ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Adaptation and Reliability of the Cinderella Story Retell Task in Canadian French Persons Without Brain Injury

Amélie Brisebois,[a,b] Simona Maria Brambati,[b,c,d] Claudie Jutras,[b] Elizabeth Rochon,[e,f,g,h] Carol Leonard,[e,g,i] Anna Zumbansen,[i,j] Carole Anglade,[a] and Karine Marcotte[a,b]

[a] École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, Québec, Canada [b] Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal, Québec, Canada [c] Département de psychologie, Faculté des arts et des sciences, Université de Montréal, Québec, Canada [d] Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Québec, Canada [e] Department of Speech-Language Pathology, Temerty Faculty of Medicine, University of Toronto, Ontario, Canada [f] KITE Research Institute, Toronto Rehabilitation Institute, Ontario, Canada [g] Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, Toronto, Ontario, Canada [h] Rehabilitation Sciences Institute, University of Toronto, Ontario, Canada [i] School of Rehabilitation Sciences, University of Ottawa, Ontario, Canada [j] Music and Health Research Institute, University of Ottawa, Ontario, Canada

ARTICLE INFO

Article History:
Received March 21, 2023
Revision received June 20, 2023
Accepted July 22, 2023

Editor-in-Chief: Katherine C. Hustad
Editor: Tamiko Azuma

https://doi.org/10.1044/2023_AJSLP-23-00101

ABSTRACT

Purpose: Main concept (MC) analysis is a well-documented method of discourse analysis in adults with and without brain injury. This study aims to develop a MC checklist that is culturally and linguistically adapted for Canadian French speakers and examine its reliability. We also documented microstructural properties and provide a normative reference in persons not brain injured (PNBIs).
Method: Discourse samples from 43 PNBIs were collected. All participants completed the Cinderella story retell task twice. Manual transcription was performed for all samples. The 34 MCs for the Cinderella story retell task were adapted into Canadian French and used to score all transcripts. In addition, microstructural variables were extracted using Computerized Language Analysis (CLAN). Intraclass correlation coefficients were computed to assess interrater reliability for MC codes and microstructural variables. Test–retest reliability was assessed using intraclass correlations, Spearman's rho correlations, and the Wilcoxon signed-ranks test. Bland–Altman plots were used to examine the agreement of the discourse measures between the two sessions.
Results: The MC checklist for the Cinderella story retell task adapted for Canadian French speakers is provided. Good-to-excellent interrater reliability was obtained for most MC codes; however, reliability ranged from poor to excellent for the "inaccurate and incomplete" code. Microstructural variables demonstrated excellent interrater reliability. Test–retest reliability ranged from poor to excellent for all variables, with the majority falling between moderate and excellent. Bland–Altman plots illustrated the limits of agreement between test and retest.
Conclusions: This study provides the MC checklist for clinicians and researchers working with Canadian French speakers when assessing discourse with the Cinderella story retell task. It also addresses the gap in available psychometric data regarding test–retest reliability in PNBIs.
Supplemental Material: https://doi.org/10.23641/asha.24171087

The study of discourse, which is language beyond a simple clause (Armstrong, 2000), has become an increasingly important area of interest in acquired neurogenic disorders. Discourse abilities are fundamental for the realization of a large range of everyday needs and social

Correspondence to Karine Marcotte: karine.marcotte@umontreal.ca.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

participation (Dipper & Pritchard, 2017), which support their increasing importance from both clinical and research points of view. According to Frederiksen's model of discourse (Frederiksen & Stemmer, 1993), discourse production is divided into three distinct stages: (a) conceptual preparation (i.e., idea generation and macrostructural processes), (b) linguistic formulation (i.e., microstructural processes that relate to sentence processing), and (c) articulation and monitoring of the verbal message. Most studies conducting discourse analysis have focused on conceptual preparation, which represents the macrostructural level of discourse, or on linguistic formulation, also known as the microstructural level, and less on the articulation of the verbal level. Macrostructural measures refer to discourse-level processing features such as informativeness, coherence, and cohesion, whereas microstructural measures refer to within-sentence features and depict discourse's lexical and grammatical components.

Main Concept Analysis

Considered a hybrid macrostructural–microstructural approach, the main concept analysis (MCA) focuses on the proposition level of knowledge expression (Richardson & Dalton, 2016b). A MC is an utterance that contains a subject, one main verb (and its subordinate clauses), and an optional object (Nicholas & Brookshire, 1993b). Each MC consists of semantic elements considered to be essential to the story gist, and the accuracy and completeness achieved in formulating these elements by the speakers are coded using a multilevel coding system (Dalton & Richardson, 2015; Kong, 2009, 2011; Nicholas & Brookshire, 1993b, 1995; Richardson & Dalton, 2016b). MCA is useful for assessing discourse in constrained discourse tasks in either a clinical or a research setting. MCA documents the ability to convey conceptual information at the macrostructural level of discourse processing as well as the accuracy of the words and sentences used to express these conceptual elements, which is at the microstructural level. Closed sets of MCs have been developed for specific discourse tasks and cultures. Namely, they were developed for the Cookie Theft picture description task (Boston Diagnostic Aphasia Examination; Goodglass et al., 2000) for English (Nicholas & Brookshire, 1993b) and Japanese (Yazu et al., 2022) speakers; the Cinderella story retell, Broken Window picture sequence narrative, and Peanut Butter and Jelly procedural tasks (Richardson & Dalton, 2016b), as well as the picture description scene of Cat in the Tree and the sequential picture description of Refused Umbrella (Richardson & Dalton, 2020) for English speakers; four sets of sequential stimuli for Cantonese speakers (Kong, 2009) and adapted to Taiwanese Mandarin speakers (Kong & Yeh, 2015), American English speakers (Kong et al., 2016), Japanese speakers (Yazu et al., 2022), Spanish speakers (Kong, 2021), and Dutch speakers (Criel et al., 2021); a set of four discourse tasks, namely, two picture scenes and two picture sequences, developed by Nicholas and Brookshire (1993b) in young English–Spanish bilinguals (Rivera et al., 2018); and the Cat in the Tree (Nicholas & Brookshire, 1993b) for English speakers (Hameister & Nickels, 2018; Richardson & Dalton, 2020).

