This article has been accepted for publication in the American Journal of Speech-Language Pathology (August 12th 2021)

Picture description of the Western Aphasia Battery Picnic scene: Reference data for the French-Canadian population

Johémie Boucher[a,b] and Amélie Brisebois[c,d*]*

Antoine Slegers[a,b], Melody Courson[b], Marianne Désilets-Barnabé[c,d], Anne-Marie Chouinard [c,d], Véronika Gbeglo[a,b], Karine Marcotte[c,d] and Simona Maria Brambati[a,b]

[a]Département de psychologie, Faculté des arts et des sciences, Université de Montréal, Montréal, Québec, Canada

[b]Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Québec, Canada.

[c]École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada.

[d]Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal (Hôpital du Sacré-Cœur de Montréal), Montréal, Québec, Canada.

**Equal status of first author*

Corresponding author:
Simona Maria Brambati
Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal (CRIUGM)
CIUSSS du Centre-Sud-de-l'Île-de-Montréal
4565 Queen-Mary, Montréal, QC, H3W 1W5, Canada
simona.maria.brambati@umontreal.ca

Abstract

Purpose: The main aim of the current study is to provide French-Canadian reference data for quantitative measures extracted from connected speech samples elicited by the Western Aphasia Battery-Revised (WAB-R) Picnic scene, a discourse task frequently used in clinical assessment of acquired language disorders.

Method: Our sample consisted of 62 healthy French-Canadian adults divided in two age groups: a 50 to 69 y.o. group and a 70 - 90 y.o. group.

Results: High inter-rater reliability scores were obtained for most of the variables. Most connected speech variables did not demonstrate an age effect. However, the 70 - 90 y.o. group produced more repetitions than the 50 to 69 y.o. group and displayed reduced communication efficiency (number of information content units per minute).

Conclusions: These findings contribute to building a reference dataset to analyze descriptive discourse production in clinical settings.

Keywords: connected speech, descriptive discourse, norms, aphasia, acquired language impairment, reference data

Introduction

Connected speech analyses assess multiple language domains and offer a relatively ecological evaluation of language in individuals with acquired language impairment. Connected speech refers to "spoken language when analyzed as a continuous sequence, as in normal utterances and conversations" (Crystal, 2008). Current research emphasizes the importance of assessing and treating language impairments beyond the single-word level and increasingly relies on discourse tasks to assess language production (Bryant, Ferguson, & Spencer, 2016). In fact, performance on speech-eliciting tasks may more accurately predict the difficulties experienced by people with language impairments such as aphasia in everyday communication contexts than scores on single-word production tasks (Bryant et al., 2016; Herbert, Hickin, Howard, Osborne, & Best, 2008). Moreover, a fine-grained analysis of connected speech can help provide valuable information about expressive language impairment and guide specific interventions (Boyle, 2020; Bryant et al., 2016).

Several tasks can be used to elicit connected speech samples. These include structured or semi-structured interviews (Glosser & Deser, 1992; Mackenzie, 2000), story-retelling procedures (Doyle et al., 2000; McNeil, Doyle, Fossett, Park, & Goda, 2001), and picture description tasks (Brookshire & Nicholas, 1994; Capilouto, Wright, & McComas Maddy, 2016; Kavé, Samuel-Enoch, & Adiv, 2009; Le Dorze & Bédard, 1998). Length and content of productions can vary significantly depending on the nature of the connected speech eliciting task (Bryant et al., 2016; Stark, 2019). Whilst thorough clinical assessment should include various discourse types (Boyle, 2020), picture descriptions are among the most widely used tasks in clinical settings. Also, while different tasks bear variations in discourse variables (Stark, 2019), data collected with structured tasks such as picture descriptions can predict performance in unstructured speech-elicited tasks (e.g., interviews), notably regarding the number of words and correct information units produced (Doyle, Goda, & Spencer, 1995). Picture descriptions consist in the detailed description of a standardized pictorial stimulus representing a complex scene. Compared to other tasks, picture descriptions present the advantage of providing a relatively constrained discourse sample with expected topics (Chenery & Murdoch, 1994) which allows a standardized approach to studying language production in context and facilitates performance comparison over time and across different groups.

Most speech-language pathologists working with patients with acquired neurogenic disorders evaluate connected speech at least at some point during their language assessment (Boyle, 2014; Bryant et al., 2016). In current clinical practices, picture description tasks are usually administered as part of larger language batteries such as the Western Aphasia Battery (WAB-R; Kertesz, 2006), the Boston Diagnostic Aphasia Examination (Goodglass, Kaplan, & Barresi, 2000), the Quick Aphasia Battery (Wilson, Eriksson, Schneck, & Lucanie, 2018), or, in French, the Montréal-Toulouse Language Battery (Nespoulous et al., 1992). These batteries generally offer qualitative grids to score connected speech. For instance, in the WAB-R, the spontaneous speech subtest qualitatively rates both the fluency (i.e., verbal productivity) and information content in structured interview and picture description contexts, on two eleven-point scales. In their Quick Aphasia Battery (QAB), Wilson et al. (2018) propose to rate connected speech elicited by an interview using a scale ranging from *severe* to *not present or within normal range* on various measures (i.e., length and complexity of utterances, speech rate, agrammatism, paragrammatism, anomia, empty speech, semantic and phonemic paraphasias, and neologisms). However, while most of these batteries can provide a quick and global approximation of the severity of language deficits, a major drawback of qualitative scoring grids is that they do not precisely and objectively describe the connected speech performance. Semi-automatic speech analysis software such as Computerized Language ANalysis (CLAN; MacWhinney, 2000) allow researchers to measure quantitative variables of connected speech production more easily.

In general, quantitative analysis of connected speech samples can inform us about the micro- and macrostructural elements of discourse production (e.g., Armstrong, 2000). Microstructural ("microlinguistic" or within-utterance) variables include lexical and grammatical variables, while macrostuctural ("macrolinguistic" or between-utterance) variables include discourse-level processing (e.g., informativeness). While interactions between the micro- and macrostructural levels (Sherratt, 2007) are inherent to connected speech production, they are not in the main scope of interest of this study and the next paragraphs will present the microstructural and macrostructural elements of discourse separately.

Several microstructural variables, such as overall verbal productivity (i.e., word quantity), utterance length, speech rate, syntactic complexity (e.g., mean number of verbs per utterance), lexical selection (e.g., open-to-closed-class words ratio, noun-to-verb ratio), lexical diversity (e.g., Moving Average Type-Token Ratio, Voc-*D*), speech errors, and disruptions to fluency (e.g., repetitions and self-corrections) can be extracted using the CLAN software (MacWhinney, 2000). Research focusing on these measures in various languages reveals significant impairments in connected speech of people with acquired language disorders relative to neurologically healthy older adults (e.g., Andreetta, Cantagallo, & Marini, 2012; Behrns, Wengelin, Broberg, & Hartelius, 2009; Boucher et al., 2020; Fergadiotis & Wright, 2016; Jaecks, Hielscher-Fastabend, & Stenneken, 2012; Marini, Caltagirone, Pasqualetti, & Carlomagno, 2007; Pashek & Tompkins, 2002; Shewan, 1988). Such detailed linguistic analyses are critical when assessing the language performance, especially among people with mild deficits, that are usually highly functional in everyday communication (Kong, 2011).

