An Analysis of Topics and Vocabulary in Chinese Oral Narratives by Normal Speakers and Speakers with Fluent Aphasia

Sam-Po Law[1], **Anthony Pak-Hin Kong**[2], and **Christy Lai**[1]

[1]Division of Speech and Hearing Sciences, University of Hong Kong, Pokfulam Road, Hong Kong SAR

[2]Department of Communication Sciences and Disorders, University of Central Florida, Orlando, FL, USA

Abstract

This study analyzed the topic and vocabulary of Chinese speakers based on language samples of personal recounts in a large spoken Chinese database recently made available in the public domain, i.e. Cantonese AphasiaBank (http://www.speech.hku.hk/caphbank/search/). The goal of the analysis is to offer clinicians a rich source for selecting ecologically valid training materials for rehabilitating Chinese-speaking people with aphasia (PWA) in the design and planning of culturally and linguistically appropriate treatments. Discourse production of 65 Chinese-speaking PWA of fluent types (henceforth, PWFA) and their non-aphasic controls narrating an important event in their life were extracted from Cantonese AphasiaBank. Analyses of topics and vocabularies in terms of part-of-speech, word frequency, lexical semantics and diversity were conducted. There was significant overlap in topics between the two groups. While the vocabulary was larger for controls than that of PWFA as expected, they were similar in distribution across parts-of-speech, frequency of occurrence, and ratio of concrete to abstract items in major open word classes. Moreover, proportionately more different verbs than nouns were employed at the individual level for both speaker groups. The findings provide important implications for guiding directions of aphasia rehabilitation not only of fluent but also non-fluent Chinese aphasic speakers.

Keywords

Chinese fluent aphasia; connected speech; vocabulary; topics; language rehabilitation

Introduction

Treatments of word retrieval difficulties as a result of brain injuries have a strong tendency to target nouns and verbs that are concrete or highly imageable for various reasons (Renvall, Nickels, & Davidson, 2013a; b). This bias may be understandable as nouns and verbs are major word classes and they arguably carry the core information in any verbal communication. Nonetheless, it is questionable whether such approaches are most effective

in enhancing the communicative abilities of individuals with acquired language deficits and whether the trained items are ecologically valid. As such, many studies of word retrieval treatments would evaluate generalization of treatment effects to discourse, most commonly in performance in conversations, in addition to story-telling and picture description (see Carragher, Convoy, Sage, & Wilkinson, 2012; Webster & Whitworth, 2012; Webster, Whitworth, & Morris, 2015, for a review). Renvall et al. (2013a; b) considered in detail the criteria for functionally relevant lexical items and methods of selecting such stimuli for language rehabilitation. Functionally relevant items could be those personally chosen by people with aphasia (PWA) or those with generally high frequency of occurrence. Based on two previous studies of topics of conversation (i.e. Stuart, Vanderhoof-Bilyeu, & Beukelman, 1994; Davidson, Worrall, & Hickson, 2003) and two large spoken corpora (i.e. SUBTLEX, Brysbaert & New, 2009; CELEX, Baayen, Piepenbrock, & Gulikers, 1995), Renvall et al. (2013a; b) compared common conversation topics among unimpaired elderly speakers with those of PWA, and compiled lists of most frequent vocabulary items. It was found that frequently used lexical items by unimpaired speakers involve word classes besides nouns and verbs, such as adverbs, adjectives and pronouns, and they are less concrete and of lower imageability than typically trained items in language therapy. Conversations of PWA tend to focus on the here and now and be restricted to topics concerning individuals socially close to them (Renvall et al., 2013a; b). The authors suggested that these observations were related in the sense that PWA were often unable to produce abstract or emotion words; consequently, they were limited in participating in a variety of topics of everyday conversation.

Drawing on results of Davidson et al. (2003) and Stuart (2000) that elderly speakers often use storytelling narratives to recount their life history, express their values and opinions on life issues, and to link the past and the present, Renvall et al. (2013b) have made a number of recommendations for future research. More specifically, similar studies should be conducted based on language samples from speakers of a wider age range and different cultural backgrounds. In rehabilitating language production of older speakers, it is preferable to also target function words and abstract content words, as well as different genres of conversations. They also called for a greater effort to include conversations besides monologues in large spoken databases such as the AphasiaBank consortium (MacWhinney, Fromm, Forbes, & Holland, 2011, http://talkbank.org/AphasiaBank/) to serve as potential source of data for selecting topic and vocabulary as trained items in language therapies.

