units (CIUs, Nicholas & Brookshire, 1993) and main-events (Capilouto et al., 2005) require subjective judgements by trained annotators. While these measures are informative in revealing different aspects of language competence and deficiencies of PWA, they often involve labour-intensive and time-consuming processes, such as transcription and annotation, and have seldom been used in real-time clinical practices. Thus, there is a need for standardized and norm-referenced measures that can be easily applied in clinical settings.

Recently core lexicon analysis has been developed and proposed as a feasible and easy-to-use measure for PWA assessment (Dalton & Richardson, 2015; Kim, Berube et al., 2022; Kim et al., 2019; Kim & Wright, 2020). The basic assumption of core lexicon analysis is that functional communication abilities during discourse production tasks can be evaluated via a set of typical vocabularies or a core lexicon (MacWhinney et al., 2010). A core lexicon list based on productions by healthy controls for a specific task could index PWA's ability to retrieve words at discourse level for that task. Such task-specific and norm-referenced core lexicon lists are useful in clinical settings as clinicians can assess PWA's spoken discourse production by checking the list in real time. Core lexicon lists have proved to be effective in distinguishing PWA from control groups as well as subtypes of aphasia (S. G. Dalton & Richardson, 2015; Kim et al., 2021, 2019). They also significantly correlate with linguistic measures, such as correct information units, lexical diversity, syntactic complexity (Kim & Wright, 2020), main concepts (Dalton & Richardson, 2015), thematic units, and coherence measures (Kim & Wright, 2020), suggesting good concurrent validity. The core lexicon lists have also achieved acceptable interrater reliability even for raters with very limited experience and training time (Kim & Wright, 2020). Thus, a well-built core lexicon list saves the time for recording, transcribing and annotating discourse production as required by most traditional measures, such as TTR and is more objective than other measures, such as main concepts, which require long-term trainings and are difficult to maintain inter-rater consistency.

Several English core lexicon lists have been constructed in previous studies for English-native PWA. Dalton and Richardson selected spoken discourse production data of the Broken Window task from the AphasiaBank by 92 healthy participants as a normative sample and created a core lexicon list of 24 lemmas that were produced by 50% or more people in the group (Dalton & Richardson, 2015). Then they calculated the number of lemmas produced by PWA (n= 235) and another healthy group (n = 166) that matched the core lexicon list (also called CoreLex) and found a significantly difference between PWA and healthy controls as well as between Broca's aphasia and other aphasia subtypes (e.g., anomic, conduction and Wernicke's, but no significant differences among these latter subtypes) in terms of CoreLex. A novel contribution of the study is that for PWA and for all within subgroups, CoreLex positively correlated with main concept scores (i.e., essential concepts of a narrative).

Kim and colleagues in another study constructed age-based core lexicon lists for content words (nouns, verbs, adjectives, and adverbs; Kim et al., 2019) and function words (Kim et al., 2021) using two storytelling tasks: Good Dog Carl (Day, 1985) and Picnic (McCully, 1984). For the content word lists, they found significant positive correlations between core verbs and Aphasia Quotients (AQs, as in Western Aphasia Battery-Revised, WAB, Kertesz, 2006) but no correlations between other word types and AQs (Kim et al., 2019). For the function words, PWA produced fewer core function words than the control group and there was a correlation between core function word use and aphasia severity. In addition, core function words can distinguish non-fluent PWA from fluent PWA, as the former group produced fewer function words than the latter (Kim et al., 2021).

Most previous discourse production studies have focused on English PWA. Only a few studies have examined Chinese PWA discourse productions. Chinese is an isolating language, in which most words cannot be analyzed into suffixes or prefixes, typologically different from English (Cui & Sung, 2022; Li & Thompson, 1989). One Chinese syllable is orthographically represented by one character and generally corresponds to one morpheme. The majority of Chinese words comprises of one or more characters/syllables: 6% of Chinese words are single-character and 73% two-character (i.e., compound words; Tse et al., 2017). In terms of grammar, unlike English, Chinese does not require verb conjugation, subject-verb agreement, pluralization, case/gender specification, and the use of articles, but it relies on context, adverbs, or nouns, such as *yijing* (已经), *jiangyao* (将要), *le* (了), *zhe* (着), and *guo* (过) to indicate tense and aspect, particles such as *a* (啊), *ya* (呀), *ba* (吧), *bei* (呗), *ma* (吗), *ne* (呢), to form questions and signpost certain emotions or tone of speech (Cui & Sung, 2022).

Most studies that investigated Chinese spoken discourse production employed classic lexical and syntactic measures (Lai et al., 2019; Law et al., 2018) or main concept analysis (Deng et al., 2021). Lai and colleagues compared discourse production of 12 anomia Mandarin PWA and healthy controls using single-picture description, sequential picture description and story narrative tasks. They found that the anomic aphasia group was lower than the control group in terms of mean length of utterance in words, the moving average type-token ratio, words per minute and proposition density. Similarly, based on a large spoken database *Cantonese AphasiaBank*, it was found that control group participants had larger vocabulary size than the fluent PWA in personal story narrations (Law et al., 2018), but the distribution of parts-of-speech, frequency of occurrence, and the ratio of concrete to abstract items in major open word classes were all similar.

So far there has been only one core lexicon study on Mandarin PWA discourse production. Jiang and colleagues developed core noun and verb lists for three different task types, i.e., picture descriptions, story narrative and procedural discourse, using Mandarin AphasiaBank data (Jiang et al., 2023). PWA were found to produce fewer core words than participants in the control group, but there was no correlation between core lexicon use and the severity of aphasia. As a pioneering work on this issue, their research generally supports that core lexicon analysis is feasible and applicable in assessing Mandarin spoken discourse production. However, it remains unresolved whether/how core lexicon scores correlate with other linguistic measures, such as lexical diversity and discourse informativeness measures, such as Correct Information Units (Nicholas & Brookshire, 1993). Methodologically speaking, high frequency words were extracted as core words (in Jiang et al., 2023), which could be biased if a single word was used significantly more frequently than others by a particular participant. In addition, core

words extracted from specific tasks within a task type were aggregated in calculating core lexicon scores. For example, in their study, the Broken Window, the Refused Umbrella, the Cat Rescue and the Flood tasks were combined for the Picture Description task and the Tortoise and the Hare as well as the Cry Wolf for the Story Narrative task. However, given that a core lexicon list is not only task-specific but also sensitive to elicitation materials. It is likely that there were important variations among different specific tasks. Furthermore, Jiang and colleagues did not include function words in their core lexicon lists, which have been shown to be relevant in Mandarin PWA production (Wang et al., 2019).