The MC coding used for the Cinderella story retell task of this study appears in Table 1. The first aspect scored is a concept's presence or absence (AB). If present, the concept receives one of the four following codes: accurate and complete (AC), accurate but incomplete (AI), inaccurate but complete (IC), and inaccurate and incomplete (II). AC, AI, IC, and II codes allow the examiner to analyze the quality of the information and provide more details on the overall informativeness. Detailed scoring guidelines for the Cinderella story retell task appear on the AphasiaBank website (AphasiaBank, 2022; MacWhinney et al., 2011; Richardson & Dalton, 2016a).

The MCA has been largely used to assess the discourse of adults with neurogenic language disorders (e.g., Adams, 2021; Dalton et al., 2020; Fromm et al., 2017; Kong et al., 2016; Kong & Yeh, 2015; Nicholas & Brookshire, 1995). Namely, persons with aphasia have demonstrated less accurate and complete MCs than persons not brain injured (PNBIs) without significant differences in the overall production of MCs (e.g., Kong et al., 2016; Kong & Yeh, 2015; Nicholas & Brookshire, 1995). Compelling results have also been obtained in persons with neurocognitive disorders and primary progressive aphasia. For instance, in a sample of individuals with fluent and nonfluent aphasia, individuals with Alzheimer's disease, and PNBIs, a lower degree of presence, completeness, accuracy, and efficiency of producing MCs was identified in all clinical groups compared to PNBIs (Kong et al., 2016). Similarly, 17 persons with primary progressive aphasia demonstrated less accurate and complete MCs compared to PNBIs (Dalton et al., 2020). MCA demonstrated high diagnostic sensitivity in 27 persons with subclinical aphasia (Fromm et al., 2017). Less accurate and complete MCs and more absent codes were also observed in 60 persons with latent aphasia compared to persons with anomic aphasia and PNBIs (Adams, 2021). Aging effects have also been observed in a large sample of 92 PNBIs; speakers less than 59 years of age produced more accurate and complete MCs in the Cinderella story retell task compared to speakers over 60 years of age (Richardson & Dalton, 2016a).

Reliability of MCA

MCA is easy and relatively rapid to score, which supports its clinical feasibility. Microstructural analyses rely on long transcriptions, which largely explains why

Table 1. Richardson and Dalton's (2016b) main concept (MC) scoring system.

Label	Score for each MC	Definition	Example The target is MC 2: "Cendrillon[1] vit[2] avec sa belle-mère/ses belles soeurs[3]" [Cinderella lives with stepmother/stepsisters]
Accurate and complete (AC)	3 points	The statements contain all correct information.	"C'est une jeune fille[1] qui a perdu sa mère qui vit[2] avec son père et son père s'est remarié (. . .) donc le père se remarie donc la belle-mère arrive à la maison avec les deux filles[3]" [It's a young woman who's lost her mother and lives with her dad and her dad got remarried (. . .) so her dad got remarried so the stepmother]
Accurate but incomplete (AI)	2 points	The statements contain correct pieces of information but fail to include one essential element.	"La jeune fille[1] vit[2] dans une maison" [The young woman lives in a house]
Inaccurate but complete (IC)	2 points	The statements contain at least one incorrect piece of information but mention all essential elements.	"C'est une jeune fille[1] qui vit[2] avec sa tante[3]" [It's a young woman who lives with her aunt]
Inaccurate and incomplete (II)	1 point	The statements contain at least one incorrect essential element and fail to include at least one essential element.	"Elle[1] visite[2] une maison" [She visits a house]
Absent (AB)	0 point	The statements are absent.	

Note. MC Composite (total composite score of all MCs) was computed according to Richardson and Dalton's (2016b) formula: MC = (3 × AC) + (2 × AI) + (2 × IC) + (1 × II). The superscript numbers refer to each element of the MC.

discourse analysis is less used in clinical settings (Bryant et al., 2017). Conversely, MC scoring is based on a finite set of themes, which makes it quicker to analyze, and thus reconciles quantifiable measures with clinical practical requirements. It is also strongly recommended due to its psychometric strengths, including good inter- and intrarater reliability (Boyle, 2014; Dalton & Richardson, 2015; Kong, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016b) and test–retest reliability (Kong, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995). More precisely, studies have reported point-to-point intrarater reliability of above 80% in transcripts of PNBIs and participants with aphasia (Nicholas & Brookshire, 1995) and above 90% in PNBIs (Richardson & Dalton, 2016b). Good point-to-point interrater reliability (i.e., > 80%) was also demonstrated in PNBIs and participants with aphasia (Boyle, 2014; Nicholas & Brookshire, 1993b, 1995; Richardson & Dalton, 2016b). Additionally, MCA demonstrated good test–retest reliability for close (i.e., < 3 weeks between sessions; Boyle, 2014; Kong, 2009; Nicholas & Brookshire, 1993b, 1995) and distant (i.e., 12–16 months between sessions; Kong, 2011) assessment for some MC codes. AC and AB codes reached sufficient test–retest reliability for use in research (> .70 recommended for research studies; Fitzpatrick et al., 1998; see also Boyle, 2014) and in clinical decision making (> .90; Kong, 2011). In contrast, poor reliability was obtained in statements including one or more pieces of inaccurate information (IN code in the scoring system of Nicholas & Brookshire, 1995), possibly because of the limited number of IN statements for this category (Boyle, 2014). However, the test–retest reliability of MC codes was mainly adequate when tested by combining multiple tasks into one sample (Boyle, 2014, 2015; Brookshire & Nicholas, 1994a, 1994b). Similarly, the test–retest reliability of microstructural variables has been mainly assessed using a combination of various discourse tasks (e.g., Boyle, 2014; Brookshire & Nicholas, 1994a). However, it has been recently reported for both the five separate monologic tasks and the combination of the five tasks (Stark, Alexander, et al., 2022) in persons with aphasia. Test–retest reliability was lower for PNBIs (Stark, Alexander, et al., 2022), which supports the need to determine the psychometric properties of MCA for specific populations.