With regard to awareness and sensitivity to linguistic and cultural diversity, the AphasiaBank consortium is pioneering in aphasia research because of its representation of languages other than English, such as Spanish, French, and more recently Cantonese (Kong, Law, & Lee, 2010; Cantonese AphasiaBank accessible at http://www.speech.hku.hk/caphbank/search/). Cantonese is the second most widely spoken dialect in the Chinese language family with over 52 million speakers distributed over southern China and overseas Chinese communities. The Chinese language has a number of linguistic characteristics which differ drastically from most European languages. Chinse has little inflectional morphology. It lacks person or tense marking on the verb, and has no number, gender, or case agreements between a noun and its modifier. There are, however, aspect markers that appear after a verb. It is important to note that the use of aspect marker is optional, particularly when a time adverbial (e.g.

yesterday, just now, already) is present to indicate the state of the action or event denoted by the verb. In short, the morphological and phonological forms of Chinese words remain unchanged during sentence construction. With respect to the inventory of word classes, two do not exist in European languages - utterance-final particles (UFPs) and classifiers. The former appear at the end of utterances, and mainly serve various pragmatic functions, such as indicating speech act types or the source of knowledge of the speaker (evidentiality), or expressing the attitude of the speaker towards what s/he is saying. Classifiers are grammatical morphemes associated with nouns. Every noun is assigned a specific classifier; its occurrence is obligatory when the noun phrase contains a numeral, a demonstrative, or a quantifier, e.g. *loeng-zi-bat* (two-classifier-pen 'two pens'). It is somewhat equivalent to *piece* as in 'two pieces of paper' in English.

Over the past decades, a small number of spoken Cantonese databases have been constructed, but many of them are based on scripted materials (Lee 2011; Lee, Lo, Ching, & Meng, 2002; Xu & Lee 1998). Speech of a more spontaneous nature can be found in Zhou, Li, Yin, and Zong (2010) from telephone conversations and Leung and Law (2001) from radio programmes; however, they lack demographic information on the speakers. In contrast, Cantonese AphasiaBank (Kong et al., 2010) contains speech samples elicited from various narrative tasks of native speakers of Cantonese including 150 neurologically unimpaired male and female speakers of varied ages and education levels, and 105 speakers of different aphasia types who suffered a single stroke as verified through neuroimaging or a clear medical diagnosis.

Although Cantonese AphasiaBank (Kong et al., 2010) does not contain conversations, we argue that it still contains valuable data that can inform clinicians the range of topics that are of interest to elderly Cantonese speakers, healthy or aphasic, and the associated vocabulary. Such information may be treated as a source for selecting culturally, linguistically and age appropriate materials in language rehabilitation of Chinese PWA. More specifically, in one of the narrative tasks, i.e. Important Event, the participants were asked to narrate an important event in their life. It was the only narrative task that did not have targeted themes or content. The requirement of the task may elicit narratives that are not restricted to the present and carry emotional expressions, differing from the tendency observed of conversations in previous studies. It is worth noting that despite the differences between personal recounts and conversations, asking participants to tell personal stories has been less frequently employed compared with other discourse tasks. Previous studies examining the linguistic properties of personal recounts were often based on language samples of healthy speakers (Armstrong, 2002; Sherratt, 2007; Whitworth, Claessen, Leitão, & Webster, 2015). A few studies in which PWA recounted personal experience of strong emotional value have shown that aphasic speakers performed better on recounting emotive than emotionally-neutral topics in terms of relevancy and quantity (Borod et al., 2000), and they utilized similar devices as healthy speakers, including direct speech, repetition for emphasis, and production of emotive words and metaphors (Armstrong & Ulatowska, 2007). The nature of the Important Event task also echoes the emphasis on storytelling narrative as a genre in language therapy of older PWA by Renvall et al. (2013b).

In this study, language samples of narratives about important events by 65 Chinese speakers with fluent aphasia (henceforth, PWFA) and their age- and education-matched controls are extracted from Cantonese AphasiaBank (Kong et al. 2010). We focused on the connected speech of speakers with fluent aphasia because their narratives tended to be longer and contained more complete sentences; consequently, it was usually easier to determine the topics or themes of the recounts, compared with those of non-fluent speakers. Analyses of topics and vocabularies in terms of parts-of-speech and word frequency were conducted. The production of nouns and verbs of the two speaker groups was further compared with respect to the concreteness of meaning denoted by the words and lexical diversity.