Thus, the present study reports our development of core lexicon lists for each specific task with inclusion of both content and function words with Mandarin PWA data from AphasiaBank. In addition, we hypothesize that core lexicon scores assess linguistic productivity and discourse informativeness at lexical level. We examined these hypotheses by conducting correlation analyses between core lexicon scores, and lexical diversity measures as well as CIUs.

2. Materials and methods

2.1 Data

Data for this study were drawn from the Mandarin section (Deng et al., 2021; Chen et al., 2018) of the AphasiaBank database (MacWhinney & Fromm, 2016; MacWhinney et al., 2011). The original English AphasiaBank protocol contains five sections: (1) Greetings (Initiating the conversation); (2) Picture Descriptions (A. Broken Window, B. Refused Umbrella, C. Cat Rescue, D. Flood); (3) Story Narrative (A. Cinderella); (4) Procedural Discourse (Peanut Butter and Jelly Sandwich); and (5) Free Speech (A. Important Event, B. Stroke Story and C. Stroke Coping). Some changes were made in the Chinese AphasiaBank protocol in which the Cinderella story was replaced by two culturally familiar stories: A. The Tortoise and the Hare and B. Cry Wolf. In addition, the Peanut Butter and Jelly Sandwich procedural discourse task was replaced with description of Egg Ham Fried Rice.

We selected all the seven semi-spontaneous discourse tasks from three major task types, i.e., picture descriptions, story narrative and procedural discourse, for core lexicon list development and for examining task effects on core lexicon analysis in Mandarin PWA. Details regarding administration of these tasks and transcription of discourse samples are included at the AphasiaBank website (https://aphasia.talkbank.org/protocol/languages/Mandarin/).

2.2 Transcripts

Transcripts from 61 neurologically intact participants and 18 PWA were selected from the Mandarin section of the AphasiaBank database (see Table 1). The PWA were diagnosed by WAB-R (Kertesz, 2006) with anomic aphasia (n = 15), Broca's aphasia (n = 1), Wernicke's aphasia (n = 1) and conduction aphasia (n = 1).

The neurologically intact participants' data were further split into two sets with age and years of education matched (see Table 2): One set (n = 43) as a normative sample which was used to generate the core lexicon lists and the other as a control group (n = 18) which served as a baseline for comparison with the PWA. There were also no significant differences between the PWA and the two neurologically intact groups in terms of age and years of education.

Table 1. Demographic information for normative, control and PWA groups.

		Age (yrs)			Education (yrs)			WAB Aphasia Quotient			Time since stroke (months)	
Groups	n	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*
Normative	43	43.1	16.2	22–70	12.6	4.4	6–20	-	-	- -	-	-
Control	18	47.9	12.7	28–67	12.8	3.1	9–20	-	-	- -	-	-
PWA	18	44.6	13.3	23–67	14.2	5.9	5–32	78.4	12	54–93	6.3	5.8

Table 2. Statistical tests of age and years of education among normative, control and PWA groups.

	Groups		*t*	*df*	*p*
Age	Normative	Control	−1.24	40.42	0.66
	Normative	PWA	−0.36	38.63	0.88
	Control	PWA	0.78	33.93	0.88
Years of education	Normative	Control	−0.22	44.35	0.95
	Normative	PWA	−1.02	25.23	0.95
	Control	PWA	−0.86	25.85	0.95

Note: *T*-tests were used to compare age and years of education among the three groups, with Holm adjustments made for the *p*-values.

We also controlled the amount of language production between the normative and control group to be comparable in terms of number of Chinese characters, number of word tokens and types (see Tables 3 and 4).

2.3 Procedures

In the present study, we used several state-of-the-art natural language processing and modelling packages in *R* (R Core Team, 2018) to analyze the data instead of the Computerized Language Analysis (CLAN) program. The *R* packages maximise the efficiency and reproducibility of data analysis with a highly integrated workflow.

First, we extracted different tasks from the transcripts. Then, we used functions in the *stringr* package with regular expressions to filter out annotations such as repetition and pauses. Only transcriptions of clear speech were kept. Then, we tokenised the transcriptions via the *jiebaR* package, which uses dynamic programming to find the most probable combination based on word frequency and has been frequently used in Chinese natural language processing (Chen et al., 2019; Qin & Wu, 2019). We also manually checked the results and modified user-defined dictionary to maximize tokenisation accuracy.

The top 30 words with the widest distribution in the normative samples, i.e., percentage of people in the normative group who used the word, were selected in the present study. This threshold was used because it yields a reasonable size of lexicon, which is manageable in clinical settings and comparable to previous research (n = 25 for only content words, Jiang et al., 2023) and includes a reasonable number of content and function words. We did not use a fixed percentage, say 50% by the normative group as in (Dalton & Richardson, 2015), because that will result in a different number of core lexicon items for each task. This would make it difficult to compare core lexicon scores across different tasks because participants who produce the same number of core words will have different core lexicon scores in different tasks. We did not base our core lexicon lists on word frequencies in the normative samples (c.f., Kim et al., 2019) because frequencies

Table 3. Number of characters, word types, and tokens for each task and each group.

		Characters		Types		Tokens	
Tasks	Groups	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pic-CatRescue	Normative	156.8	95.7	59.2	22.6	115.3	68.1
	Control	148.5	64.2	59.2	18.1	109.6	45.9
	PWA	83.8	54.5	35.1	16.6	66.7	45.1
Pic-Flood	Normative	108	80.8	44.2	22.9	73.5	54.3
	Control	99.8	67.8	41.3	19.0	67.9	44.4
	PWA	57.9	55.0	26.2	19.5	45.0	40.6
Pic-Umbrella	Normative	142.4	84.2	55.9	23.8	104.0	59.6
	Control	131.6	47.3	53.3	14.6	99.0	32.7
	PWA	76.8	41.2	32.3	13.9	59.5	33.5
Pic-Window	Normative	120.4	97.8	47.6	24.1	87.2	69.6
	Control	120.3	64.5	48.3	14.1	87.7	46.4
	PWA	53.0	32.1	24.3	10.5	40.1	24.5
Proc-FriedRice	Normative	127.7	84.8	48.0	22.1	88.6	62.2
	Control	119.7	60.0	47.6	18.0	83.5	41.5
	PWA	52.1	34.4	23.7	14.1	39.2	26.8
S-CryWolf	Normative	247.9	168.4	80.4	37.4	183.3	122.6
	Control	246.8	126.2	78.7	29.7	185.8	95.6
	PWA	100.6	76.8	38.6	24.0	76.4	60.1
S-TortioseHare	Normative	256.4	162.1	86.5	37.3	176.5	112.5
	Control	240.6	113.2	84.7	29.9	166.1	79.0
	PWA	100.3	65.7	35.1	19.1	71.1	45.6

Note: Pic = picture descriptions; Proc = procedural discourse; S = story narrative.

may be biased if a single word was used significantly more than others by a particular participant. In English core lexicon studies, lemmas were used to include different forms of a word. In Chinese, there is no inflection for words. But we grouped synonyms as a core lexicon in generating the core lexicon list. For example, "小孩" (small kid) and "孩子" (kid) were considered the same.