Together with microstructural variables, assessing the performance of people with acquired language disorders on macrostructural measures such as informativeness is also crucial because information content of language is closely tied to communication needs (Pritchard, Hilari, Cocks, & Dipper, 2017). Indeed, these measures seize the ability, or lack thereof, to convey relevant information (Armstrong, 2000). In a picture description task, informativeness can be quantified using a checklist of key elements (e.g., objects, people, places, actions), or *information content units*, represented in the pictorial stimulus (Ahmed, Haigh, de Jager, & Garrard, 2013). It has been demonstrated that people with aphasia convey less relevant information in speech production tasks when compared with non-brain damaged controls (e.g., Boyle, 2014; Gordon, 2008; Nicholas & Brookshire, 1995) and display reduced communication efficiency, i.e. the rate at which the relevant information is conveyed (content units/duration; e.g., Gordon, 2008; Kavé & Goral, 2017; Shewan, 1988). However, conflicting evidence have emerged concerning the impact of healthy aging on connected speech production. In fact, while previous research suggests that most healthy adults have well-preserved language production abilities after 50 y.o. (Boone, Bayles, & Koopmann, 1982; Ryan, Hutchinson, & Hull, 1980), some subtle changes are expected to occur across the adult lifespan (e.g., Capilouto et al., 2016; Kavé & Goral, 2017; Le Dorze & Bédard, 1998). These changes could be reflected in connected speech, which supports the development of reference data for specific age ranges. For instance, existing literature indicates that lexical diversity remains stable through time (Fergadiotis, Wright, & Capilouto, 2011) whereas communication efficiency (Arbuckle, Nohara-LeClair, & Pushkar, 2000; Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Mackenzie, 2000) and the proportion of main events (Capilouto et al., 2016) be influenced by some changes associated with healthy aging.

(Boone et al., 1982; Ryan et al., 1980)(Boone et al., 1982; Ryan et al., 1980)(Boone et al., 1982; Ryan et al., 1980)Additionally, one of the most reported language changes associated with aging is word finding difficulty (Abrams & Farrell, 2011). A common explanation would be supported by the transmission defect hypothesis (Burke, MacKay, Worthley, & Wade, 1991; Le Dorze & Bédard, 1998; Spieler & Griffin, 2006; Thornton & Light, 2006). According to this hypothesis, aging weakens the connection between a word's semantic (i.e., meaning of a word) and phonological (i.e., sound or appearance of a word) forms, causing some word production failures. However, the criteria for determining when word-findings difficulties become pathological remains unclear. Some connected speech analyses may offer a solution as they could be sensitive enough to detect subtle changes associated with mild language impairments (Taler & Phillips, 2008).

For pathological language behavior to be properly understood, it is fundamental to also document normal language production (Sherratt, 2007). Thus, it is crucial to collect reliable quantitative reference data from healthy controls for the various features extracted from connected speech samples. For instance, *AphasiaBank*, a shared database of language samples, provides powerful tools to analyze discourse samples (MacWhinney, Fromm, Forbes, & Holland, 2011) and has yield important work, i.e. at least 45 published papers on both pathological and healthy components of connected speech (MacWhinney & Fromm, 2016). These findings support again the relevance of collecting reference data in healthy adults, which are the backbone of standardized clinical assessment. Interestingly, AphasiaBank includes connected speech samples in many languages including a small European French dataset. However, no such data are available in French-Canadian. Even if the present study does not currently contribute to AphasiaBank, this database demonstrates the importance of collecting culturally sound reference and normative data. Despite widespread use and clinical utility of connected speech production tasks in clinical settings, current valid tools for discourse assessment in French-Canadian remain scarce and somewhat outdated (Bryant, Spencer, & Ferguson, 2017). To our knowledge, only two validated tasks exist. The Montreal-Toulouse Language Battery (Nespoulous et al., 1992) offers a quantitative scoring grid that has been validated in French and consists of a check list of information content units (Béland, Lecours, Giroux, & Bois, 1993). Also, the *Protocole Montréal d'Évaluation de la Communication* includes a grid to score conversational speech (Joanette, Ska, & Côté, 2004). Sound discourse assessment should be supported by culturally relevant reference data to support clinical advances for the French-Canadian communities. The lack of standardized tools in French-Canadian means that the interpretation of picture description productions is largely based on the clinical judgment of SLPs and neuropsychologists because it is presently based on subjective and often qualitative criteria (Garcia, Paradis, Sénécal, & Laroche, 2006), which may introduce biases in longitudinal evaluations of language. Obviously, reference data in connected speech is language dependant, i.e., a French-Canadian sample cannot be compared to data in another language. Also, as highlighted by previous normalization in French-Canadian such as with the Pyramid and Palm Trees Test (Callahan et al., 2010), it is crucial to establish normative data adapted to the cultural and linguistic reality of the target population.

Thus, the first aim of the present study is to provide reference data, including coding reliability measures, for quantitative micro- (i.e., duration, total number of words, mean length of utterance, speech rate, syntactic complexity, speech errors, lexical selection, and lexical diversity) and macrostructural (i.e., informativeness and communication efficiency) measures extracted from connected speech samples elicited by the Picnic scene picture description task in a group of healthy adults between 50 and 90 years old. Linguistic measures that are relevant in the context of language evaluation in acquired language disorders were derived from existing literature. This specific age range was chosen considering that aphasia's prevalence is highly related to age (Engelter et al., 2006; Grossman, 2010). Recent clinical guidelines (S. J. Wallace et al., 2019) recommend using the Western Aphasia Battery-Revised (WAB-R), which includes the Picnic scene picture description task, for the measurement of aphasia outcome. As opposed to story-retelling tasks and interviews, static picture description tasks provide patients with visual support which can help reduce memory load for people with severe memory deficits. Moreover, the WAB Picnic scene is useful because the key vocabulary used to describe the stimulus is believed to be acquired early in life, hence familiar to most speakers (Giles, Patterson, & Hodges, 1996).

The second aim is to determine whether there are differences between a 40 - 69 y.o. group and a 70 - 90 y.o. group in connected speech production. These age categories were determined in line with those of Capilouto et al. (2016). Considering the task and variables studied, it is expected that some subtle differences might appears between the groups. There should be no differences in content related variables (e.g., content units, lexical diversity) across groups (Fergadiotis et al., 2011). However, in line with the transmission defection hypothesis, age should likely affect time-dependent variable such as communication efficiency (e.g., Capilouto et al., 2016).