Method

Ethics statement

Informed written consent was obtained from all participants before the study began. The tasks were performed in accordance with the Declaration of Helsinki with approval of the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster.

Participants

Inclusion criteria for individuals with aphasia were (i) a single stroke resulting in left cortical or subcortical lesion(s), (ii) post-onset for at least six months, (iii) premorbidly right-handed, and (iv) Cantonese as the mother tongue or their primary language of daily communication without an accent. Exclusion criteria included (i) multiple strokes, (ii) severe problems with articulation/oro-motor control/voice production/swallowing, (iii) dementia/brain trauma/ brain tumor, and (iv) a history of drug abuse/alcoholism or mental disease. Language samples of all 65 speakers with fluent aphasia from Cantonese AphasiaBank (Kong et al., 2010) were analyzed (44 males; mean age (in years) $\pm$ SD: 55 ± 10, range: $30 – 85$; mean years of education: 9 ± 4, range: $0 – 16$; mean post-onset period (in months): 59.5 ± 52.7, range: $6–185$). For each PWFA, an age- and education level-matched control was selected (33 males; mean age $\pm$ SD: 53 ± 10, range: $30 – 75$; mean education: 10 ± 4, range: $2–19$), totaling 65 healthy speakers. The two groups did not differ in age ($t(128) = .947$, $p = .345$) or education ($t(128) = -1.226$, $p = .222$). The PWFA group included 57 speakers with anomic aphasia (mean AQ = 90.4), six with Transcortical sensory aphasia and two with Wernicke's aphasia (mean AQ = 71.7), according to the Cantonese version of the Western Aphasia Battery (CAB; Yiu, 1992).

Procedure and data analysis

The language samples in Cantonese AphasiaBank are orthographically transcribed with part-of-speech (POS) annotation. The transcriptions are formatted using the Codes for the Human Analysis of Transcripts (CHAT; MacWhinney, 2000) and linked to audio and video recordings through a computerized analytic program named Child Language Analyses (CLAN; MacWhinney, 2003). CLAN allows one to carry out a variety of linguistic analyses, such as frequency count, lexical diversity, mean length of utterance (MLU), as well as searches for user-specified combinations of words, character strings, or words in context, etc.

In the Important Event narrative task, the participant was given the instruction - 'Please talk about an important event in your life. It can be about anything.' If the participant had difficulty coming up with one, the examiner would provide a prompt – 'Perhaps talk about a memorable event'. A second prompt would be given if the participant still showed difficulty – 'This event can take place a while ago or recently. It can be happy or sad. Try your best to think of an important or memorable event'.

The themes of the 130 samples were categorized based on the topics in Stuart et al. (1994) and those of Davidson (2004) as listed in Renvall et al. (2013b). If a suitable topic could not be found in the list, a new topic would be created based on the theme of the narrative. The search functions in CLAN were used to compile lists of different lexical items produced by the PWFA and control participants, respectively. Only real (meaningful) words produced by the participants would be identified for further analyses. In other words, data analyses included verbal paraphasias, words in incomplete utterances, and words in extraneous comments. In contrast, unintelligible segments, jargon/neologism, incomplete words during self-correction, and repetition were excluded. We then obtained the type frequencies of all parts-of-speech to find out the relative dominance of different word classes. The lexical items were further classified into (i) words produced by both groups (common lexicon), (ii) words only by PWFA (unique to PWFA), and (iii) words only by controls (unique to control). Frequency counts of lexical items in the three lists were retrieved from Cheung and Chan (1997), and the Kruskal-Wallis test was employed to compare the frequency of occurrence of items among the lists. Significance threshold was Bonferroni-corrected to p = .0125 (or .05/4) according to the number of comparisons. This approach of comparing the vocabularies of the two speaker groups, as opposed to contrasting the mean frequencies of occurrence of all words produced by PWFA vs. control participants, was motivated by our aim to understand the differences in the properties of actual lexical items uttered by the speakers, which might serve as potential training materials. Since the vocabulary of control participants is expected to be larger than that of PWFA, distinguishing common and unique items would prevent the situation of having the common set exerting different degrees of influence on the overall performance by speaker group. Consequently, one can gain a more in-depth understanding of whether differential lexical retrieval exists as a function of usage or familiarity between healthy and language-impaired speakers.