For both PWA and control groups, we calculated agreement percentage as core lexicon scores by dividing the number of matched core lexicon items by the total number of core lexicon items, i.e., 30. Each word was counted only one time in each task regardless of how many times it may have been used.

The core lexicon scores were evaluated for two aspects of validity. For known-groups validity, we hypothesized that PWA would have significantly lower core lexicon scores than the control group. For convergent validity, we conducted correlation analysis of core lexicon scores with aphasia severity scores (AQs of the WAB-R), lexical diversity measures and informativeness measures. We hypothesize that core lexicon scores will have moderate correlations with AQs. We did not anticipate high correlations because AQ scores were also affected by PWA's comprehension competence.

Lexical diversity represents the breadth of vocabulary utilized in a text by a speaker, reflecting their ability to access and retrieve words for constructing discourse. We expect that core lexicon scores will have strong correlations with lexical diversity. Note that there exist many leixcal diversity measures. The most traditional one is TTR, which calculates the ratio of word types to tokens. However, TTR is influenced by sample length. As the sample size grows, the likelihood of introducing new words diminishes, causing the number of word types to grow at a slower rate than the number of tokens, which consistently increases with every additional word. Alternative measures have been explored to mitigate the influence of text length. This includes the application of transformations to TTR,

Table 4. Statistical comparisions of linguistic productivity variables among normative, control and PWA goups.

Measures	Tasks	Groups		*w*	*p*
Characters	Pic-CatRescue	Normative	Control	351	0.569
		Normative	PWA	617	< 0.01*
		Control	PWA	259	< 0.01*
	Pic-Flood	Normative	Control	383	0.95
		Normative	PWA	598	< 0.01*
		Control	PWA	247	0.02*
	Pic-Umbrella	Normative	Control	340	0.462
		Normative	PWA	604	< 0.01*
		Control	PWA	255	< 0.01*
	Pic-Window	Normative	Control	348	0.543
		Normative	PWA	658	< 0.01*
		Control	PWA	282	< 0.01*
	Proc-FriedRice	Normative	Control	395	0.912
		Normative	PWA	666	< 0.01*
		Control	PWA	275	< 0.01*
	S-CryWolf	Normative	Control	351	0.575
		Normative	PWA	634	< 0.01*
		Control	PWA	272	< 0.01*
	S-TortioseHare	Normative	Control	381	0.931
		Normative	PWA	684	< 0.01*
		Control	PWA	274	< 0.01*
Types	Pic-CatRescue	Normative	Control	335	0.411
		Normative	PWA	647	< 0.01*
		Control	PWA	269	< 0.01*
	Pic-Flood	Normative	Control	397	0.881
		Normative	PWA	614	< 0.01*
		Control	PWA	250	0.01*
	Pic-Umbrella	Normative	Control	368	0.770
		Normative	PWA	616	< 0.01*
		Control	PWA	263	< 0.01*
	Pic-Window	Normative	Control	321	0.300
		Normative	PWA	693	< 0.01*
		Control	PWA	303	< 0.01*
	Proc-FriedRice	Normative	Control	382	0.937
		Normative	PWA	651	< 0.01*
		Control	PWA	278	< 0.01*
	S-CryWolf	Normative	Control	365	0.728
		Normative	PWA	649	< 0.01*
		Control	PWA	274	< 0.01*
	S-TortioseHare	Normative	Control	365	0.728
		Normative	PWA	728	< 0.01*
		Control	PWA	294	< 0.01*
Tokens	Pic-CatRescue	Normative	Control	358	0.652
		Normative	PWA	590	< 0.01*
		Control	PWA	250	0.01*
	Pic-Flood	Normative	Control	385	0.975
		Normative	PWA	559	0.02*
		Control	PWA	234	0.05*
	Pic-Umbrella	Normative	Control	336	0.420

(*Continued*)

Table 4. (Continued).

Measures	Tasks	Groups		*w*	*p*
		Normative	PWA	589	< 0.01*
		Control	PWA	257	< 0.01*
	Pic-Window	Normative	Control	346	0.517
		Normative	PWA	654	< 0.01*
		Control	PWA	281	< 0.01*
	Proc-FriedRice	Normative	Control	388	1.000
		Normative	PWA	634	< 0.01*
		Control	PWA	267	< 0.01*
	S-CryWolf	Normative	Control	349	0.553
		Normative	PWA	634	< 0.01*
		Control	PWA	266	< 0.01*
	S-TortioseHare	Normative	Control	379	0.899
		Normative	PWA	672	< 0.01*
		Control	PWA	274	< 0.01*

Wilcoxon tests were used to compare linguistic productivity variables among the three groups, with Holm adjustments made for the p-values.

such as the log transformation known as Herdan's C (Herdan, 1964, also denoted as LogTTR), or the root transformation applied in Guiraud's R (Root TTR).

Additionally, the moving average type–token ratio (MATTR) uses a moving window of a specific size (say 500 words) and provides a more fine-grained analysis of lexical diversity compared to a single TTR value for an entire text. MATTR is thus more suitable for analyzing long texts, such as long articles or novels (Chen & Chang, 2023). A recent measure D, which is reported to be relatively robust to text length variation, combines an algebraic transformation model and curve fitting to estimate lexical diversity (Fergadiotis & Wright, 2011). The curve-fitting procedure for this measure, however, is carried out on a segment of 35–50 tokens (Malvern et al., 2004). Values for D could not be obtained for participants, especially PWA, who produced less than 35 valid words. Here, we calculated TTR, Herdan's C and Guiraud's R and compared their effectiveness in distinguishing between control and PWA groups. These comparisons allow us to select the most suitable lexical diversity measure for our further correlation analyses with core lexicon scores.

For discourse informativeness, we followed the previous study (Nicholas & Brookshire, 1993) to calculate CIUs, which are words that are intelligible in context, and are accurate in relation to the picture(s) or topic, and are informative about their content. Words do not have to be grammatically correct to be counted as CIUs. It is expected that core lexicon scores will have relative strong correlations with CIU scores.

3. Results

3.1 Core lexicon lists

Table 5a and 5b show the top 30 core lexicon items generated from the normative samples in each task. For example, the function word *le* "了" (FC1) ranked very high in all tasks, which was shared by 84%-100% participants in the normative group. We included a very brief table (Table 6) to list the basic functions/meanings of each function

Table 5a. Top 30 core lexicon items generated by the normative sample for the four picture description tasks (Cat Rescue, Flood, Refused Umbrella and Broken Window) with simple English glosses. Function words with multiple functions were marked with FC plus a numeric ID with detailed explanations in Table 6.