Cinderella Story Retell Task

The retell task of the Cinderella story is a semispontaneous discourse elicitation method that has been primarily studied in English speakers with and without brain injury (e.g., Fergadiotis & Wright, 2011; Fromm et al., 2017; Greenslade et al., 2020; Richardson & Dalton, 2016b; Stark, 2019). The procedure (see the AphasiaBank website; Richardson & Dalton, 2016a) requires the participant to generate a story after looking at a wordless book of the Cinderella tale. Compared to single picture stimuli, sequential pictures elicited more relational ideas in PNBIs (Capilouto et al., 2005) and more story grammar episodes in individuals with and without closed head injury (Coelho, 2002). Also, the Cinderella task has elicited unique microstructural features compared to expositional and procedural discourse tasks in a large aphasia group (N = 90) and a matched-PNBI group (Stark, 2019). For instance, contrary to expositional and procedural tasks, the Cinderella task elicited the densest but least lexically diverse speech in participants with aphasia and matched PNBIs as well as the most tokens in PNBIs (Stark, 2019). These results highlight the importance of investigating the different types of discourse separately as they are mediated by different variables, such as long-term memory and executive functions in the case of Cinderella, and some tasks might be more sensitive than others on different language aspects (Stark, Alexander, et al., 2022). Similarly, in a group of 27 PNBIs, lexical diversity was significantly larger in the Cinderella story retell task compared to results obtained with single and sequential picture descriptions (Fergadiotis & Wright, 2011). In addition, using the Cinderella story retell task, a group of 27 poststroke participants who were not aphasic according to the Western Aphasia Battery–Revised (Kertesz, 2006) performed significantly different from 92 participants with anomic aphasia and 177 PNBIs on several measures, including number of words per minute, moving-average type–token ratio (MATTR; a measure of lexical diversity), and MCs (Fromm et al., 2017). In summary, discourse performance in the Cinderella story retell task has been documented at the micro- and macrostructural levels of discourse processing in adults with and without brain injury, including people with subclinical language difficulties in English.

A recent international survey of current practices in discourse assessment identified a lack of linguistic and culturally specific discourse assessment methods (Stark et al., 2021). Indeed, the scarcity of discourse protocols and normative data, including psychometric properties, was identified as a barrier to discourse assessment in nondominant languages. Although using other tasks, MCA (Kong, 2009) has been adapted, along with its respective stimuli, from Cantonese speakers to Taiwanese Mandarin speakers (Kong & Yeh, 2015), American English speakers (Kong et al., 2016), Japanese speakers (Yazu et al., 2022), Spanish speakers (Kong, 2021), and Dutch speakers (Criel et al., 2021). However, no such MC list exists in Canadian French. The Cinderella story is well known in the Canadian French culture; thus, MCA of the Cinderella story retell task is well suited for cultural and linguistic adaptation. There is also a growing need to document the psychometric properties of discourse measures, which are often influenced by the nature of discourse tasks (e.g., Capilouto et al., 2005; Stark, 2019; Stark, Alexander, et al., 2022). Additionally, knowledge about typical variability in performance in both micro- and macrostructural measures allows clinicians to differentiate "normal" fluctuations between two assessments from variations attributed to significant language changes (Boyle, 2014). Hence, the main aim of this study is to adapt MCA for the Cinderella story retell task for PNBIs who are speakers of Canadian French and examine its reliability. We also extend our work with the secondary aims of reporting microstructural measures and providing Canadian French norms for these measures in PNBIs. Similar to previous studies, good interrater reliability is expected, but lower test–retest reliability is probable in PNBIs (Stark, Alexander, et al., 2022). We believe that making this information available will improve future studies using MCA with Canadian French speakers and contribute to the advances in culturally adapted, psychometrically sound discourse analysis methods for both research and clinical settings.

Method

This project is part of a larger study approved by the ethics review board of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal (No. 2020–1900), which sought to investigate longitudinal discourse changes following a stroke and to include PNBIs. Written informed consent was obtained from all participants. We report best practice guidelines for spoken discourse research in aphasia (Stark, Bryant, et al., 2022; see Supplemental Material S1). Currently, our ethics committee does not grant permission to share individual raw data (i.e., videos and language sample transcriptions).

Participants

Initial recruitment was performed between May and August 2020 in the Montréal, Québec, area. Forty-three participants were included: 28 women, 15 men; mean age of 64.2 years (SD = 6.5); mean education level of 16.4 years (SD = 2.7). All participants performed an online assessment twice (days between sessions: M = 241.8, SD = 56.6). The inclusion criteria for this study were (a) to be at least 50 years of age and (b) to have Canadian (Québec) French as their primary language. The exclusion criteria for this study were (a) presenting a severe mental illness, (b) presenting an acquired or a developmental language impairment, (c) suffering from a neurological impairment, (d) having suffered from a traumatic brain injury, and/or (e) self-reporting uncorrected visual or auditory deficits. Cognitive screening using the videoconference version of the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), with instructions for remote administration (https://www.mocatest.org/remote-moca-testing/), was completed (M = 27.7, SD = 1.6). All participants scored within the normal range on the videoconference version of the MoCA according to French Québec normative data of videoconference administration, adjusted for age and education (Gagnon et al., 2022). Participant characteristics appear in Table 2.