Method

Participants

A total of 62 native French-Canadian speakers, 40 women; mean age: 70.95 ± 9.43 years; mean education: 15.56 ± 4.05 years were recruited through the participants' bank of the *Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal* (CRIUGM)*,* which includes approximately 1000 adults of various ages recruited on a voluntary basis (e.g., using posters, social media, or in-person recruitment). Before they could register in the participants' bank, participants had to read the consent form available on the CRIUGM website and answer a short registration form including sociodemographic information (more information regarding the participants' bank is available online : http://www.criugm.qc.ca/en/participate.html). Participants from the bank who met the inclusion and exclusion criteria for the present study were first selected by the bank administrator. Inclusion criteria for the present study included being at least 40 years old and being fluent in French-Canadian. All participants were recruited in the Montreal (Quebec) area. Exclusion criteria included severe mental illness, acquired or developmental language impairments, neurological impairments (including neurocognitive disorders), traumatic brain injury, and self-reported uncorrected visual or auditive deficits. The selected participants were then contacted by a research assistant and asked whether they wanted to participate in a study aiming to collect normative data regarding language production in healthy older adults. These participants were included as healthy controls in larger projects directed by K.M and S.M.B. Forty-two out of the 62 participants were recruited for a project which sought to establish normative data for picture description tasks and had been approved by the ethical committee as a multicentric project at *Comité* *d'éthique de la recherche — Vieillissement et neuroimagerie du Centre intégré universitaire de santé et de services sociaux du Centre-Sud-de-l'Île-de-Montréal* (CER VN 17-18-12). The 20 remaining participants were recruited in a project which sought to investigate longitudinal changes in post-stroke aphasia (Boucher et al., 2020; Brisebois et al., 2020; Osa García et al., 2020). This project was approved by the ethical committee as a multicentric project at *Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal* (Project #MP-32-2018-1478). Written informed consent was obtained from all participants for the two projects.

Participants were divided into two age groups such as in Capilouto et al. (2016), who had formed a middle-aged (40 - 69) and an older-aged (70-89) group. In the present study, the two groups were divided as follow: a 50 - 69 y.o. group (n = 28), 18 women; mean age: 62.4 ± 5.7 years; mean education: 15.36 ± 2.9 years, and a 70 - 90 y.o. group (n = 34), 22 women; mean age 78.00 ± 4.97 years, mean education 15.74 ± 4.83 years. No significant differences were found between the groups for education, t (60) = -0.363, p = .718. Sociodemographic data are reported in Table 1 and individual sociodemographic variables of all participants are reported in Appendix 1. All participants answered a questionnaire about their health, including questions about their sight and hearing, medication, the possibility of having any mental or neurological illness, and any other health problem. To identify potential language impairments, the participants also completed at least one confrontation naming task that has been validated in French (DO80: n = 35; 30-item Boston Naming Test: n = 40; 60-item Boston Naming Test: n = 16) and a semantic association task (Pyramids and Palm Trees Test). All participants performed within normal range on these tasks, according to published norms (PPTT : Callahan et al., 2010; DO-80 : Deloche &

Hannequin, 2007; BNT-60 : Roberts & Doucet, 2011; BNT- 30 : Slegers et al., 2018). Means and standard deviations for confrontation naming and Pyramids and Palm Trees tests are presented in Appendix 2.

Table 1

Sociodemographic variables of group participants

	50 - 69 y.o. group Mean (SD)	70 - 90 y.o. group Mean (SD)	Difference test
Male (Female)	10 (18)	12 (22)	
Age (years)	62.39 (5.67)	78.00 (4.97)	t (60) = -11.54, p<0.001
Education (years)	15.36 (2.93)	15.74 (4.83)	t (60) = -0.363, p =.718

Procedure

All participants completed various language tasks, including the Picnic scene picture description task from the WAB-R. Testing took place at the *Centre de recherche de l'Institut universitaire de gériatrie de Montréal* and lasted approximately one hour, during which the participant was seated and alone with the examiner. For 42 out of the 62 participants, connected speech audio samples were recorded using a Sony IC recorder icd-px312. For the remaining participants, 20 out of the 62, the picture description samples were filmed using Sony HDR-PJ540 camera (9.2 mega pixels). Before the picture description task, the instruction given to the participants was to describe everything they saw happening in the picture, using complete sentences (« *Décrivez en détail tout ce qui se passe sur cette image en utilisant des phrases complètes* »). If the participants remained silent for more than ten seconds, they were prompted one time by the examiner with the following sentence: « Is there something you would like to add ? » *(« Avez-vous quelque chose à ajouter ? »).*

Transcriptions

For 42 out of the 62 participants, audio recordings were transcribed by the first author, a Ph.D. student in neuropsychology (J.B.), and a research assistant (V.G.) using the CLAN program (MacWhinney, 2000). As for the other 20 participants (out of 62), videos of each connected speech sample were first transcribed in ELAN (Sloetjes & Wittenburg, 2008) and imported in the CLAN software (MacWhinney, 2000) by the other first author (A.B.), who is an experienced speech and language pathologist and Ph.D. student, and students in speech-language pathology (M.D.-B. and A.-M.C.). For all samples, A.B. and J.B. trained the transcribers, transcription and utterance segmentation was made using CHAT conventions (MacWhinney, 2000), with additional guidance from french users of the program (Colin & Le Meur, 2016).

Connected Speech Measures

Various measures were extracted and analyzed. These include measures of overall verbal productivity (duration, total number of words), mean length of utterance, speech rate (words per minute), syntactic complexity (verbs/utterance), speech errors (repetitions, self-corrections, and word errors), lexical selection (open-to-closed class ratio and noun-to-verb ratio), lexical diversity (VOC-D measure), informativeness (information content units), and communication efficiency (information units/duration, information units/total number of words, and information units/total number of utterances).

Microstructural variables

All microstructural variables were extracted using the EVAL program in CLAN (MacWhinney, 2000) for each speech sample. Utterances segmentation, transcription, scoring for utterances and lexical errors was conducted following the CHAT manual guidelines (MacWhinney, 2000) with additional guidance of French users (Colin & Le Meur, 2016). Productivity measures extracted were duration of the sample, total utterances, mean length of utterance (in words), types (number of different words), tokens (total number of words), and number of words per minute. Grammatical and syntactic complexity were measured with number of verbs per utterance, noun to verb ratio, open class word to closed class word ratio. Dysfluencies were also computed: a self-correction is counted by the CLAN program every time a modification is made to one or more previous words (Schmitter-Edgecombe, Vesneski, & Jones, 2000; e.g., "*elle a* ***bloqué bouché*** *[le renvoi d'eau*]" ; she **blocked clogged** [the backwater])) and a repetition is counted by the CLAN program every time a word is inappropriately uttered more than one time (e.g., "*c'est* ***le le le*** *ballon*"; it is **the the the** ball). Lexical diversity was estimated using the Voc-*D* program in CLAN. It provides a measure of lexical diversity that is considered more robust to differences in sample length than the Token Type Ratio (TTR) (Capilouto et al., 2016). Essentially, this measure is calculated by comparing randomly sampled data from the transcript to a mathematical model representing how TTR varies with token size (cf., McKee, Malvern, & Richards, 2000 for a detailed description).