	Pic-CatRescue			Pic-Flood			Pic-Umbrella			Pic-Window		
Rank	Chinese	%	English	Chinese	%	English	Chinese	%	English	Chinese	%	English
1	了	100	FC1	了	91	FC1	了	100	FC1	了	100	FC1
2	树	95	tree	的	86	FC2	伞	100	umbrella	踢	100	kick
3	上	93	up	在	84	in	妈妈	95	mum	球	95	ball
4	在	88	in	一个	81	classifier (one)	他	91	he	在	91	in
5	下来	84	come down	这个	77	this	下雨	88	raining	的	86	FC2
6	猫	84	cat	小孩	74	kid	的	86	FC2	到	79	arrive
7	的	84	FC2	是	74	link verb	不	81	negative marker	把	77	FC3
8	狗	81	dog	救	72	save	带	81	carry	窗户	72	window
9	不	79	negative marker	洪水	67	flood	上学	79	go to school	这个	72	this
10	把	77	FC3	有	60	have	着	79	FC5	是	70	link verb
11	小	74	small	然后	58	then	就	74	FC4	然后	70	then
12	来	74	come	这	58	this	小孩	70	kid	一个	65	classifier (one)
13	梯子	74	ladder	她	51	she	雨	70	rain	他	63	he
14	然后	74	then	着	51	FC5	然后	67	then	看	60	look
15	爬	74	climb	把	47	FC3	要	65	want	不	53	negative marker
16	救	72	save	树枝	47	branch	走	65	go	就	53	FC4
17	是	72	link verb	一	44	one	去	60	go	玻璃	53	glass
18	这个	72	this	水	44	water	时候	60	time	一	51	one
19	有	70	have	上	42	up	淋	60	get wet (by rain)	上	51	up
20	到	67	arrive	吧	37	FC7	在	58	in	小孩	51	kid
21	着	67	FC5	就	37	FC4	有	53	have	吧	49	FC7
22	一个	65	classifier (one)	不	35	negative marker	说	53	say	电视	49	TV
23	就	65	FC4	个	35	classifier	这个	53	this	这	49	this
24	下	60	down	去	35	go	上	51	up	有	44	have
25	也	60	too	来	35	come	是	51	link verb	着	44	FC5
26	这	60	this	被	35	FC6	跑	51	run	人	42	man
27	去	58	go	小	33	small	把	49	FC3	个	40	classifier
28	她	58	she	里	33	inside	没	49	none	那个	40	that
29	小孩	58	kid	中	28	in the middle	到	47	arrive	里面	40	inside
30	想	58	think	到	28	arrive	回	47	back, return	家	37	home

word that appeared in Table 5a and 5b with some examples. It should be noted that function words in Chinese are extremely complex in terms of their functions and their interpretations in different structures and it is beyond this paper to elaborate on grammar details of these function words (for information regarding functional categories, see Cui & Sung, 2022; Huang et al., 2014; Li & Thompson, 1989).

Table 5b. Top 30 core lexicon items generated by the normative sample for Procedural Discourse (Fried Rice) and Story Narrative (Cry Wolf and Tortoise and Hare) with simple English glosses. Function words with multiple functions were marked with FC plus a numeric ID with detailed explanations in Table 6.

	Proc-FriedRice			S-CryWolf			S-TortoiseHare		
Rank	Chinese	%	English	Chinese	%	English	Chinese	%	English
1	火腿肠	100	sausage	了	100	FC1	乌龟	100	tortoise
2	炒	98	stir fry	他	100	he	了	100	FC1
3	把	95	FC3	来	98	come	兔子	100	rabbit
4	米饭	95	rice	狼	98	wolf	就	98	FC4
5	放	88	put	的	98	FC2	的	98	FC2
6	了	84	FC1	就	95	FC4	在	95	in
7	油	81	oil	在	93	in	它	93	it
8	然后	81	then	羊	93	sheep	跑	93	run
9	蛋炒饭	81	egg fried rice	没	91	none	快	79	quick
10	先	79	first	放	86	put	赛跑	79	race
11	就	74	FC4	有	86	have	过	79	FC9
12	锅	70	wok	一个	84	classifier (one)	一	77	one
13	的	67	FC2	小孩	79	kid	到	74	arrive
14	再	63	then	人	77	man	得	74	FC8
15	鸡蛋	63	egg	喊	77	shout	我	74	I
16	切	60	cut	是	77	link verb	时候	74	time
17	可以	56	can	都	77	all	睡	74	sleep
18	好	53	good	真的	72	real	不	72	negative marker
19	是	53	link verb	不	70	negative marker	有	72	have
20	做	51	do	时候	70	time	是	67	link verb
21	在	51	in	说	70	say	然后	67	then
22	倒	49	pour	又	65	again	动物	65	animal
23	盐	44	salt	然后	65	then	听	65	hear
24	里面	44	inside	吃	63	eat	和	65	and
25	我	42	I	这个	60	this	很	65	very
26	进去	42	into	听	58	hear	比赛	63	games
27	一下	40	one time	呢	58	FC10	爬	63	climb
28	之后	37	then	把	58	FC3	终点	63	end
29	以后	37	after this	次	58	(this) time	开始	60	begin
30	就是	37	be exactly	上	56	up	没	60	none

3.2 Comparisions of core lexicon scores between PWA and healthy control groups

Next, we calculated core lexicon scores for both PWA and healthy control groups (see Figure 1).

We compared the scores of PWA with those of controls for each task. We used *Wilcoxon* tests and reported here both *p* values and effect sizes. The results indicate that core lexicon scores of PWA and those of controls were significantly different for all the tasks and the effect sizes were all strong (See Table 7).

3.3 Correlations of core lexicon scores with aphasia severity, lexical diversity and discourse informativeness

To investigate the relationship between core lexicon scores and aphasia severity, Spearman's correlations, a nonparametric test corresponding to Pearson correlations, were performed

Table 6. A summary of Mandarin function words in the core lexicon lists.

ID	Character(s)	Pinyin	Function (s)	Examples
FC1	了	*le*	le1: a grammatical marker indicating an event is bounded temporally (located in the past) le2: a sentence-final particle expressing a currently relevant state	看了电影(watched a movie) 刮风了。(The wind is blowing.)
FC2	的	*de*	a structural particle used after an attribute	好的 (good)
FC3	把	*ba*	an object marker, often used in the well-known *ba*-structure	把门打开(open the door)
FC4	就	*jiu*	an adverb, indicating imminent future or proximity of location	马上就来(about to come); 就在这里(right here)
FC5	着	*zhe*	an aspect marker indicating a continuous state	打着雨伞(using an umbrella).
FC6	被	*bei*	a passive marker	被打 (be beaten)
FC7	吧	*ba*	a sentence-final particle	我们走吧! (Let's go!)
FC8	得	*de*	a particle used after a verb showing effect	跑得快 (run fast)
FC9	过	*guo*	a grammatical marker showing something has happened at least once in the past	吃过午饭 (have had lunch)
FC10	呢	*ne*	a sentence-final particle	你呢? (And you?)