Data Collection

The procedures for virtual assessment are reported in a previously published article by our team (see Marcotte et al., 2022). The retell task of the Cinderella story was administered following the AphasiaBank protocol (https://aphasia.talkbank.org/protocol/english/materials-aphasia/instructions.pdf). Participants were shown wordless images of the Cinderella book on their computer screen and asked to remember the story as they went on. The research assistant oversaw sharing and advancing the pictures, which were each presented for 10 s. Pictures were presented a second time if participants wanted to revisit previously shown pictures. Then, images were withdrawn from the screen, and participants were asked to retell the story. The instruction was, "Racontez-moi l'histoire de Cendrillon du mieux que vous pouvez. Vous pouvez utiliser tous les détails que vous connaissiez déjà de l'histoire ainsi que les images que vous venez de regarder" [Tell me the Cinderella story as well as you can. You can use any details you know about the story as well as the pictures you just looked at]. In cases where participants produced less than three utterances or remained silent for more than 10 s, the examiner prompted them with, "Que s'est-il ensuite passé?" [What happened next?] or "Allez-y" [Go on]. Participants' productions were recorded via the Zoom platform (http://www.zoom.us).

Table 2. Participant characteristics.

Variable	Values
Age	
M (*SD*)	64.23 (6.54)
Mdn [min–max]	62 [55–79]
Gender	
Female	28 (65.12%)
Male	15 (34.88%)
Handedness	
Right	39 (90.70%)
Left	4 (9.30%)
Education	
M (*SD*)	16.44 (2.73)
Mdn [min–max]	16.0 [11–25]
Time between sessions (days)	
M (*SD*)	241.77 (56.61)
Mdn [min–max]	253.0 [162–373]
Montreal Cognitive Assessment	
M (*SD*)	27.7 (1.64)
Mdn [min–max]	28.0 [24–30]
Naming score (Test de dénomination de Québec [TDQ-30])	
M (*SD*)	28.88 (1.10)
Mdn [min–max]	29.00 [26–30]

Note. min = minimum; max = maximum.

Transcription

Videos/audios of each discourse sample were imported and transcribed in ELAN (Sloetjes & Wittenburg, 2008) using CHAT conventions. Complete orthographic transcriptions were conducted, and the transcription was verbatim. The CHAT manual (MacWhinney, 2000) was used for utterance segmentation, transcription, and scoring, with additional guidance for French speakers (Colin & Le Meur, 2016). Transcriptions were performed by an experienced speech-language pathologist (A.B.) and an undergraduate student in psychology (C.J.). The same transcriber transcribed both test and retest samples from the same participant for consistency. Transcribers were blind to patient identity.

Microstructural Variable Extraction

Once the transcription was completed, morphological and grammatical information coding was conducted using the CLAN (Computerized Language Analysis) program called MOR, which tags morphemes and words under each utterance in the transcripts. Subsequently, all microstructural variables were extracted for each sample using the program EVAL of CLAN (MacWhinney, 2000). Specific CLAN commands for each variable are provided in Supplemental Material S2 (Table S1).

MC List Adaptation in Canadian French

The MCA of the Cinderella story retell task was developed originally for American English speakers (Richardson & Dalton, 2016a), and cultural adaptation requires that the target population shares a similar cultural background with the initial sample. Cultural and linguistic sound adaptations usually involve modifications, that is, developing an entirely new task (Kong, 2009) or refining the scoring protocol (Criel et al., 2021; Yazu et al., 2022). Considering that Canadian French speakers share a similar cultural background with American English speakers regarding Cinderella, an adaptation was made by refining the scoring protocol. Thus, the MC checklist was translated and adapted from Richardson and Dalton's (2016a) original list. First, we used the online, free version of DeepL Translator (DeepL Traduction—DeepL Translate, 2022) to translate the first draft of the 34 MCs in French. Second, a research assistant (C.J.), who was a native Canadian French speaker with advanced knowledge of written English, reviewed the first draft to ensure that each element was as semantically similar as possible to the original version. Third, final adjustments were made via discussion between the research assistant (C.J.), the first author (A.B.), and the principal investigator (K.M.). The final reconciled translation of the MC list is reported in the Results section.

MC Scoring

MC scoring was performed with the training materials and scoring guidelines (Richardson & Dalton, 2016a) provided on the AphasiaBank website (https://aphasia.talkbank.org/discourse/MainConcepts/) and on the repository of the Main Concept Analysis web-app (https://github.com/aphasia-apps/mainConcept; Cavanaugh et al., 2021) The transcripts were used to score MCs manually using a Microsoft Excel spreadsheet (the template is available in Supplemental Material S3). All transcripts were reviewed to identify potentially relevant MCs that were not present in the original list. None were identified.

Dependent Variables

MCs

We used Richardson and Dalton's (2016b) MC scoring system, as depicted in Table 1. The variables are MC Composite, AC, AI, IC, II, and AB.

Microstructural Variables

The initial selection of microstructural variables was based on Stark (2019) and recent literature reviews on neurocognitive disorders (Filiou et al., 2020; Slegers et al., 2018). These variables are described in Table 3 and include mean length of utterance (MLU), duration of samples, propositional density (Fromm et al., 2016), number of words per minute, number of verbs per utterance, open/closed class ratio, noun/verb ratio, number of tokens, percentage of correct information units (CIUs [hereinafter referred to as %CIU]; Nicholas & Brookshire, 1993a), CIUs per minute, and MATTR (Covington, 2007).