Macrostructural variables

Information Content Units (ICUs), prespecified units of accurate and relevant information conveyed by the speaker (Cooper, 1990), were also computed. ICUs were calculated by two teams of two independent examiners (J.B. and M.C, and A.B. and M.D.-B.), using a list of 30 predefined ICUs, separated in places (e.g., at the beach), people (e.g., the mother), objects (e.g., a kite), and actions (e.g., pouring [a drink]) adapted from Jensen, Chenery, and Copland (2006), The list of chosen ICUs is presented in Table 2. Examiners were all trained by J.B. and A.B. using the methods described in Jensen, Chenery, and Copland (2006). Communication efficiency was also calculated as ICUs/duration (mean number of ICU conveyed per second), ICUs/token (number of ICUs divided by total number of words), and ICUs/utterance (mean number of ICUs produced per utterance).

Table 2

List of 30 Information Content Units (ICUs) adapted from Jensen, Chenery and Copland (2006)

Key category	**Semantic Units**	**Frequency (%)**
Subjects	*Père / homme* (man 1 reading)	100.00
	Pêcheur / homme (man 2 fishing)	96.77
	Mère / femme / dame (woman pouring drink)	100.00
	Garçon / enfant (boy / child flying kite)	100.00
	Fillette / enfant / sœur (girl / child / sister playing in sand)	96.72
	Gens sur le bateau (people sailing)	46.30
	Chien (dog)	91.94
Places	*lac / eau / rivière* (in the water / on the water's edge)	85.00
	plage / sable / grève / terre / rivage / berge (on the beach)	87.10
	couverture / nappe / tapis (blanket / tablecloth / mat)	61.40
	maison / chalet (house)	93.44
	quai (on the jetty)	53.57
Objects	*Cerf-volant* (kite)	96.77
	Livre / volume (book)	28.85
	Voiture / auto (car)	78.69
	Bateau / voilier (boat / sailing ship)	90.16

Key category	Semantic Units	Frequency (%)
	Drapeau (flag)	66.67
	Radio	75.41
	Sandales / chaussures / souliers (shoes)	67.24
	Arbre (tree)	73.77
	Poisson / prise (fish, catch)	72.88
	Boisson / bouteille / vin / verre / bière / quelque chose à boire / liqueur / de l'eau / "drink" / alcool / (servir) à boire (drink)	91.53
	Château de sable (sandcastle)	89.66
Actions	*Lire (un livre) / faire la lecture* (man reading)	96.77
	Pêcher / attraper / prendre (un poisson) (man fishing)	93.44
	Verser / servir / vider (un verre de vin) / mettre de l'eau (girl pouring / having a drink)	90.32
	Jouer au / s'amuser avec / faire du / tenir un (cerf-volant) / (Le cerf-volant) vole / courir (garçon) (boy flying a kite)	93.44
	Jouer / Construire / faire / fabriquer (un château de sable) (child playing on the beach)	95
	Faire un pique-nique (couple having a picnic)	91.38
	Courir (chien) / suivre / accompagner (dog following the boy)	72.13

Data analysis

All statistical analyses were done using SPSS® v25.0 and the significance level was set at $p < .05$. Adjustment for multiple comparison was made using planned Bonferroni correction (Weisstein, 2004): we adjusted *p*-values for each level (microstructural and macrostructural analyses), as scores obtained by a participant within each dimension are considered interdependent.

Participant's *z* scores for each connected speech measure were first calculated to detect extreme scores and assumptions of normality were verified. Independent samples *t*-tests were conducted for each micro- and macrostructural variable to evaluate age-group differences. Inter-rater coding reliability was assessed using two-way random effects intraclass correlations (ICC) with a consistency model.

Inter-rater coding reliability

All variables were tested for inter-rater coding reliability. ICUs were independently scored by four of the authors (J.B., A.B., M.C., and M.D.-B.). Two-way random effects intraclass correlations (ICC) with a consistency model (McGraw & Wong, 1996; Shrout & Fleiss, 1979) were performed on all microstructural variables to determine inter-rater coding reliability. ICC is a widely used statistical approach to assess inter-rater reliability in different fields including language tasks (Marcotte et al., 2017). A subset of 19 participants were randomly selected to perform these analyses: 11 women, mean age: 68.7 ± 9.0 years; mean education: 15.1 ± 3.1 years.

Most of the variables met the threshold of high reliability (ICC > .80; Streiner & Norman, 2008). ICCs for macrostructural measures were .997 for ICU score, .984 for ICUs per minute, .992 for ICUs per word, and .922 for ICUs per utterance. Cronbach's alpha (α) was also above .80 for all microstructural variables, except for number of word errors, that reached α = .660. Detailed results are reported in Supplemental material (Appendix 3).

Results

Reference data

This section presents a summary of descriptive statistics for quantitative micro- and macrostructural measures extracted from the connected speech samples. Complete reference data are presented in Table 3. Mean length of utterance was 9.61 (*SD* = 1.93) words for the 50 - 69 y.o. group and 8.70 (*SD* = 2.47) words for the 70 - 90 y.o. group. On average, the 50 - 69 y.o. participants produced 160.91 (*SD* = 40.69) words per minute, with 2.04 (*SD* = 2.10) repetitions and 3.07 (*SD* = 2.50) self-corrections and the 70 - 90 y.o. participants produced 140.86 (*SD* = 34.81) words per minute, with 6.03 *(SD* = 4.27) repetitions and 4.06 *(SD* = 2.96) self-corrections. Mean noun to verb ratio and open to close word category ratio were respectively 6.04 (*SD* = 2.77) and 0.50 (*SD* = 0.09) for the 50 - 69 y.o. group and 6.24 (*SD* = 3.87) and 0.47 (*SD*= 0.09) for the older group. Regarding lexical diversity, mean VocD scores were 48.64 (*SD* = 9.96) and 48.64 (*SD* = 12.42) for each group. As for mascrostructural measures, the 50 - 69 y.o. participants produced on average 24.86 (*SD* = 3.20) ICUs, at a mean rate of 0.37 (*SD* = 0.17) ICUs per second and 0.14 (*SD* = 0.05) ICUs per word while the 70 - 90 y.o. participants produced on average 23.12 ICUs (*SD* = 4.46) at a mean rate of 0.28 (*SD* = 0.12) ICUs per second and 0.12 (*SD* = 0.04) ICUs per word.