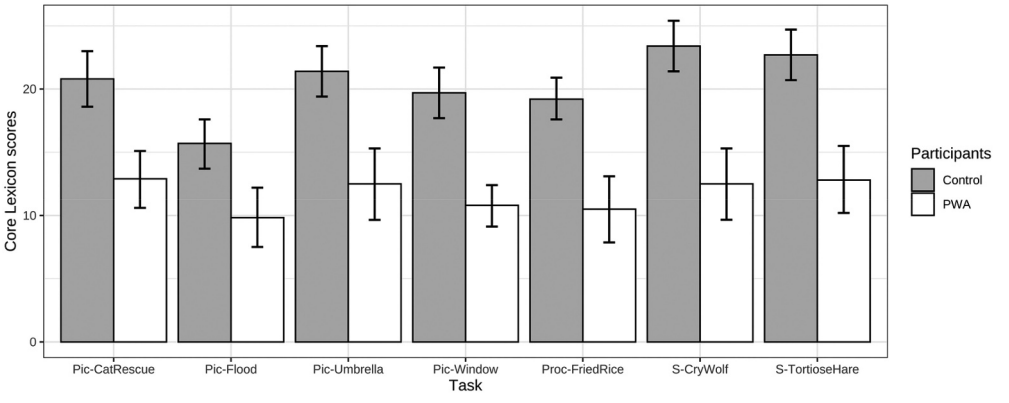

Figure 1. Core lexicon scores of different tasks by Mandarin healthy controls and PWA. Note: Pic = picture descriptions; Proc = procedural discourse; S = story narrative.

Table 7. Statistical comparisons of core lexicon scores between healthy controls and PWA in each discourse production task.

Tasks	*w*	*p*	Effect sizes	
Pic-CatRescue	291	< 0.001	0.68	large
Pic-Flood	268	< 0.001	0.56	large
Pic-Umbrella	290	< 0.001	0.68	large
Pic-Window	311	< 0.001	0.79	large
Proc-FriedRice	291	< 0.001	0.68	large
S-CryWolf	308	< 0.001	0.77	large
S-TortioseHare	300	< 0.001	0.73	large

Note: Pic = picture descriptions; Proc = procedural discourse; S = story narrative. Effect sizes: < 0.3 (small), 0.3 ~ 0.5 (moderate), > 0.5 (large).

Table 8. Correlations of core lexicon scores with aphasia severity.

Tasks	$r_{spearman}$	*p*
Pic-CatRescue	0.46	0.06
Pic-Flood	0.39	0.11
Pic-Umbrella	0.23	0.36
Pic-Window	0.07	0.78
Proc-FriedRice	0.41	0.09
S-CryWolf	0.47	0.05*
S-TortioseHare	0.51	0.03*

Note: Pic = picture descriptions; Proc = procedural discourse; S = story narrative.

between core lexicon scores and WAB-based Aphasia Quotients (AQs) for each task as in a previous study (Kim et al., 2019). Significant correlations between core lexicon scores and WAB AQs were found only for the two story narrative tasks (see Table 8).

As mentioned in section 2.3, we calculated the type-token ratios (TTR), log transformed TTR (i.e., C) and root transformed TTR (i.e., R) for each group and task (see Table 9) to quantify lexical diversity. PWA showed significantly and unexpectedly higher TTR than both normative and control group for the Cry Wolf task and there were no signiticant differences among the three groups for all the others tasks (see Table 10). This suggests that TTR as affected by text sample size cannot accurately quantify lexical diversity in our dataset. On the other hand, PWA

Table 9. Lexical diversity measures of each group per task.

		TTR		C		R	
Tasks	Groups	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pic-CatRescue	Normative	0.6	0.1	0.9	0	5.6	0.7
	Control	0.6	0.1	0.9	0	5.7	0.7
	PWA	0.6	0.1	0.9	0	4.3	0.8
Pic-Flood	Normative	0.7	0.1	0.9	0	5.2	0.9
	Control	0.6	0.1	0.9	0	5.0	1.0
	PWA	0.7	0.2	0.9	0.1	3.9	1.1
Pic-Umbrella	Normative	0.6	0.1	0.9	0	5.5	0.9
	Control	0.6	0.1	0.9	0	5.3	0.7
	PWA	0.6	0.1	0.9	0	4.2	0.8
Pic-Window	Normative	0.6	0.1	0.9	0	5.2	0.8
	Control	0.6	0.1	0.9	0	5.3	0.4
	PWA	0.7	0.2	0.9	0.1	3.9	0.8
Proc-FriedRice	Normative	0.6	0.1	0.9	0	5.1	0.8
	Control	0.6	0.1	0.9	0	5.2	0.7
	PWA	0.7	0.2	0.9	0.1	3.7	1.0
S-CryWolf	Normative	0.5	0.1	0.9	0	6	0.9
	Control	0.5	0.1	0.8	0	5.8	0.8
	PWA	0.6	0.1	0.9	0	4.4	1.2
S-TortioseHare	Normative	0.5	0.1	0.9	0	6.6	0.9
	Control	0.5	0.1	0.9	0	6.6	0.9
	PWA	0.5	0.1	0.8	0	4.1	1.1

Note: TTR = type-token ratios; Pic = picture descriptions; Proc = procedural discourse; S = story narrative.

Table 10. Statistical comparisions of lexical diversity measures among normative, control and PWA goups per task.