Data Analysis

Analysis of MC Frequency

Previous test adaptation in Canadian French has demonstrated cultural differences in performance on specific task items (e.g., Callahan et al., 2010). Hence, the frequency of each MC was computed at test and retest. As recommended by Richardson and Dalton (2016b), only the MCs that were produced by a minimum of 33% of the sample were kept in the final adaptation of the MC checklist.

Interrater Reliability

To determine interrater reliability in transcription, 19 transcripts per rater (representing 22% of the transcripts each) were selected for each of the two raters randomly. In other words, C.J. transcribed samples that were initially transcribed by A.B. and vice versa. Two-way mixed intraclass correlation coefficients (ICCs) with absolute agreement were calculated on tokens, total number of utterances, and %CIU. The total number of tokens represents the accuracy of the transcription. The number of

Table 3. Definition of the microstructural variables.

Measure	Definition	Language dimension
Duration	Duration of the sample in seconds	Corpus size
Tokens	Total number of words produced	Corpus size
MLU	Average number of words per utterance	Productivity
Propositional density	Number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words	Content richness
Words per minute	Total number of tokens divided by the duration (converted from seconds to minutes)	Fluency
Verbs per utterance	Average number of verbs (verbs, copulas, auxiliaries followed by past or present participles) per utterance	Syntactic complexity
Open/closed class ratio	Ratio of open-class words (all nouns, verbs, copulas, adjectives, and adverbs) divided by closed-class words (all other words)	Syntactic complexity
Noun/verb ratio	Ratio of nouns to verbs, excluding auxiliaries and modals	Syntactic complexity
MATTR	Average of estimated type–token ratios for successive nonoverlapping windows of fixed length	Lexical diversity
%CIU	Total number of words relevant to the stimulus and informative (CIUs) divided by the total number of words	Lexical informativeness
CIUs per minute	Total number of CIUs divided by the duration (converted from seconds to minutes)	Lexical informativeness

Note. Data derived from the CLAN software (MacWhinney et al., 2010). MLU = mean length of utterance; MATTR = moving-average type–token ratio; %CIU = percentage of correct information units; CIUs = correct information units.

utterances is critical in CHAT format since it relies uniquely on the transcriber's competence to distinguish utterance boundaries. Reliability on this measure suggests consistency in utterance segmentation throughout the samples.

To determine test–retest consistency between the two raters (A.B. and C.J.) who scored the MCs, samples from 10 participants were randomly selected. ICCs with complete agreement were calculated for all MC codes: MC Composite, AC, AI, IC, II, and AB.

Statistical Analysis of Test–Retest Reliability

Data distribution was analyzed using the Kolmogorov–Smirnov test for all dependent variables, for each session. Consistent with Stark, Alexander, et al. (2022), more than 70% of the data were not normally distributed; as a result, nonparametric tests were used throughout. Although correlation is one of the most common statistical methods used to investigate test–retest reliability, the sole use of correlations in studies dealing with replicate data is insufficient as it does not test agreement (Bland & Altman, 1986). Test–retest reliability refers to the capacity of a test or measure to replicate the same ordering between participants when tested twice (Kottner et al., 2011), whereas agreement refers to the capacity of a test or measure to provide the same result twice (Berchtold, 2016). Following the guidelines of Koo and Li (2016) in selecting the appropriate ICC, reliability between test and retest sessions was evaluated using two-way mixed ICCs with absolute agreement.

Agreement was tested using the Wilcoxon signed-ranks test to evaluate if there was a statistically significant difference between test and retest. We also measured the strength of association using Spearman's rho to assess the similarity between test and retest. The significance level was set at $p < .05$. Regarding agreement, Bland–Altman plots were produced to allow for a visual inspection of the data by examining the limits of agreement between testing points (Altman & Bland, 1983). Bland–Altman plots are scatter plots, with the y-axis representing the difference between results at test and retest and the x-axis representing the mean test–retest results. The scatter plot also illustrates the limits of agreement, with horizontal dashed lines at ±1.96 *SD*s of the mean of differences. Good agreement between test and retest is obtained if 95% of the data fall between these limits (Bland & Altman, 1999). These plots were created for the variables that obtained the best test–retest ICC.

As MCA could be useful in detecting subclinical language or cognitive deficits, we also provided minimally detectable change (MDC) for each dependent variable. MDC at a 90% confidence interval (CI; MDC90) was computed to assess the approximate change needed to be associated with clinical change, given the variance from the test–retest results (Donoghue et al., 2009). MDC90 includes the standard error of measurement (*SEM*), computed with the following formula: $SEM = SD\sqrt{1} - r$, where *SD* is the standard deviation for the obtained score distribution and r is the correlation coefficient (i.e., ICC). The formula to calculate MDC90 is $\text{MDC90} = SEM \times 1.65 \times \sqrt{(2)}$.

Analysis Software

All statistical analyses were performed using SPSS (Version 26.0). Bland–Altman plots were computed using RStudio 2022.07.2.

Results

Development of the Adapted MC List

The frequency of each MC was computed at test and retest and appears in Table 4. MCs 9, 11, and 12 did not reach the 33% frequency threshold suggested by Richardson and Dalton (2016b) and, therefore, were not included in the statistical analyses (see the Excel sheet "Modèle à copier" in Supplemental Material S3 for the checklist adapted in Canadian French). The final adapted list of MCs with the detailed scoring guide appears in Supplemental Material S4.