Table 3

Connected speech characteristics

	50 - 69 y.o. group						70 - 90 y.o. group						Paired *t*-test
Variables	Min.	Max.	mean (SD)	median	asymetry	kurtosis	Min.	Max.	mean (SD)	median	asymetry	kurtosis	
Duration (seconds)	18	287	88.11 (59.28)	70.5	1.86	3.93	30	271	100.12 (50.14)	89.5	1.40	3.09	t (60) = -.86, p= .391
Total number of utterances	8	63	22.43 (13.65)	17.0	1.86	3.32	10	48	24.82 (10.12)	23	0.88	0.45	t (60) = -.79, p= .431
Mean lenght of utterance (words)	7.04	14.14	9.61 (1.93)	9.22	1.11	0.57	5.13	15.20	8.70 (2.47)	8.11	1.03	0.74	t (60) = 1.59, p= .117
Types (number of different words)	48	189	98.64 (36.80)	91.0	1.13	0.92	50	204	99.38 (31.76)	94	1.12	2.35	t (60) = -.01, p= .990
Tokens (total number of words)	72	572	221.04 (131.71)	176.0	1.61	2.11	68	562	224.47 (104.19)	206.5	1.32	2.62	t (60) = -.09, p= .933
Number of words per minute	78.07	246.67	160.91 (40.69)	165.76	0.12	-0.10	70.30	223.90	140.86 (34.81)	141.03	0.08	-0.45	t (60) = 2.09, p= .041

Number of verbs per utterance	0.18	0.87	0.50 (0.20)	0.51	0.25	-0.99	0.10	1.00	0.44 (0.23)	0.39	0.71	-0.13	*t* (60) = .97, *p*= .339
Noun to verb ratio	2.80	12.67	6.04 (2.77)	5.62	1.05	0.32	1.89	20.32	6.24 (3.87)	5.54	1.92	4.94	*t* (60) = -.233, *p*= .816
Open to close word category ratio	0.39	0.72	0.50 (0.09)	0.49	0.91	0.94	0.30	0.79	0.47 (0.09)	0.47	1.25	3.95	*t* (60) = 1.37, *p*= .175
Self-corrections	0	9	3.07 (2.50)	3.00	1.04	0.70	0	11	4.06 (2.96)	4.3	0.81	-0.32	*t* (60) = -1.40, *p*= .167
Repetitions	0	8	2.04 (2.10)	1.00	1.63	2.08	0	14	6.03 (4.27)	5.0	0.56	-0.91	*t* (60) = -4.51, *p*< .001***
Number of word errors	0	2	0.21 (0.50)	0.00	2.38	5.42	0	2	0.41 (0.74)	0.0	1.35	0.31	*t* (60) = -1.20, *p*= .235
VocD	28.98	68.28	48.64 (9.96)	50.07	-0.59	-0.19	30.04	79.03	48.64 (12.42)	44.38	0.52	-0.43	*t* (60) = .00, *p*= .999
Information Content Unit (ICU)	17.00	29.00	24.86 (3.20)	25.5	-0.81	0.25	12.00	29.00	23.12 (4.46)	25.0	-1.02	0.38	*t* (60) = 1.73, *p*= .089
ICUs/duration (ICUs per second)	0.10	0.84	0.37 (0.17)	0.33	1.02	1.41	0.11	0.62	0.28 (0.12)	0.26	1.24	2.26	*t* (60) = 2.58, *p*= .012*

ICUs/ token (Number of ICU per word)	0.05	0.24	0.14 (0.05)	0.14	-0.16	-0.66	0.05	0.21	0.12 (0.04)	0.12	0.30	-0.38	*t* (60) = 1.74, *p*= .086
ICUs/utt (Number of ICU per utterance)	0.43	2.27	1.35 (0.47)	1.43	-0.15	-0.38	0.53	1.71	1.05 (0.37)	1.04	0.26	-0.93	*t* (60) = 2.82, *p*= .006*

Effects of age on quantitative measures of connected speech

Table 3 shows the range, mean values, and standard deviations for all connected speech variables for both groups. The results of independent samples *t*-tests between the two age groups are summarized in the next sections and detailed in Table3.

Microstructural variables

Independent samples *t*-tests revealed no significant differences between the 50 - 69 y.o. group and the 70 - 90 y.o. group for duration, t (60) = -.86, p = .391, total utterances, t (60) = -.79, p = .431, types, t (60) = -.01, p = .990, tokens, t (60) = -.09, p = .933, number of verbs per utterance, t (60) = .97, p = .339, noun to verb ratio, t (60) = -.233, p = .816, open to close category ratio, t (60) = 1.37, p = .175, number of word errors, t (60) = -1.20, p = .235, and VocD, t (60) = .00, p = .999, mean length of utterance, t (60) = 1.59, p = .117, and number of self-corrections, t (60) = -1.40, p = .167.

While participants the 50 - 69 y.o. group were slightly more time efficient in their speech samples, producing on average 20.05 more words per minute, t (60) = 2.09, p = .041, than participants in the 70 - 90 y.o. group, this result did not survive Bonferroni correction for multiple comparisons. The 70 - 90 y.o. group produced on average 3.85 more repetitions, t (60) = -4.51, $p < .001$, than the 50 - 69 y.o. group and after adjusting for multiple comparison using Bonferroni correction this group difference remained significant ($p < .001$).

Macrostructural variables

The analyses revealed no significant differences between the groups for total ICUs and ICUs per words. Significant differences were found for number of ICUs per second and number of ICUs per utterance - in both cases, the 50 - 69 y.o. group results showed more communication efficiency when transmitting information than the 70 - 90 y.o. group. More precisely, they show a mean advantage of 0.09 ICU per second, t (60) = 2.58, p = .012, and a mean advantage of 0.30 ICU per utterance, t (60) = 2.82, p = .006, compared to the 70 - 90 y.o. group, both significant after Bonferroni correction (p =.048; p =.024).

Discussion

In the present study, we present reference data for a picture description task for healthy older French-Canadian speakers between 50 and 90 years old on an array of micro- and macrostructural measures that are relevant for aphasia assessment, demonstrate their reliability, and highlight the effects of healthy aging on connected speech production.

Previous literature had suggested that connected speech of people with acquired language disorders is characterized by significant impairments in various language domains, in comparison with healthy older adults (e.g., Andreetta et al., 2012; Behrns et al., 2009; Fergadiotis & Wright, 2016; Jaecks et al., 2012; Marini et al., 2007; Pashek & Tompkins, 2002; Shewan, 1988). That being said, in current clinical practice, because of time constraints, assessment of connected speech production is mainly based on qualitative rating scales (Bryant et al., 2017). The use of semi-automatic programs such as CLAN allows researchers to extract quantitative measures of connected speech production more easily, but the lack of reference data for healthy older adults is a major limitation that prevents from conducting this in-depth evaluation in clinical settings. It is accepted that some subtle changes in connected speech production occur during healthy aging (Capilouto et al., 2016; Kavé & Goral, 2017; Le Dorze & Bédard, 1998), but normal intraindividual fluctuations of language performance also exist (G. L. Wallace, 1999). This should be considered when assessing language in clinical populations (Sherratt, 2007). Our first aim precisely addresses to this issue. Indeed, we presented reference data of quantitative micro- and macrostructural variables for a widely used elicitation task (i.e., the WAB-R Picnic scene) in a group of healthy French-Canadian speakers between 50 and 90 years old. The psychometric properties of quantitative connected speech measures was also assessed, which is crucial in order to legitimate their use with healthy subjects as well as with clinical populations (Stark & Fukuyama, 2021). Very high inter-rater reliability scores, namely for all the variables that differentiated the 50 - 69 y.o. group from the 70 - 90 y.o. group, constitute a strength of this study. Capilouto et al. (2016) also documented such results, based on the analysis of 10% of all the transcriptions, whereas the present study presents IRR for 31% of the transcripts (19 out of 62). Indeed, disruptions of fluency (repetitions), number of words per minute, and ICUs per second all obtained high reliability scores. ICUs also obtained very high IRR scores, which supports its value in quantifying semantic content in production of descriptive discourse. In other studies, similar variables assessing informativeness yielded very good reliability scores (e.g., Correct Information Units developed by Nicholas & Brookshire, 1993). However, to our knowledge, ICUs for the WAB-R Picnic scene stimulus had not been tested for inter-judge reliability.