Measures	Tasks	Groups		*w*	*p*
TTR	Pic-CatRescue	Normative	Control	375	1.000
		Normative	PWA	344	1.000
		Control	PWA	151	1.000
	Pic-Flood	Normative	Control	442	1.000
		Normative	PWA	358	1.000
		Control	PWA	132	1.000
	Pic-Umbrella	Normative	Control	440	1.000
		Normative	PWA	379	1.000
		Control	PWA	129	1.000
	Pic-Window	Normative	Control	405	0.867
		Normative	PWA	320	0.867
		Control	PWA	131	0.867
	Proc-FriedRice	Normative	Control	333	0.470
		Normative	PWA	269	0.186
		Control	PWA	124	0.470
	S-CryWolf	Normative	Control	445	0.363
		Normative	PWA	230	0.027*
		Control	PWA	79	0.027*
	S-TortioseHare	Normative	Control	364	1.000
		Normative	PWA	367	1.000
		Control	PWA	166	1.000
C	Pic-CatRescue	Normative	Control	352	0.768
		Normative	PWA	443	0.768
		Control	PWA	199	0.756
	Pic-Flood	Normative	Control	455	0.867
		Normative	PWA	415	1.000
		Control	PWA	159	1.000
	Pic-Umbrella	Normative	Control	464	0.458
		Normative	PWA	465	0.315
		Control	PWA	190	0.458
	Pic-Window	Normative	Control	387	1.000
		Normative	PWA	408	1.000
		Control	PWA	167	1.000
	Proc-FriedRice	Normative	Control	308	0.651
		Normative	PWA	370	1.000
		Control	PWA	168	1.000
	S-CryWolf	Normative	Control	471	0.561
		Normative	PWA	338	0.561
		Control	PWA	126	0.561
	S-TortioseHare	Normative	Control	357	0.644
		Normative	PWA	543	0.039*
		Control	PWA	234	0.045*
R	Pic-CatRescue	Normative	Control	336	0.428
		Normative	PWA	685	< 0.01*
		Control	PWA	294	< 0.01*
	Pic-Flood	Normative	Control	398	0.869
		Normative	PWA	666	< 0.01*
		Control	PWA	259	< 0.01*

(*Continued*)

Table 10. (Continued).

Measures	Tasks	Groups		*w*	*p*
	Pic-Umbrella	Normative	Control	432	0.487
		Normative	PWA	635	< 0.01*
		Control	PWA	261	< 0.01*
	Pic-Window	Normative	Control	332	0.389
		Normative	PWA	689	< 0.01*
		Control	PWA	315	< 0.01*
	Proc-FriedRice	Normative	Control	363	0.713
		Normative	PWA	657	< 0.01*
		Control	PWA	281	< 0.01*
	S-CryWolf	Normative	Control	432	0.482
		Normative	PWA	657	< 0.01*
		Control	PWA	270	< 0.01*
	S-TortioseHare	Normative	Control	351	0.578
		Normative	PWA	742	< 0.01*
		Control	PWA	308	< 0.01*

Note: Pic = picture descriptions; Proc = procedural discourse; S = story narrative.

showed significantly lower C scores than both normative and control groups but only for the Tortiose and Hare task. This suggests that although C scores can estimate lexical diversity for that task, it is not sensitive enough for other tasks.

R scores outperformed both TTR and C in that for all tasks both normative and control groups were higher than PWA, suggesting higher lexical diversity as we expected. Additionally, there was no significant differences between normative and control groups (see Table 10). Thus, R scores was the best indicator for the present study and will be used as the lexical diversity measure for the following correlation analyses with core lexicon scores.

To investigate the relationship between core lexicon and lexical diversity, Spearman's correlations were performed between core lexicon scores and R scores for each task. Significant correlations between the two scores were found for all the tasks (See Table 11).

We also did Spearman's correlation analyses between core lexicon scores and CIUs with PWA production data for each task to investigate the relationship between core lexicon and discourse informativeness. Significant correlations were found for all the discourse tasks, suggesting that core lexicon scores can be effective estimates of informativeness of PWA discourse production (see Table 12).

Table 11. Correlations of core lexicon scores with lexical diversity R.

Tasks	$r_{spearman}$	*p*
Pic-CatRescue	0.80	< 0.01*
Pic-Flood	0.83	< 0.01*
Pic-Umbrella	0.81	< 0.01*
Pic-Window	0.61	< 0.01*
Proc-FriedRice	0.87	< 0.01*
S-CryWolf	0.87	< 0.01*
S-TortioseHare	0.88	< 0.01*

Table 12. Correlations of core lexicon scores with Correct Information Units.

Tasks	$r_{spearman}$	*p*
Pic-CatRescue	0.86	< 0.01*
Pic-Flood	0.82	< 0.01*
Pic-Umbrella	0.71	< 0.01*
Pic-Window	0.56	0.02*
Proc-FriedRice	0.86	< 0.01*
S-CryWolf	0.90	< 0.01*
S-TortioseHare	0.94	< 0.01*

4. Discussion

In the present study, we have developed core lexicon lists for Mandarin PWA data for seven discourse production tasks. Core lexicon scores based on these lists distinguished discourse produced by PWA and controls in all seven tasks and significantly correlated with aphasia severity (AQs) for two story narrative tasks and with lexical diversity (R scores) and discourse informativeness (CIUs) for all seven tasks.

First, core lexicon scores of Mandarin PWA were significantly lower than those of controls across all the tasks as we expected, in line with previous studies on English PWA (Dalton & Richardson, 2015; Kim et al., 2021, 2019) and Mandarin PWA (Jiang et al., 2023). This suggests that lexical retrieval difficulties in discourse production by Mandarin PWA can be assessed with task-specific core lexicon lists based on data produced by an age- and education-matched normative sample group. Different from the previous research on Mandarin PWA (Jiang et al., 2023) which aggregated different specific tasks within a task type, our core lexicon lists for each task can be used directly by clinicians to assess PWA discourse production.

Second, core lexicon scores positively correlated with AQs in the two story narrative tasks. This is consistent with two previous studies (Kim et al., 2021, 2019) on English PWA in which significant correlations were found between AQs and core verbs (Kim et al., 2019) and function words (Kim et al., 2021) in two story narrative tasks. However,the previous study on Mandarin PWA (Jiang et al., 2023) did not find significant correlations between AQs and core nouns or verbs. This difference could be accounted for by the fact that participants in our study included more subtypes than just anomia, and thus their AQ variations were larger than those of Jiang et al. (2023).

Moreover, we have included both content and function words whereas only content words (i.e., core nouns and verbs) were considered in Jiang et al. (2023), suggesting that core function words have an important role to play in assessing PWA discourse production. The core function words in our core lexicon lists revealed distinctive linguistic characteristics of Chinese. For example, *le* "了" ranked first in several lists, suggesting the importance of this function category. The fact that *le* can serve as either a sentence-final particle (sentential-le) or a particle that appears immediately after the verb (verbal-le) may explain its wide distribution across different speakers. Other aspect makers such as "着" *zhe* (imperfective aspect markers) and "过" *guo* (perfective aspect markers) were also identified. Another important function category is numeral classifiers, such as *ge* "个", which frequently appeared in several lists. In Chinese, classifiers are obligatory in cases

where a noun is preceded by a number (one, *yi ge* "一个"), a demonstrative (this/that, *zhe ge/na ge* "这个/那个"), and certain quantifiers (some, *ji ge* "几个"; Huang et al., 2014). Thus, our study emphasised the importance of including function words in assessing Mandarin PWA discourse production.