Interrater Reliability

Koo and Li's (2016) interpretation guidelines were used for all ICCs (interrater and test–retest reliability): below .50 = poor, between .50 and .75 = moderate, between .75 and .90 = good, and above .90 = excellent. Transcription reliability on the first assessment was excellent for total number of utterances, ICC(2, 1) = .901, 95% CI [.732, .963]; tokens, ICC(2, 1) = .997, 95% CI [.991, .999]; and %CIU, ICC(2, 1) = .985, 95% CI [.861, .994]. MC Composite scoring reliability was excellent at both test, ICC(2, 1) = .941, 95% CI [.783, .985], and retest, ICC(2, 1) = .965, 95% CI [.866, .991]. Excellent interrater reliability was also found for AC at both test, ICC(2, 1) = .932, 95% CI [.753, .983], and retest, ICC(2, 1) = .976, 95% CI [.906, .994]. IC scoring reliability was excellent at both test, ICC(2, 1) = .951, 95% CI [.815, .987], and retest, ICC(2, 1) = .915, 95% CI [.696, .978]. AB scoring reliability was excellent at both test, ICC(2, 1) = .952, 95% CI [.821, .988], and retest, ICC(2, 1) = .950, 95% CI [.813, .987]. AI scoring reliability was good at test, ICC(2, 1) = .800, 95% CI [.382, .983], whereas it was excellent at retest, ICC(2, 1) = .914, 95% CI [.694, .978]. Interrater reliability of II was, on average, poor at test, ICC(2, 1) = .533, 95% CI [−.101, .859], but excellent at

Table 4. Frequency for each main concept.

Main concept	Frequency			
	Test		Retest	
	n	%	*n*	%
1. Le père marie une femme avec deux filles. *Dad remarried a woman with two daughters.*	16	37.2	18	41.9
2. Cendrillon vit avec sa belle-mère/ses belles-sœurs. *Cinderella lives with stepmother/stepsisters.*	22	51.2	25	58.1
3. La belle-mère et les demi-sœurs étaient méchantes avec Cendrillon. *Stepmother/stepsisters were mean to Cinderella.*	32	74.4	28	65.1
4. Cendrillon était la servante de la belle-mère et des demi-sœurs. *Cinderella was a servant.*	30	65.1	23	53.5
5. Cendrillon doit faire le ménage. *Cinderella has to do the housework.*	30	69.8	31	72.1
6. Le roi pense que le prince devrait se marier. *The king thinks the prince should get married.*	24	55.8	25	58.1
7. Le roi annonce qu'il va y avoir un bal en l'honneur de son fils qui doit trouver une épouse. *King announces there is going to be a ball in honor of son who needs to find a wife.*	33	76.7	34	79.1
8. Elles ont eu une invitation au bal. *They got an invitation to the ball.*	22	51.2	22	51.2
9. Elles sont excitées à l'idée d'aller au bal.[a] *They are excited about the ball.*	10	23.3	5	11.6
10. La belle-mère dit à Cendrillon qu'elle ne peut pas aller au bal à moins que/parce que (insérer la raison). *Cinderella is told by the stepmother she cannot go to the ball unless/because (insert reason).*	32	74.4	29	67.4
11. Les demi-sœurs abîment la robe de Cendrillon.[a] *The stepsisters tore Cinderella's dress.*	13	30.2	12	27.9
12. La belle-mère et les belles-sœurs sont allées au bal.[a] *Stepmother/stepsisters went to the ball.*	14	32.6	14	32.6
13. Cendrillon était triste. *Cinderella was upset.*	20	46.5	14	32.6
14. Une fée marraine est apparue à Cendrillon. *A fairy godmother appeared to Cinderella.*	29	67.4	30	69.8
15. La fée marraine fait en sorte que {éléments} se transforment en {éléments}. *The fairy godmother makes {item(s)} turn into {items}.*	29	67.4	30	69.8
16. La fée marraine fait de Cendrillon une belle princesse. *The fairy godmother makes Cinderella into a beautiful princess.*	38	88.4	38	88.4
17. Cendrillon est allée au bal en carrosse. *Cinderella went to the ball in the coach.*	36	83.7	37	86.0
18. Elle savait qu'elle devait être à la maison parce que tout va se retransformer à minuit. *She knew she had to be home by midnight because everything will turn back at midnight.*	39	90.7	39	90.7
19. Le prince et Cendrillon ont dansé dans la salle/toute la nuit/sans personne d'autre. *The prince and Cinderella danced around the room/all night/with no one else.*	28	65.1	31	72.1
20. Le prince tombe amoureux de Cendrillon. *Prince falls in love with Cinderella.*	20	46.5	17	39.5
21. Cendrillon a réalisé qu'il est minuit. *Cinderella realized it is midnight.*	34	79.1	34	79.1
22. Elle a descendu les escaliers. *She ran down the stairs.*	40	93.0	36	83.7
23. En courant dans les escaliers, elle a perdu une de ses pantoufles de verre. *As she was running down the stairs, she lost one of her glass slippers.*	40	93.0	42	97.7
24. Le prince trouve la chaussure de Cendrillon. *The prince finds Cinderella's shoe.*	15	34.9	18	41.9
25. Tout retourne à sa forme originale. *Everything turns back to its original form.*	17	39.5	12	27.9
26. Elle est rentrée à la maison à temps. *She returned home in time.*	27	62.8	16	37.2

(*table continues*)