Focusing on the individual usage of the two main word classes, mean proportions of unique nouns and verbs were computed with reference to the unique items by group (i.e. the list of lexical items uniquely produced by either the PWFA or control group) and unique items by individual participants (i.e. lists of lexical items individually produced by the participants excluding the common lexicon list). Two-way mixed effects ANOVA (word class x speaker group) were conducted to detect differences between form classes and between participant groups. We additionally examined the percentages of abstract nouns and verbs in the respective 'unique' lists of the two participant groups. Applying a similar distinction as in Barca, Burani, and Arduino (2002), words denoting physical entities (e.g. objects and living beings), and actions (including verbs of saying), that can be experienced by the senses are considered "concrete"; words referring to concepts that cannot be experienced by the senses would be considered "abstract".

Results

The important events narrated by the participants in both groups can be classified into 21 topics. The list of topics and their frequency (and percentage) by speaker group are shown in table 1. The personal recounts of PWFA and those of control participants overlapped in 12 of the 21 topics, accounting for 75% and 83% of the narratives in the PWFA and control groups, respectively. It is not surprising that some themes are unique to PWFA, such as ones about their experience of stroke.

With respect to the words produced in the narratives of the two groups, the vocabulary of the control speakers is expectedly larger than that of the PWFA, 1530 vs. 1048 different lexical items, of which 450 are in common, as indicated in table 2. The table also shows the distribution of items across different parts-of-speech (POS). Since Cantonese AphasiaBank recognizes a total of 38 POS, only the most frequently produced 17 categories are listed in table 2, for illustration purposes. It was found that verbs and nouns were the most numerous POS, 49.6% of the common lexicon, 57.3% of the vocabulary unique to PWFA, and 64.3% of words unique to Control; they were followed by adverbs and adjectives, accounting for an additional 19.1%, 16.3%, and 17.1% in the respective lists. PWFA produced many grammatical words in various POS in common with the control participants, although they employed fewer different function words than control speakers in most types. Among the various function word categories, utterance final particles (UFP) and classifiers are the larger categories for both groups.

To evaluate any differential access to the lexicon between the speaker groups, word frequencies of the three lists, i.e. common, unique to PWFA and unique to control, were contrasted. Note that as Cantonese is essentially a spoken language, many colloquial lexical items do not have frequency counts in Cheung and Chan (1997), which is based on printed materials. In order to include as many Cantonese lexical items as possible in the analysis, we used the frequencies of the closest standard forms as substitutes, e.g. the standard form 爲什麼 *wai6sam6mo1* 'why' for the colloquial form 點解 *dim2gaai2* 'why' (the number in the phonetic transcription represents the lexical tone). The mean frequency ranks of the three lists are shown in table 3. Results of the Kruskal-Wallis test revealed that the common vocabulary list has a significantly higher frequency than either unique list ($p < .001$), while the two unique lists did not differ from each other after correction for multiple comparisons ($p > .0125$). Nonetheless, control speakers tend to utter words of lower frequency than PWFA.

Further analyses focused on nouns and verbs. Table 4 shows that verbs account for a higher proportion of the vocabulary than nouns for both participant groups. When the group unique list was used as a reference, the two-way mixed effect ANOVA showed a significant main effect of word class ($F(1, 128) = 29.144$, $p < .001$), but no effects of group ($p > .5$) or interaction between word class and group ($p > .4$). When the reference was the list unique to each individual participant, both effects of word class ($F(1, 128) = 9.942$, $p < .005$) and group ($F(1, 128) = 28.824$, $p < .001$) were significant, without an interaction effect ($p > .4$). More different verbs than nouns were produced by both groups, and proportionately more different nouns and verbs were employed by control than PWFA participants. Therefore, it

can be said that using individually unique items as a reference has provided greater insights into the properties of vocabulary produced by the two types of speakers. An examination of table 4 revealed that the standard deviations of PWFA of both nouns and verbs with the group unique list as a reference (table 4a) were larger than the means, suggesting great individual variations in the number of different lexical items produced. This was not the case for the control group or when the individual unique list served as the reference. Thus, the latter results (shown in table 4b) may be taken as more reliable. Finally, regarding the lexico-semantic effect of concreteness, the percentages of abstract nouns and verbs were highly comparable between the PWFA and control speakers (PWFA: 33.3% for nouns, 58.1% for verbs; Control: 37.6% for nouns, 58.6% for verbs), using the group unique noun and verb lists as references, respectively.