The second aim was to determine whether there are differences between the 50 - 69 y.o. group and the 70 - 90 y.o. group in connected speech production. The few significant effects of age found in the present study are consistent with available literature studying connected speech production in various languages and showing that healthy aging can be accompanied by slight changes in language production, mainly resulting from an increase in lexical-retrieval difficulties (e.g., Capilouto et al., 2016; Kavé & Goral, 2017). These subtle changes might be explained by word finding difficulties associated with normal aging, which are most commonly explained by the transmission defect hypothesis (Burke et al., 1991). In the present study, the 70 - 90 y.o. group produced significantly more repetitions than the 50 - 69 y.o. group. These disruptions to fluency are generally considered evidence for word-finding difficulties (Kavé & Goral, 2017). Importantly, the older group's tendency to repeat the same words is compatible with the hypothesis of word-finding difficulties originating from a "transmission defect" (Burke et al., 1991), according to which aging weakens the connection between semantic and phonological nodes. Older adults, who experience more difficulty in retrieving new words, may then more readily re-use the words that have been recently activated, hence the repetitions.

As for the reduction of communication efficiency found in the older-aged group, it is consistent with prior evidence suggesting that healthy older adults usually take more time to convey the same amount of information in connected speech production tasks (Arbuckle et al., 2000; Capilouto et al., 2016; Le Dorze & Bédard, 1998). However, it remains unclear as to whether this results from an increase in lexical retrieval difficulties (Le Dorze & Bédard, 1998) or from other age-related factors that may not be specific to language, such as general cognitive slowing or inhibition difficulties. For instance, Le Dorze and Bédard (1998) identified word-finding comments, or ''tip-of-the tongue moments'', in the picture descriptions of older subjects, which may have in some cases resulted in a reduction of communication efficiency. Moreover, in that same study, older adults produced as many content units (i.e., their speech was as informative) but more repetitions than younger adults. Thus, in the present study, older adults in the 70 - 90 y.o. group are less efficient in their overall content production than adults in the 50 - 69 y.o. group. In contrast, the reduction of communication efficiency could be explained by a general decline in the ability to inhibit irrelevant information, which results in an increase of off-topic speech with advancing age (Arbuckle et al., 2000). Interestingly, this explanation accounts for the discrepancies between tasks, i.e., the age-related decline in communication efficiency is generally more subtle in constrained tasks (e.g., picture descriptions) than in less structured tasks such as interviews (Arbuckle et al., 2000; Bortfeld et al., 2001; Mackenzie, 2000). Indeed, the latter may offer more opportunities for off-topic speech for older adults (James, Burke, Austin, & Hulme, 1998).

This study has a few limitations, which should be acknowledged. First, the sample was relatively small and had an overrepresentation of women (65% of the sample). Namely, the sample of the normative study for the pyramids and palm trees test in the Quebec-French population (Callahan et al., 2010) includes 64% of women. Also, inter-rater reliability for number of word errors (e.g., phonological and semantic errors, neologisms; see MacWhinney, 2017) was below expectations. When compared to single-word production tasks, picture descriptions implicate more elaborate language production and the set of target words is not closed. Thus, the identification of error in this context is more complex and subjective, which might have led to variability between raters. Then, as mentioned in the aims, this study contributes to a more standardized assessment of connected speech variables in French-Canadian. It remains unclear whether the investigated measures remain stable when connected speech sample are collected at multiple time points (e.g., Boyle, 2014). Thus, test-retest stability should be investigated in future studies as it was not accounted for in the present article. Another limitation would be the educational homogeneity of our sample. The participants of this study had a mean education of 15.74 ± 4.83 years, thus most of them achieved high school and some acquired a high education diploma. Even if high education is a common sampling bias(e.g., Callahan et al., 2010; Marcotte et al., 2017), it needs to be considered since education has clearly an impact on connected speech performance. Indeed, previous research in French-Canadian (e.g., Le Dorze & Bédard, 1998) has suggested that individuals with fewer years of education produced less informative speech than subjects with higher levels of education. Future studies should account for this factor. The present study extracted microstructural data using the CLAN software. However, such analyses require precise transcription using the CHAT format, which is not common practice in clinical settings, for obvious reasons including time management. Direct transfer of microstructural results into clinical practice may therefore be limited. That being said, findings regarding the overall stability of microstructural variables in adults between 50 and 90 y.o., with an expected increase however in word repetition after 70 y.o. will be useful for clinicians. Also, similarly to another discourse task with a content unit list available in French-Canadian (i.e., Montreal-Toulouse Language Battery), the ICU list for the WAB-R picnic picture could easily be used in clinical settings. For instance, clinicians will be able to compare the number of ICUs produced by a patient during the WAB-R picture description and the data provided in the present article (i.e., mean and SD of ICUs in Table 3).

This study is, to our knowledge, the first to provide reference data for several measures of connected speech elicited by the WAB-R Picnic scene that are relevant for aphasia assessment in the older French-Canadian population. As mentioned previously, picture description tasks are frequently used in clinical settings because they elicit more constrained productions and allow easier comparison across assessments. Even though the ecological value of picture descriptions is not as high as spontaneous speech, it offers a good compromise while offering more analytical dimensions compared to single-word production tasks. Assessments using quantitative measures provide detailed and essential information about language performance and are particularly important for clinicians working with people with acquired language disorders to plan for treatment, document changes, but also to identify language production difficulties, especially the milder ones (cf., Boyle, 2020; Mueller et al., 2018 for a review). A better knowledge of expected expressive language changes associated with typical aging in French-Canadian speakers also contributes to a better detection of atypical language changes that could be early indicators of language impairments related to degenerative disease. This study contributes to the first steps towards building a larger reference dataset in French-Canadian, which could be used to describe a complete and precise language profile of people with acquired language disorders in several language domains, indicating for which measure, and to which extent, they display impairments relative to healthy controls, which could help setting therapy goals and measuring outcomes from treatments. Future work in French-Canadian could include more discourse tasks, expanded reliability analyses and more diverse sociodemographic backgrounds.