Table 4. *(Continued).*

Main concept	Frequency			
	Test		Retest	
	n	%	*n*	%
27. Le prince fait du porte à porte pour trouver Cendrillon. *The prince searched door to door for Cinderella.*	41	95.3	40	90.7
28. Le prince vient à la maison de Cendrillon. *Prince comes to Cinderella's house.*	17	39.5	20	46.5
29. Les demi-sœurs essayent la pantoufle de verre. *The stepsisters try on the glass slipper.*	15	34.9	16	37.2
30. La pantoufle ne faisait pas aux demi-sœurs. *The slipper didn't fit the stepsisters.*	18	41.9	17	39.5
31. Il a mis la pantoufle au pied de Cendrillon. *He puts the slipper on Cinderella's foot.*	19	44.2	21	48.8
32. La pantoufle convient parfaitement à Cendrillon. *The slipper fits Cinderella perfectly.*	34	79.1	31	72.1
33. Cendrillon et le prince se sont mariés. *Cinderella and the prince were married.*	36	83.7	29	67.4
34. Cendrillon et le prince vécurent heureux pour toujours. *Cinderella and the prince lived happily ever after.*	30	69.8	28	65.1

[a]Main Concepts 9, 11, and 12 were produced by less than 33% of the sample and were not considered in the statistical analyses.

retest, ICC(2, 1) = .950, 95% CI [.813, .987]. Supplemental Material S2 (Table S2), provides ICC interrater reliability results for MC Composite as well as for AC, AI, IC, II, and AB codes.

Test–Retest Reliability

Considering the extensiveness of the results, a summary is presented in Table 5. No systematic differences were obtained for all of the MC codes and microstructural variables, except for the coding of II, which showed a significant test–retest difference (p = .007). The strengths of the relationship between sessions ranged from weak to strong. The MC codes AC, IC, and AB as well as MC Composite obtained moderate associations between test and retest, demonstrating the highest strength of relationship. Microstructural variables demonstrated associations ranging from very weak to strong relationships between test and retest. Duration, tokens, number of words per minute, propositional density, noun/verb ratio, and CIUs per minute demonstrated strong associations, and number of verbs per utterance demonstrated moderate associations between sessions.

A summary of test–retest results, ICCs, Spearman's rho correlations, and absolute value differences is reported in Table 6. MDC90 is also presented in Table 6. The MC codes AC, AB, and MC Composite obtained good interrater reliability. The interrater reliability for IC ranged from moderate to good. For the interrater reliability of the coding of AI and II, a poor ICC was obtained. As for the microstructural variables, %CIU obtained an excellent ICC, with the CI ranging between moderate and excellent. In addition, the measures of duration, number of tokens, and number of words per minute all obtained good ICCs, with 95% CIs ranging from moderate to good.

Bland–Altman plots were created for the MC variables and the microstructural variable that obtained the best test–retest ICCs. Figure 1 illustrates the limits of agreement for the variables MC Composite, AC, and AB, whereas Figure 2 represents the limits of agreement for %CIU. Mean differences of agreement were close to zero for both AC and AB at 0.95 and 0.81, respectively. MC Composite presented a mean of differences of 2.21 between test and retest. MC Composite and AC demonstrated good agreement according to the standards of Bland and Altman (1999), with 95% of data (i.e., 41 out of 43) within ±1.96 *SD*s of the mean of differences. AB obtained 90% of the values (i.e., 40 out of 43) within limits of agreement of ±1.96 *SD*s. The mean difference of agreement between test and retest was also close to zero for %CIU, more precisely at −0.22. The variable %CIU also obtained good agreement according to the standards espoused by Bland and Altman (1999), with 95% of the data (i.e., 41 out of 43) within ±1.96 *SD*s of the mean of differences.

Discussion

This study aimed to document the test–retest reliability of MCA for the Cinderella story retell task by Canadian French speakers. To begin, a cultural and linguistic adaptation of the MC checklist of Richardson and

Table 5. Descriptive statistics of the main concept codes and microstructural variables.

Variable	Test (*n* = 43)		Retest (*n* = 43)		Statistics		Interpretation
	M **(*SD*)**	***Mdn*** **[min–max]**	***M*** **(*SD*)**	***Mdn*** **[min–max]**	***V*** **(*p* value)**	**Spearman's rho (*p* value)**	
Main concept codes							
MC Composite	55.4 (15.95)	57 [2–80]	53.2 (15.59)	55 [6–77]	355.0 (.154)	.644 (< .001)	No systematic difference, moderate relationship between sessions
AC	15.7 (5.12)	16 [0–26]	14.8 (5.21)	16 [1–25]	316.5 (.090)	.646 (< .001)	No systematic difference, moderate relationship between sessions
AI	1.2 (0.85)	1 [0–4]	1.2 (1.08)	1 [0–4]	214.0 (.790)	.286 (.063)	No systematic difference, weak relationship between sessions
IC	2.7 (1.91)	3 [0–8]	3.1 (2.24)	3 [0–12]	317.5 (.168)	.535 (< .001)	No systematic difference, moderate relationship between sessions
II	0.4 (0.54)	0.0 [0–2]	0.1 (0.41)	0.0 [0–2]	42.0 (.007)	–.070 (.655)	Significant difference between sessions, weak relationship between sessions
AB	10.9 (5.52)	10 [3–30]	11.7 (5.39)	11 [4–28]	490.5 (.159)	.640 (< .001)	No systematic difference, moderate relationship between sessions
Microstructural variables							
Duration (s)	184.5 (74.86)	186 [21–423]	180.9 (63.33)	174 [50–395]	410.5 (.608)	.722 (< .001)	No systematic difference, strong relationship between sessions
Tokens	758.2 (331.44)	688 [123–1,937]	737.9 (301.99)	685 [43–1,843]	398.0 (.365)	.765 (< .001)	No systematic difference, strong relationship between sessions
MLU (words)	14.05 (1.86)	13.94 [10.38–19.65]	13.69 (2.98)	13.33 [8.5–21.25]	412.0 (.461)	.105 (.504)	No systematic difference, very weak relationship between sessions
Propositional density	0.50 (0.03)	0.50 [0.42–0.54]	0.50 (0.03)	0.50 [0.43–0.56]	414.000 (.476)	.722 (< .001)	No systematic difference, strong relationship between sessions
Words per minute	250.76 (45.77)	248.28 [116.13–351.43]	246.00 (46.83)	241.17 [124.09–392.40]	410.0 (.447)	.722 (< .001)	No systematic difference, strong relationship between sessions
Verbs per utterance	2.31 (0.73)	2.33 [0.52–4.28]	2.27 (0.83)	2.38 [0.51–4.09]	465.0 (.923)	.503 (.001)	No systematic difference, moderate relationship between sessions
Open/closed class ratio	1.16 (0.10)	1.16 [0.89–1.37]	1.16 (0.09)	1.15 [1.02–1.39]	475.500 (.976)	.165 (.289)	No systematic difference, very weak relationship between sessions
Noun/verb ratio	1.76 (0.71)	1.56 [1.04–4.93]	1.79 (0.64)	1.61 [1.00–3.91]	488.500 (.644)	.722 (< .001)	No systematic difference, strong relationship between sessions
MATTR	0.95 (0.01)	0.95 [0.93–0.98]	0.96 (0.01)	0.96 [0.94–0.98]	633.500 (< .001)	.446 (.003)	No systematic difference, weak relationship between sessions
%CIU	58.99 (8.40)	57.24 [52.17–95.55]	59.20 (8.44)	57.45 [52.91–96.01]	485.000 (.885)	.239 (.122)	No systematic difference, very weak relationship between sessions
CIUs per minute	145.74 (22.13)	141.99 [105.96–217.14]	143.80 (25.14)	141.11 [99.69–252.00]	426.000 (.570)	.722 (< .001)	No systematic difference, strong relationship between sessions