Discussion

The current study analyzed language samples from the Important Event narrative task in Cantonese AphasiaBank produced by speakers with fluent aphasia and their controls in terms of topic and vocabulary, with the aim to identify topics of interest to elderly speakers and the lexical items employed in talking about these topics that may be used as potential training materials for language rehabilitation. It was found that the two speaker groups had a high proportion of overlap in topics. As the task required the participants to describe an important or memorable event, the themes were not restricted to the here and now. As expected, the vocabulary of PWFA was smaller than that of control speakers; however, the lexical items employed by the two groups were similar in word frequency and in proportion of distribution across POS. Verbs, nouns, adverbs and adjectives were the four most numerous POS for both groups. In addition, PWFA and control speakers produced proportionately more different verbs than nouns, with comparable ratios of concrete and abstract words. The present finding regarding production of nouns and verbs differs from the frequently reported pattern of better performance on nominal than verbal items in the literature (e.g. Mätzig, Druks, Masterson, & Vigliocco, 2009), but is compatible with recent results from connected speech of Cantonese speakers with anomic aphasia and healthy controls when confounding psycholinguistic variables have been controlled for (Law, Kong, Lai, & Lai, 2015).

The picture that has emerged from the topic and vocabulary analyses has important implications for language rehabilitation of Cantonese speakers with aphasia. Reinforcing Davidson et al.'s (2003) finding and Renvall et al.'s (2013b) recommendation, personal recount seems to be a fruitful genre for eliciting connected speech, particularly for older speakers. Narrating important events revealed PWFA's ability to engage in topics of personal significance not limited to the present context. Although PWFA do not seem to exhibit specific deficits to lexical production as a function of concreteness and grammatical class, their far more reduced vocabulary, compared to healthy speakers, across most of the POS is highly suggestive of a direction for language training of these speakers. Comparing the lists of unique words of PWFA and control participants, healthy speakers produced almost twice as many items as PWFA in the categories of nouns, verbs, adverbs, adjectives for content words, and utterance final particles, connectives and locatives for function words. These findings may be taken to suggest that in rehabilitating PWFA, the focus is on

expanding the vocabulary of the client in major content word classes not limited to nouns and verbs, echoing the conclusion in Renvall et al. (2013a; b). Moreover, the greater lexical diversity of verbs than nouns in both PWFA and healthy speakers would suggest that training of both functionally relevant nouns and verbs be comparably emphasized for PWFA. Finally, specific function word types such as UFP and connectives may also be targeted for training. Conjunctions are known to contribute to cohesion (e.g. Sherratt, 2007) and even local coherence (Whitworth et al., 2015) of discourse. Competent employment of UFP is particularly important in spoken Chinese. As suggested in Matthews and Yip (1994), many utterances in casual conversation of Cantonese seem incomplete without UPFs. The ability to produce a greater variety of these items may, therefore, enrich and improve the interpersonal function and cohesion of PWA's discourse.

Although the current observations are based on narratives of fluent aphasic speakers, the list of topics and the results regarding vocabulary use may also provide useful guidance to clinicians for rehabilitating non-fluent aphasic speakers. For the former, clinicians may consider first focusing on themes that are most relevant to the client, such as those related to stroke. The stroke story has been proposed to have specific therapeutic value to the person with aphasia and his/her family, through providing PWA an opportunity to cope with identity and social issues (Armstrong & Ulatowska, 2007). This can be followed by discussion of topics that are common to both PWFA and their controls, starting from the most popular theme of travel. Referring to table 1, other suitable topics include recreation activities, children/grandchildren, and childhood memories. Alternatively, taking a more client-oriented approach, the clinician may consult the client for selecting the most interesting topic(s) to him/her from the list. As for remediating word retrieval, the initial focus may be on nouns and verbs, followed by other content word classes and function words. In light of previous findings of better performance reflected in type/token ratio on nouns than verbs in discourse production of Mandarin speakers with non-fluent aphasia (Packard, 1990; 1993), and the current finding of greater lexical diversity of verbs than nouns in PWFA and healthy speakers, perhaps a greater emphasis may be placed on verb training. This approach would also be consistent with the general trend to remediate verb impairment commonly observed in PWA of the non-fluent types (e.g. Bastiaanse & Jonkers, 1998; Berndt, Haendiges, Mitchum, & Sandson, 1997; Conroy, Sage, & Lambon Ralph, 2009; Thompson & Lee, 2009; see Marshall, 2015, for a review).