Acknowledgments

This project was funded by a grant-in-aid from the Heart and Stroke Foundation (grant number G-16-00014039) to K.M. and S.M.B. K.M. and S.M.B. hold a Career Award from the "Fonds de Recherche du Québec – Santé”. A.B. holds a scholarship from the "Fonds de Recherche du Québec – Santé” and J.B. from the Canadian Institute on Health Research (CIHR).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abrams, L., & Farrell, M. T. (2011). Language processing in normal aging.

Ahmed, S., Haigh, A.-M. F., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain, 136*(12), 3727-3737. doi:10.1093/brain/awt269

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia, 50*(8), 1787-1793.

Arbuckle, T. Y., Nohara-LeClair, M., & Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychology and Aging, 15*(1), 65.

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875-892. doi:10.1080/02687030050127685

Behrns, I., Wengelin, Å., Broberg, M., & Hartelius, L. (2009). *Clinical Linguistics & Phonetics, 23*(null), 507.

Béland, R., Lecours, A. R., Giroux, F., & Bois, M. (1993). The MT-86 β aphasia battery: A subset of normative data in relation to age and level of school education (Part II). *Aphasiology, 7*(4), 359-382.

Boone, D., Bayles, K., & Koopmann, J. C. (1982). Communicative aspects of aging. *Otolaryngologic Clinics of North America, 15*(2), 313-327.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech, 44*(2), 123-147. doi:10.1177/00238309010440020101

Boucher, J., Marcotte, K., Brisebois, A., Courson, M., Houzé, B., Desautels, A., . . . Brambati, S. M. (2020). Word-finding in confrontation naming and picture descriptions produced by individuals with early post-stroke aphasia. *The Clinical Neuropsychologist*, 1-16.

Boyle, M. (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research, 57*(3), 966-978.

Boyle, M. (2020). *Choosing discourse outcome measures to assess clinical change.* Paper presented at the Semin Speech Lang.

Brisebois, A., Brambati, S. M., Désilets-Barnabé, M., Boucher, J., García, A. O., Rochon, E., . . . Marcotte, K. (2020). The importance of thematic informativeness in narrative discourse recovery in acute post-stroke aphasia. *Aphasiology*, 1-20.

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research, 37*(2), 399-407.

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics, 30*(7), 489-518. doi:10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology, 31*(10), 1105-1126.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*(5), 542-579. doi:10.1016/0749-596x(91)90026-g

Callahan, B. L., Macoir, J., Hudon, C., Bier, N., Chouinard, N., Cossette-Harvey, M., . . . Potvin, O. (2010). Normative data for the pyramids and palm trees test in the Quebec-French population. *Archives of Clinical Neuropsychology, 25*(3), 212-217.

Capilouto, G. J., Wright, H. H., & McComas Maddy, K. (2016). Microlinguistic processes that contribute to the ability to relay main events: influence of age. *Aging, Neuropsychology, and Cognition, 23*(4), 445-463.

Chenery, H. J., & Murdoch, B. E. (1994). The production of narrative discourse in response to animations in persons with dementia of the Alzheimer's type: Preliminary findings. *Aphasiology, 8*(2), 159-171. doi:10.1080/02687039408248648

Colin, C., & Le Meur, C. (2016). Adaptation du projet Aphasiabank à la langue française.

Cooper, P. V. (1990). Discourse production and normal aging: Performance on oral picture description tasks. *Journal of Gerontology, 45*(5), P210-P214.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Malden, MA: Oxford: Blackwell Pub.

Deloche, G., & Hannequin, D. (2007). *Test de dénomination orale d'images: DO 80*: Éd. du Centre de psychologie appliquée.

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology, 4*(4), 130-134.

Doyle, P. J., McNeil, M. R., Park, G., Goda, A., Rubenstein, E., Spencer, K., . . . Szwarc, L. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology, 14*(5-6), 537-549. doi:10.1080/026870300401306

Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., . . . Lyrer, P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke, 37*(6), 1379-1384.

Fergadiotis, G., & Wright, H. H. (2016). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology, 30*(4), 364-380.

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology, 25*(10), 1261-1278.

Garcia, L. J., Paradis, J., Sénécal, I., & Laroche, C. (2006). Utilisation et satisfaction à l'égard des outils en français évaluant les troubles de la communication/Use of and satisfaction with assessment tools for evaluating communication disorders in French. *Journal of Speech-Language Pathology and Audiology, 30*(4), 239-249.

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

Glosser, G., & Deser, T. (1992). *Journal of Gerontology: Psychological Sciences, 47*(null), null.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination Record Booklet*: Lippincott Williams & Wilkins.

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

Grossman, M. (2010). Primary progressive aphasia: clinicopathological correlations. *Nature Reviews Neurology, 6*(2), 88-97.

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology, 22*(2), 184-203. doi:10.1080/02687030701262613

Jaecks, P., Hielscher-Fastabend, M., & Stenneken, P. (2012). *Aphasiology, 26*(null), 953.

James, L. E., Burke, D. M., Austin, A., & Hulme, E. (1998). Production and perception of" verbosity" in younger and older adults. *Psychology and Aging, 13*(3), 355-367. doi:10.1037/0882-7974.13.3.355

Jensen, A. M., Chenery, H. J., & Copland, D. A. (2006). A comparison of picture description abilities in individuals with vascular subcortical lesions and Huntington's disease. *Journal of Communication Disorders, 39*(1), 62-77.

Joanette, Y., Ska, B., & Côté, H. (2004). *Protocole MEC, Protocole Montréal d'évaluation de la communication*: Ortho éd.

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Aging, Neuropsychology, and Cognition, 24*(5), 508-527. doi:10.1080/13825585.2016.1226249

Kavé, G., Samuel-Enoch, K., & Adiv, S. (2009). The association between age and the frequency of nouns selected for production. *Psychology and Aging, 24*(1), 17.

Kertesz, A. (2006). Western aphasia battery-revised (WAB-R). *Austin, TX: Pro-Ed*.

Kong, A. P.-H. (2011). The main concept analysis in Cantonese aphasic oral discourse: External validation and monitoring chronic aphasia. *Journal of Speech, Language, and Hearing Research*.

Le Dorze, G., & Bédard, C. (1998). Effects of age and education on the lexico-semantic content of connected speech in adults. *Journal of Communication Disorders, 31*(1), 53-71.

Mackenzie, C. (2000). Adult spoken discourse: the influences of age and education. *International Journal of Language & Communication Disorders, 35*(2), 269-285.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. transcription format and programs* (Vol. 1): Psychology Press.

MacWhinney, B. (2017). Tools for analyzing talk part 1: The chat transcription format. In: Carnegie.

MacWhinney, B., & Fromm, D. (2016). *AphasiaBank as BigData.* Paper presented at the Semin Speech Lang.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286-1307.

Marcotte, K., McSween, M.-P., Pouliot, M., Martineau, S., Pauzé, A.-M., Wiseman-Hakes, C., & MacDonald, S. (2017). Normative study of the functional assessment of verbal reasoning and executive strategies (FAVRES) test in the French-Canadian population. *Journal of Speech, Language, and Hearing Research, 60*(8), 2217-2227.