In addition, our correlation analysis of core lexicon scores and AQs suggested that the story narrative task is the most revealing task type in assessing discourse production performance. In the picture description tasks, especially for single picture descriptions, PWA tended to list all objects in the pictures with very simple words and sentences. Thus, the variations in total lexical items and core lexicon scores can be very small in general among PWA with differing AQs. In the procedure task, PWA were constrained by the experience they had with the procedure, i.e., making fried rice, which rendered the correlation between core lexicon scores and AQs to be weak. In the story narrative tasks, however, participants had to describe temporal, spatial and/or causal information to form storylines. They generally produced more words with greater variations among PWA with differing AQs (see Table 3). In fact, in the two English PWA studies (Kim et al., 2021, 2019) that found correlations between core lexicon scores and AQs, their data were elicited with two 30-page wordless picture books with storylines, similar to the story narrative tasks in our study.

Third, core lexicon scores of Mandarin PWA discourse production also significantly correlated with lexical diversity R scores for all tasks, indicating the underlying common discourse features, such as lexical semantics, shared by both measures. The advantage of the core lexicon score is that it can be obtained in real time without transcriptions relative to other lexical diversity measures. The seven core lexicon lists in our study, covering the commonly used task types in aphasia research, serve as a response to the need of a transcription-less approach in analysing discourse production in clinical settings (Kintz & Wright, 2018; De Riesthal & Diehl, 2018). These strong correlations across all tasks are different from a previous English study (Kim & Wright, 2020), in which separate lists were generated for core nouns, verbs, adjectives, adverbs and function words for two story narrative tasks, *Good Dog, Carl* and *Picnic*. In that study, core nouns, verbs, adjectives and function words generated with the *Picnic* task correlated positively with lexical diversity measures, whereas for the *Good Dog, Carl* task, there is no positive correlation between any core lexicon lists and lexical diversity, suggesting task-induced variations in core lexicon measures and/or lexical diversity. Our core lexicon lists did not separate word classes but incorporated both content and function words and showed more stable and consistent correlations with lexical diversity across task types. Another cause for the different results could be the different ways of calculating lexical diversity (MATTR as in Kim & Wright, 2020). As stated in the methodology section, no single best measure of lexical diversity exists for all tasks and all languages. In our case, the measure we chose is valid in the context of our study as it effectively distinguished lexical differences between healthy and PWA participants and at the same time remained stable across the normative and control groups.

Fourth, core lexicon scores also significantly correlated with the discourse informativeness measure, CIU scores, for all tasks, suggesting that core lexicon scores can

roughly assess the informativeness of discourse production. We attribute this to the fact that our distribution-based core lexicon lists captured content words that are closely related to the content of each task. For example, in the Cat Rescue picture description task, core nouns are "小孩" (kid), "猫" (cat), "狗"(dog), "树"(tree), "梯子"(ladder) and core verbs are "爬" (climb), "下" (go down), "救" (save). These words can, therefore, serve as a baseline for quick real-time assessment of informativeness in each task. Our findings diverge from the previous English study (Kim & Wright, 2020), in which only the core adverb list in the *Good Dog, Carl* task and the core adjective list in the *Picnic* task exhibited correlations with CIUs. It should be noted that the participant cohort in (Kim & Wright, 2020) was smaller (N = 11) with a wider range of aphasia types (3 anomia, 2 conduction, 5 Broca and 1 transcortical motor), thus leading to potentially greater variations and more intricate dynamics between core lexicon and discourse informativeness compared to our study. Additionally, the inclusive incorporation of all word classes in our core lexicon lists appears to enhance their sensitivity in capturing the content of discourse and diminish susceptibility to task variations.

Limitations and future directions

There are several limitations to our study that need to be considered in future investigations. First, although core lexicon scores are informative in assessing aphasia discourse production, the best approach to construct a core lexicon list is still inconclusive. There are two traditional approaches in constructing core lexicon lists: The frequency-based approach and the percentage (or distribution)-based approach. Kim, Schoemann et al. (2022) investigated the differences between the two approaches with English aphasia data via structural equation modelling and found that these two different approaches measured the same construct and were highly correlated with each other, suggesting that core lexicon lists developed using these two approaches can be functionally interchangeable. However, there were variations in measurement errors between the two approaches across word classes. Lower measurement errors were found in core nouns, verbs, adjectives, and adverbs developed via the frequency-based approach whereas the percentage-based approach was more suitable for function words. Additionally, lexical retrieval is influenced by discourse task types, consequently impacting the optimal approach for generating core lexicon lists (Kim, Schoemann et al., 2022). Our present study employed a consistent approach across picture descriptions, procedural discourse and story narratives without accounting for the distinctive characteristics inherent in each task type. Future research may explore the most effective strategies for developing core lexicon checklists for these task types.

A second pressing methodological issue is the lack of uniformity in determining the appropriate number of lexical items for inclusion in checklists. The previous frequency-based Mandarin core lexicon lists (Jiang et al., 2023) employed a criterion of the top 25 content words. As we intended to include both content and function words, we expanded our percentage-based list to include the top 30 words. In both cases the choices were arbitrary. Subsequent research is thus warranted to systematically investigate and ascertain the optimal threshold for item selection across different approaches and tasks.

Although we have explored the correlation between core lexicon scores and linguistic productivity and discourse informativeness, more systematic investigation of psychometric properties of Mandarin core lexicon lists with consideration of both microlignuistic and macrolinguistic measures (as demonstrated by Kim & Wright, 2020) is needed. It will not only augment the methodological rigor in the generation of core lexicon lists but also contribute to the attainment of more accurate core lexicon lists, thereby assisting clinicians in selecting the most appropriate list for their assessments.

Most PWA in the present study were identified as anomic aphasia. Future studies could include a sufficient number of participants with different aphasia subtypes to test if core lexicon scores can distinguish different subtypes of aphasia in Mandarin. Our analysis was based on Mandarin production. Future research can develop local dialectal version of core lexicon lists at least for major dialects in China, which would require collaboration between linguistics and local speech/language pathologists.

Conclusions

Seven Mandarin Chinese core lexicon lists were developed for three major discourse elicitation task types and core lexicon scores generated with reference to these lists can distinguish discourse production between healthy controls and PWA. Story narrative tasks, compared with picture and procedural description tasks, were particularly revealing in assessing discourse production within PWA as core lexicon scores for these tasks strongly correlated with AQs. Significant correlations between core lexicon scores and R scores and CIU scores for all seven tasks indicate that core lexicon scores can be used to estimate lexical diversity and discourse informativeness. These lists, in general, serve as an alternative transcription-less approach in assessing Mandarin PWA discourse production.

Disclosure statement

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

Funding

This work was supported by the Project of Humanities and Social Sciences Fundation of the Ministry of Education of PRC [23YJC740002].