Finally, the content and structure of Cantonese AphasiaBank (Kong et al., 2010) render it particularly suitable for an approach to assessment and treatment of aphasia that is gaining more attention in recent years – multi-level analysis of discourse (Marini, Andreetta, del Tin, & Carlomagno, 2011; Sherratt, 2007) and multi-level treatment (Milman, Vega-Mendoza, & Clendenen, 2014; Whitworth et al., 2015). The term "multi-level" essentially refers to structural and functional analyses of connected speech and therapeutic methods targeting lexical, syntactic and discourse levels in one treatment program. The language samples of various discourse tasks including picture description, story-telling, and personal recount from healthy Cantonese speakers stratified in age, gender, and education subgroups constitute normative data against which performance of Cantonese-speaking PWA of different types can be compared. Such analyses may reveal the relationships in performance across linguistic levels and discourse genres at the individual level, which may then enable

us to make predictions or hypotheses about generalization of treatment effects both within and across linguistic levels, as advocated in Webster et al. (2015).

In summary, this short report highlights the potential of a language-specific corpus-based approach to investigate linguistic profiles of PWA for guiding language rehabilitation. A large database of language samples of PWA of different aphasia types in principle allows one to contrast characteristics of production as a function of type more reliably. Language-specific observations can facilitate the design and planning of culturally and linguistically appropriate treatments that are sensitive to the needs of individuals with communication disorders. One major challenge remains to be devising effective techniques to train and elicit lexical items of various POS and degrees of abstractness or emotions.

Acknowledgments

References

Armstrong E (2002). Variation in the discourse of non-brain-damaged speakers on a clinical task. Aphasiology, 16, 647–658.

Armstrong E, & Ulatowska H (2007). Making stories: Evaluative language and the aphasia experience. Aphasiology, 21, 763–774.

Baayen RH, Piepenbrock R, & Gulikers L (1995). The CELEX lexical database (CD-ROM) Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Barca L, Burani C, & Arduino LS (2002). Word naming times and psycholinguistic norms for Italian nouns. Behavior Research Methods, Instruments, & Computers, 34, 424–434.

Bastiaanse R, & Jonkers R (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. Aphasiology, 12, 951–969.

Berndt RS, Haendiges A, Mitchum C, & Sandson J (1997). Verb retrieval in aphasia 2: Relationship to sentence processing. Brain and Language, 56, 68–106. [PubMed: 8994699]

Borod JC, et al. (2000). Verbal pragmatics following unilateral stroke: Emotional content and valence. Neuropsychology, 14, 112–124. [PubMed: 10674803]

Brysbaert M, & New B (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior Research Methods, 41, 977–990. [PubMed: 19897807]

Carragher M, Conroy P, Sage K, & Wilkinson R (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions. Aphasiology, 26, 895–916.

Cheung Y-S, & Chan S-D (1997). Contemporary varieties of Chinese as found in Hong Kong, Taiwan, Mainland: A study of lexical difference established through analysis of a composite corpus from regional print media Hong Kong: Research Grant Council.

Conroy P, Sage KE, Lambon Ralph MA (2009). Errorless and errorful therapy for verb and noun naming in aphasia. Aphasiology, 23, 1311–1337.

Davidson B (2004). The impact of aphasia on the everyday communication of older people (Unpublished doctoral thesis) The University of Queensland, Brisbane.

Davidson B, Worrall L, & Hickson L (2003). Identifying the communication activities of older people with aphasia: Evidence from naturalistic observation. Aphasiology, 17, 243–264.

Kong AP-H, Law S-P, & Lee A (2010). Toward a multi-modal and multi-level analysis of Chinese aphasic discourse. A grant funded by the National Institutes of Health, USA.

Law S-P, Kong AP-H, Lai LW-S, & Lai C (2015). Effects of context and word class on lexical retrieval in Chinese speakers with anomic aphasia. Aphasiology, 29, 81–100. [PubMed: 25505810]

Lee J (2011). Toward a Parallel Corpus of Spoken Cantonese and Written Chinese. Proceedings of the 5th International Joint Conference on Natural Language Processing, 1462–1466.

Lee T, Lo WK, Ching PC, & Meng H (2002). Spoken Language Resources for Cantonese Speech processing. Speech Communication, 36, 327–342.