Marini, A., Caltagirone, C., Pasqualetti, P., & Carlomagno, S. (2007). Patterns of language improvement in adults with non-chronic non-fluent aphasia after specific therapies. *Aphasiology, 21*(2), 164-186. doi:10.1080/02687030600633799

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*(1), 30.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Digital Scholarship in the Humanities, 15*(3), 323-338. doi:10.1093/llc/15.3.323

McNeil, M. R., Doyle, P. J., Fossett, T. R. D., Park, G. H., & Goda, A. J. (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology, 15*(10-11), 991-1006. doi:10.1080/02687040143000348

Mueller, K. D., Koscik, R. L., Clark, L. R., Hermann, B. P., Johnson, S. C., & Turkstra, L. S. (2018). The Latent Structure and Test-Retest Stability of Connected Language Measures in the Wisconsin Registry for Alzheimer's Prevention (WRAP). *Arch Clin Neuropsychol, 33*(8), 993-1005. doi:10.1093/arclin/acx116

Nespoulous, J.-L., Lecours, A. R., Lafond, D., Lemay, A., Puel, M., Joanette, Y., . . . Rascol, A. (1992). Protocole Montréal-Toulouse d'examen linguistique de l'aphasie (MT86). *Isbergues, France: L'Ortho-Edition*.

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*(null), 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.

Osa García, A., Brambati, S. M., Brisebois, A., Désilets-Barnabé, M., Houzé, B., Bedetti, C., . . . Marcotte, K. (2020). Predicting Early Post-stroke Aphasia Outcome From Initial Aphasia Severity. *Front Neurol, 11*, 120.

Pashek, G. V., & Tompkins, C. A. (2002). Context and word class influences on lexical retrieval in aphasia. *Aphasiology, 16*(3), 261-286. doi:10.1080/02687040143000573

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders, 52*(6), 689-732.

Roberts, P. M., & Doucet, N. (2011). Performance of French-speaking Quebec adults on the Boston Naming Test. *Canadian Journal of Speech-Language Pathology & Audiology, 35*(3).

Ryan, W., Hutchinson, J., & Hull, R. (1980). Conversation: the aging speaker. *ASHA, 22*(6), 423.

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). *Archives of Clinical Neuropsychology, 15*(null), 479.

Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology, 21*(3-4), 375-393.

Shewan, C. M. (1988). *Journal of Communication Disorders, 21*(null), 103.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin, 86*(2), 420.

Slegers, A., Cole, J., Joubert, S., Escudier, F., Seni, A. G., Charbonneau, S., . . . Rouleau, I. (2018). Normes québécoises pour une version abrégée de l'Échelle de Dénomination de Boston à 30 items. *Neuropsychologie clinique et appliquée / Applied and Clinical Neuropsychology, 2*, 96-109.

Spieler, D. H., & Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes, 21*(1-3), 291-321. doi:10.1080/01690960400002133

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology, 28*(3), 1067-1083.

Stark, B. C., & Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. Language, Cognition and Neuroscience, 36(5), 562-585.

Streiner, D. L., & Norman, G. R. (2008). Reliability. In *Health Measurement Scales: A practical guide to their development and use*. Oxford, England: Oxford University Press.

Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of Clinical and Experimental Neuropsychology, 30*(5), 501-556.

Thornton, R., & Light, L. L. (2006). Language Comprehension and Production in Normal Aging. In J. E. Birren, K. W. Schaie, R. P. Abeles, M. Gatz, & T. A. Salthouse (Eds.), *Handbook of the Psychology of Aging* (pp. 261-287). Cambridge, MA: Elsevier.

Wallace, G. L. (1999). *An inclusive management approach for individuals with right hemisphere deficits.* Paper presented at the Semin Speech Lang.

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., . . . Cherney, L. R. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International journal of stroke, 14*(2), 180-185.

Weisstein, E. W. (2004). Bonferroni correction. *https://mathworld. wolfram. com/.*

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS One, 13*(2), e0192773.

Appendix 1

Individual sociodemographic variables of all participants

Participant	Sex	Age (years)	Educ.(years)
1	F	77	16
2	F	65	20
3	F	76	12
4	F	62	18
5	M	82	18
6	F	52	14
7	F	64	14
8	F	76	16
9	F	82	19
10	F	62	17
11	F	90	12
12	M	80	18
13	F	83	17
14	M	82	37
15	M	80	16
16	M	77	11
17	F	77	15
18	M	57	18
19	F	76	18
20	F	86	12
21	F	78	15
22	F	85	17
23	F	68	18
24	F	79	17
25	F	77	19
26	F	86	15
27	F	68	16
28	F	65	11
29	F	84	12
30	M	69	11

FRENCH-CANADIAN PICNIC SCENE REFERENCE DATA

Participant	Sex	Age (years)	Educ.(years)
31	F	71	11
32	M	75	11
33	F	78	12
34	F	68	13
35	F	74	11
36	F	74	12
37	F	79	11
38	M	71	18
39	M	75	16
40	M	71	15
41	M	52	19
42	F	58	17
43	F	68	15
44	F	55	14
45	M	67	15
46	M	56	13
47	M	54	11
48	F	71	16
49	F	65	18
50	F	68	13
51	F	59	22
52	F	65	13
53	M	65	15
54	M	67	19
55	M	73	16
56	M	53	16
57	F	70	16
58	F	60	14
59	M	67	15
60	F	68	11
61	M	82	14
62	M	75	24

Appendix 2

Standardized language assessment scores

	50 - 69 y.o. group Mean (SD)	70 - 90 y.o. group Mean (SD)
DO80 (*n* = 35)	78.30 (1.87; *n* = 20)	76.20 (2.57; *n* = 15)
BNT-30 (*n* = 40)	29.55 (0.99; *n* = 11)	28.79 (1.88; *n* = 29)
BNT-60 (*n* = 16)	54.09 (4.66; *n* = 11)	57.80 (1.92; *n* = 5)
PPTT (*n* = 62)	48.25 (0.89)	47.61 (1.15)

Note. DO80 = 80 items Picture Naming Test; BNT-30 = 30 items Boston Naming Test; BNT-60 = 60 items Boston Naming Test; PPTT = Pyramids and Palm Trees Test excluding items 12, 16 and 40 as suggested by Callahan et al. (2010).

Appendix 3

Inter-rater reliability for all variables (two-way random effects intraclass correlation)

Variables	Cronbach's alpha *(*α)
Duration	.993
Total number of utterances	.860
Mean lenght of utterance (MLU)	.860
Types (number of different words)	.998
Tokens (total number of words)	.998
Number of words per minute	.986
Number of verbs per utterance	.943
Noun to verb ratio	.950
Open to close word category ratio	.927
Retracings (self-corrections)	.912
Repetitions	.932
Number of word errors	.660
VocD	.874
Information Content Unit (ICU)	.997
ICUs/duration (ICUs per second)	.984
ICUs/ token (Number of ICU per word)	.992
ICUs/utt (Number of ICU per utterance)	.922