ORCID

Juqiang Chen http://orcid.org/0000-0001-8151-3964
Hui Chang http://orcid.org/0009-0003-1497-5853

References

Capilouto, G., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*, *38*(6), 431–444. https://doi.org/10.1016/j.jcomdis.2005.03.005

Chen, J., & Chang, H. (2023). Testing simplification in translated and creative writing texts: The case of Robert van Gulik's Judge Dee detective stories. *Across Languages and Cultures*, *24*(2), 268–282. https://doi.org/10.1556/084.2023.00375

Chen, C.-M., Chen, Y.-T., & Liu, C.-Y. (2019). Development and evaluation of an automatic text annotation system for supporting digital humanities research. *Library Hi Tech*, *37*(3), 436–455. https://doi.org/10.1108/LHT-10-2017-0219

Chen, Z. Z., Lin, F., & Deng, B. (2018). Construction of Chinese Mandarin corpus with anomic aphasia. *Chinese Journal of Rehabilitation Medicine*, *33*(6), 669–674. https://doi.org/10.3969/j.issn.1001-1242.2018.06.010

Cui, S., & Sung, K. (2022). *A reference grammar for teaching Chinese: syntax and discourse*. Springer Singapore. https://doi.org/10.1007/978-981-33-4207-1

Dalton, S. G. H., & Richardson, J. D. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology*, *28*(1S), 293–320. https://doi.org/10.1044/2018_AJSLP-17-0166

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology*, *24*(4), 923–938. https://doi.org/10.1044/2015_AJSLP-14-0161

Day, A. (1985). *Good Dog Carl*. Simon & Schuster.

Deng, B., Lin, F., Lai, Q., Chen, Z., Zhou, L., & Jiang, Z. (2021). Applying main concept analysis to discourse assessment in aphasia: Research on anomic aphasia based on normal controls. *Chinese Journal of Rehabilitation Medicine*, *36*(4), 418–425. https://doi.org/10.3969/j.issn.1001-1242.2021.04.006

de Riesthal, M., & Diehl, S. K. (2018). Conceptual, methodological, and clinical considerations for a core outcome set for discourse. *Aphasiology*, *32*(4), 469–471. https://doi.org/10.1080/02687038.2017.1398805

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

Huang, C.-T. J., Li, Y. H. A., Simpson, A., Huang, C.-T. J., Li, Y. H. A., & Simpson, A. (Eds.). (2014). *The handbook of Chinese linguistics*. John Wiley & Sons. http://ebookcentral.proquest.com/lib/sjtu/detail.action?docID=1634265

Jiang, Y., Liao, X., & Liu, N. (2023). Applying core lexicon analysis in patients with anomic aphasia: Based on Mandarin AphasiaBank. *International Journal of Language and Communication Disorders*, *58*(6), 1875–1886. *n/a*(n/a. https://doi.org/10.1111/1460-6984.12864

Kertesz, A. (2006). *Western Aphasia Battery-Revised*. Pearson.

Kim, H., Berube, S., Hillis, A. E., & Castro, N. (2022). Core lexicon in aphasia: A longitudinal study. *Aphasiology*, *38*(1), 1–13. https://doi.org/10.1080/02687038.2022.2121598

Kim, H., Kintz, S., & Wright, H. H. (2021). Development of a measure of function word use in narrative discourse: Core lexicon analysis in aphasia. *International Journal of Language and Communication Disorders*, *56*(1), 6–19. https://doi.org/10.1111/1460-6984.12567

Kim, H., Kintz, S., Zelnosky, K., & Wright, H. H. (2019). Measuring word retrieval in narrative discourse: Core lexicon in aphasia. *International Journal of Language and Communication Disorders*, *54*(1), 62–78. https://doi.org/10.1111/1460-6984.12432

Kim, H., Schoemann, A. M., & Wright, H. H. (2022). Quality of measurement in core lexicon ceasures. *Journal of Speech, Language, and Hearing Research*, *65*(8), 2919–2930. https://doi.org/10.1044/2022_JSLHR-20-00722

Kim, H., & Wright, H. H. (2020). Concurrent validity and reliability of the core Lexicon measure as a measure of word retrieval ability in aphasia narratives. *American Journal of Speech-Language Pathology*, *29*(1), 101–110. https://doi.org/10.1044/2019_AJSLP-19-0063

Kintz, S., & Wright, H. H. (2018). Discourse measurement in aphasia research. *Aphasiology*, *32*(4), 472–474. https://doi.org/10.1080/02687038.2017.1398807

Kong, A. P.-H. (2009). The use of main concept analysis to measure discourse production in cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders*, *42*(6), 442–464. https://doi.org/10.1016/j.jcomdis.2009.06.002

Lai, Q., Jiang, Z., Chen, Z., Deng, B., Zhou, L., & Lin, F. (2019). Exploratory analysis of discourses in subjects with anomic aphasia and health controls based on Mandarin AphasiaBank. *Chinese Journal of Rehabilitation*, *34*(10), 507–512.

Law, S.-P., Kong, A. P.-H., & Lai, C. (2018). An analysis of topics and vocabulary in Chinese oral narratives by normal speakers and speakers with fluent aphasia. *Clinical Linguistics & Phonetics*, *32* (1), 88–99. https://doi.org/10.1080/02699206.2017.1334092

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press.

MacWhinney, B., & Fromm, D. (2016). AphasiaBank as bigData. *Seminars in Speech and Language*, *37* (1), 010–022. 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

Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan.

McCully, E. (1984). *Picnic*. Harper Collins.

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, Language, 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, Language, and Hearing Research*, *38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Qin, W., & Wu, Y. (2019). *jiebaR: Chinese Text Segmentation* (0.11) [Computer software]. https://cran.r-project.org/web/packages/jiebaR/index.html

R Core Team. (2018). *R: A language and environment for statistical computing* [Manual]. https://www.R-project.org/

Richardson, J. D., & Dalton, S. G. (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., Greenslade, K. J., Jacks, A., Haley, K. L., & Adams, J. (2021). Main concept, sequencing, and story grammar analyses of cinderella narratives in a large sample of persons with aphasia. *Brain Sciences*, *11*(1), 110. https://doi.org/10.3390/brainsci11010110

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*, *49*(4), 1503–1519. https://doi.org/10.3758/s13428-016-0810-5

Wang, H., Liu, X., Wang, X., Cui, G., Xie, X., Zhong, X., Zhai, S., Dong, L., Sui, S., Xin, H., Ge, L., & Yu, C. (2019). Functional category production and degrees of severity: Findings from Chinese agrammatism. *Aphasiology*, *33*(10), 1227–1247. https://doi.org/10.1080/02687038.2019.1626081

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, *45*(1), 27–36. https://doi.org/10.1044/jshd.4501.27