Leung M-T, & Law S-P (2001). HKCAC: The Hong Kong Cantonese Adult Language Corpus. International Journal of Corpus Linguistics, 6, 135–158.

MacWhinney B (2000). The CHILDES Project: Tools for Analyzing Talk Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney B (2003, online). Child Language Analyses (CLAN) Available at: http://childes.psy.cmu.edu/clan/ (accessed April 2015).

MacWhinney B, Fromm D, Forbes M, & Holland A (2011). AphasiaBank: Methods for studying discourse. Aphasiology, 25, 1286–1307. [PubMed: 22923879]

Marini A, Andreetta S, Del Tin S, & Carlomagno S (2011). A multi-level approach to the analysis of narrative language in aphasia. Aphasiology, 25, 1372–1392.

Marshall J (2015). Assessment and treatment of sentence processing disorders. In Hillis AE (Ed.), The Handbook of Adult Language Disorders (pp. 375–401). New York: Psychology Press.

Matthews S, & Yip V (1994). Cantonese. A Comprehensive Grammar London: Routledge.

Mätzig S, Druks J, Masterson J, & Vigliocco G (2009). Noun and verb differences in picture naming: Past studies and new evidence. Cortex, 45, 738–758. [PubMed: 19027106]

Milman L, Vega-Mendoza M, & Clendenen D (2014). Integrated training for aphasia: An application of part–whole learning to treat lexical retrieval, sentence production, and discourse-level communications in three cases of nonfluent aphasia. American Journal of Speech-Language Pathology, 23, 105–119. [PubMed: 24686892]

Packard JL (1990). Agrammatism in Chinese: A case study. In Menn L, & Obler L (Eds.), Agrammatic aphasia: A cross-language narrative sourcebook (pp. 1191–1223). Philadelphia: John Benjamins.

Packard JL (1993). A linguistic investigation of aphasic Chinese speech Dordrecht: Kluwer Academic Publishers.

Renvall K, Nickels L, & Davidson B (2013a). Functionally relevant items in the treatment of aphasia (part I): Challenges for current practice. Aphasiology, 27, 636–650.

Renvall K, Nickels L, & Davidson B (2013b). Functionally relevant items in the treatment of aphasia (part I): Further perspectives and specific tools. Aphasiology, 27, 651–677.

Sherratt S (2007). Multi-level discourse analysis: A feasible approach. Aphasiology, 21, 375–393.

Stuart S (2000). Understanding the storytelling of older adults for AAC system design. Augmentative and Alternative Communication, 16, 1–12.

Stuart S, Vanderhoof-Bilyeu D, & Beukelman D (1994). Differences in topic reference of elderly men and women. Journal of Medical Speech-Language Pathology, 2, 89–104.

Thompson C, & Lee M (2009). Psych verb production and comprehension in agrammatic Broca's aphasia. Journal of Neurolinguistics, 22, 354–369. [PubMed: 20174592]

Webster J, & Whitworth A (2012). Treating verbs in aphasia: Exploring the impact of therapy at the single word and sentence levels. International Journal of Language and Communication Disorders, 47, 619–636. [PubMed: 23121523]

Webster J, Whitworth A, & Morris J (2015). Is it time to stop "fishing"? A review of generalisation following aphasia intervention. Aphasiology, 29, 1240–1264.

Whitworth A, Claessen M, Leitão S, & Webster J (2015). Beyond narrative: Is there an implicit structure to the way in which adults organise their discourse? Clinical Linguistics and Phonetics, 29, 455–481. [PubMed: 25774761]

Whitworth A, Leitão S, Cartwright J, Webster J, Hankey GJ, Zach J, Howard D, & Wolz V (2015). NARNIA: a new twist to an old tale. A pilot RCT to evaluate a multilevel approach to improving discourse in aphasia, Aphasiology, 29, 1345–1382.

Xu L-J, & Lee T (1998). Parametric Variation in Three Chinese Dialects, Cantonese, Shanghainese and Mandarin Hong Kong: Research Grant Council.

Zhou K, Li A, Yin Z, & Zong C (2010). CASIA-CASSIL: a Chinese Telephone Conversation Corpus in Real Scenarios with Multi-leveled Annotation. The 7th International Conference on Language Resources and Evaluation, 2407–2413.

Table 1

Distribution of topics of 'Important Event' by participant group

Topic	PWFA		Control	
	Frequency	%	Frequency	%
Travel	13	20.0%	8	12.3%
Recreation activities	7	10.8%	4	6.2%
Children/grandchildren	6	9.2%	5	7.7%
Marriage	4	6.2%	2	3.1%
Health/Illness	4	6.2%	5	7.7%
Accident	3	4.6%	7	10.8%
Work/colleague	3	4.6%	5	7.7%
Pet	2	3.1%	1	1.5%
Childhood memories	2	3.1%	8	12.3%
Childbirth	2	3.1%	4	6.2%
Opinion on life	2	3.1%	1	1.5%
Death	1	1.5%	4	6.2%
Recovery after stroke	6	9.2%		
Voluntary work	3	4.6%		
Difficulties after stroke	2	3.1%		
Birthday	1	1.5%		
Humorous events			3	4.6%
Public examination (e.g. Hong Kong Advanced Level Examination, HKALE)			3	4.6%
Religion			3	4.6%
Grateful event			1	1.5%
Unfortunate events			1	1.5%
Minimal production	4	6.2%		
Total	**65**		**65**	

Table 2

Distribution of 17 parts-of-speech (POS) of vocabularies by speaker group

POS	Common lexicon		Unique to PWFA		Unique to Control		Example
	Frequency	%	Frequency	%	Frequency	%	
Verb	134	29.8%	156	25.8%	337	31.1%	打交 'to fight'
Noun	89	19.8%	189	31.5%	359	33.2%	同學 'classmate'
Adverb	51	11.1%	42	7.0%	86	8.0%	非常之 'very'
Adjective	36	8.0%	55	9.3%	98	9.1%	難忘 'unforgettable'
Utterance final particle	20	4.4%	11	1.8%	22	2.0%	㗎 indicating statement
Classifier	16	3.6%	14	2.3%	19	1.8%	次 'time'
Connective	11	2.4%	9	1.5%	16	1.5%	不過 'but'
Numeral	11	2.4%	6	1.0%	5	0.5%	兩 'two'
Direction Verb	9	2.0%	8	1.3%	12	1.1%	落嚟 'down'
Locative	9	2.0%	6	1.0%	21	1.9%	裡便 'inside'
Quantifier	7	1.6%	1	0.2%	3	0.3%	所有 'all'
Auxiliary	6	1.3%	2	0.3%	3	0.3%	可能 'can'
Pro form	6	1.3%	4	0.7%	2	0.2%	我哋 'we'
Aspect Marker	5	1.1%	3	0.5%	1	0.1%	緊 e.g. 食緊飯 progressive aspect
Interjection	5	1.1%	3	0.5%	6	0.6%	哎吔 'ouch'
Preposition	5	1.1%	9	1.5%	9	0.8%	被 passive marker
Verb particle	5	1.1%	0	0.0%	2	0.2%	定 e.g. 諗定 modifying the verb by adding a sense of certainty
Total	**425**	**94.4%**	**518**	**86.6%**	**1001**	**92.7%**	
Grand Total	450		598		1080		

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Results of Kruskal-Wallis test comparing frequency ranks of vocabularies by speaker type

	Common	Unique to PWFA	Unique to Control	*p* value
Total items	450	598	1080	
Items with available frequency count (original + substitute)	381	345	713	
Items with frequency count of substitute	75	47	131	
Mean frequency rank (original + substitute)	887.93	702.57	638.70	
Common vs. PWFA unique vs. Control unique				$\chi^2(2) = 90.118$, $p = .000$ *
Common vs. PWFA unique				$\chi^2(1) = 34.944$, $p = .000$ *
Common vs. Control unique				$\chi^2(1) = 90.249$, $p = .000$ *
PWFA unique vs. Control unique				$\chi^2(1) = 5.244$, $p = .022$

Note. "Substitute" refers to standard/formal words with written representations that are generally considered equivalent to colloquial Cantonese words without standard written forms. A high frequency rank corresponds to high word frequency.

*
Significant at Bonferroni-adjusted significance threshold of .0125 (= .05/4).

Table 4

Mean proportion (standard deviation) of individually unique nouns and verbs

a. By group unique list

	PWFA	Control
Noun	.0245 (.0299)	.0229 (.0164)
Verb	.0317 (.0346)	.0283 (.0181)

b. By individual unique list

	PWFA	Control
Noun	.0728 (.0430)	.0962 (.0361)
Verb	.0838 (.0437)	.1149 (.0346)