Note. Statistical testing used the Wilcoxon signed-ranks test for paired samples (*V* = test statistic) comparing test and retest and Spearman's rho correlation assessing the strength of association between test and retest. min = minimum; max = maximum; MC Composite = main concept total composite score; AC = accurate and complete; AI = accurate but incomplete; IC = inaccurate but complete; II = inaccurate and incomplete; AB = absent; MLU = mean length of utterance; MATTR = moving-average type–token ratio; %CIU = percentage of correct information units; CIUs = correct information units.

Table 6. Summary of test–retest results.

Measure	ICC	95% CI Low–high	Koo & Li (2016) ICC quality (CI quality)	Spearman's rho		Absolute value difference between test and retest		MDC90
				r	*p*	*M* (*SD*)	Range	
Main concepts								
MC Composite	.775	[.622, .871]	Good (Moderate–good)	.644	< .001	8.77 (6.04)	1–23	17.40
AC	.707	[.521, .830]	Good (Moderate–good)	.646	< .001	3.23 (2.34)	0–10	5.71
AI	.213	[–.096, .483]	Poor	.286	.063	0.86 (0.86)	0–1	1.07
IC	.563	[.323, .736]	Moderate (Poor–moderate)	.535	< .001	1.47 (1.32)	0–4	2.30
II	.132	[–.127, .391]	Poor	–.070	.655	0.47 (1.32)	0–4	0.55
AB	.790	[.644, .880]	Good (Moderate–good)	.640	< .001	2.86 (2.12)	0–8	6.02
Microstructural								
Duration (s)	.806	[.670, .890]	Good (Moderate–good)	.722	.001	35.19 (25.18)	0.00–102	76.32
Tokens	.791	[.646, .881]	Good (Moderate–good)	.765	< .001	153.74 (116.97)	3–468	349.05
MLU (words)	.147	[–.160, .427]	Poor (Poor)	.105	.504	2.69 (1.80)	0.07–7.07	2.74
Propositional density	.538	[.284, .721]	Moderate (Poor–moderate)	.722	< .001	0.02 (0.02)	0.00–0.10	0.03
Words per minute	.747	[.579, .854]	Good (Moderate–good)	.722	< .001	26.30 (20.07)	0.43–77.26	51.01
Verbs per utterance	.566	[.322, .740]	Moderate (Poor–moderate)	.503	.001	1.78 (0.81)	0.09–3.55	0.86
Open/closed ratio	.165	[–.146, .444]	Poor (Poor)	.165	.289	0.10 (0.07)	0.01–0.31	0.10
Noun/verb ratio	.675	[.472, .810]	Moderate (Poor–good)	.722	< .001	0.34 (0.43)	0.00–1.76	0.74
MATTR	.343	[.043, .585]	Poor (Poor–moderate)	.446	.003	0.01 (0.01)	0.00–0.03	0.01
%CIU	.929	[.873, .961]	Excellent (Good–excellent)	.239	.122	2.32 (2.17)	0.06–11.33	9.26
CIUs per minute	.742	[.571, .851]	Moderate (Moderate–good)	.722	< .001	13.19 (10.86)	0.28–35.57	26.08

Note. Koo and Li (2016) give the following suggestion for interpreting intraclass correlation coefficients (ICCs), including confidence intervals (CIs): below .50 = poor, between .50 and .75 = moderate, between .75 and .90 = good, and above .90 = excellent. MDC90 = minimally detectable change at a 90% confidence interval; MC Composite = main concept total composite score; AC = accurate and complete; AI = accurate but incomplete; IC = inaccurate but complete; II = inaccurate and incomplete; AB = absent; MLU = mean length of utterance; MATTR = moving-average type–token ratio; %CIU = percentage of correct information units; CIUs = correct information units.

Figure 1. Bland–Altman plots for the main concept variables (A) MC Composite (main concept total composite score), (B) AC (accurate and complete), and (C) AB (absent).

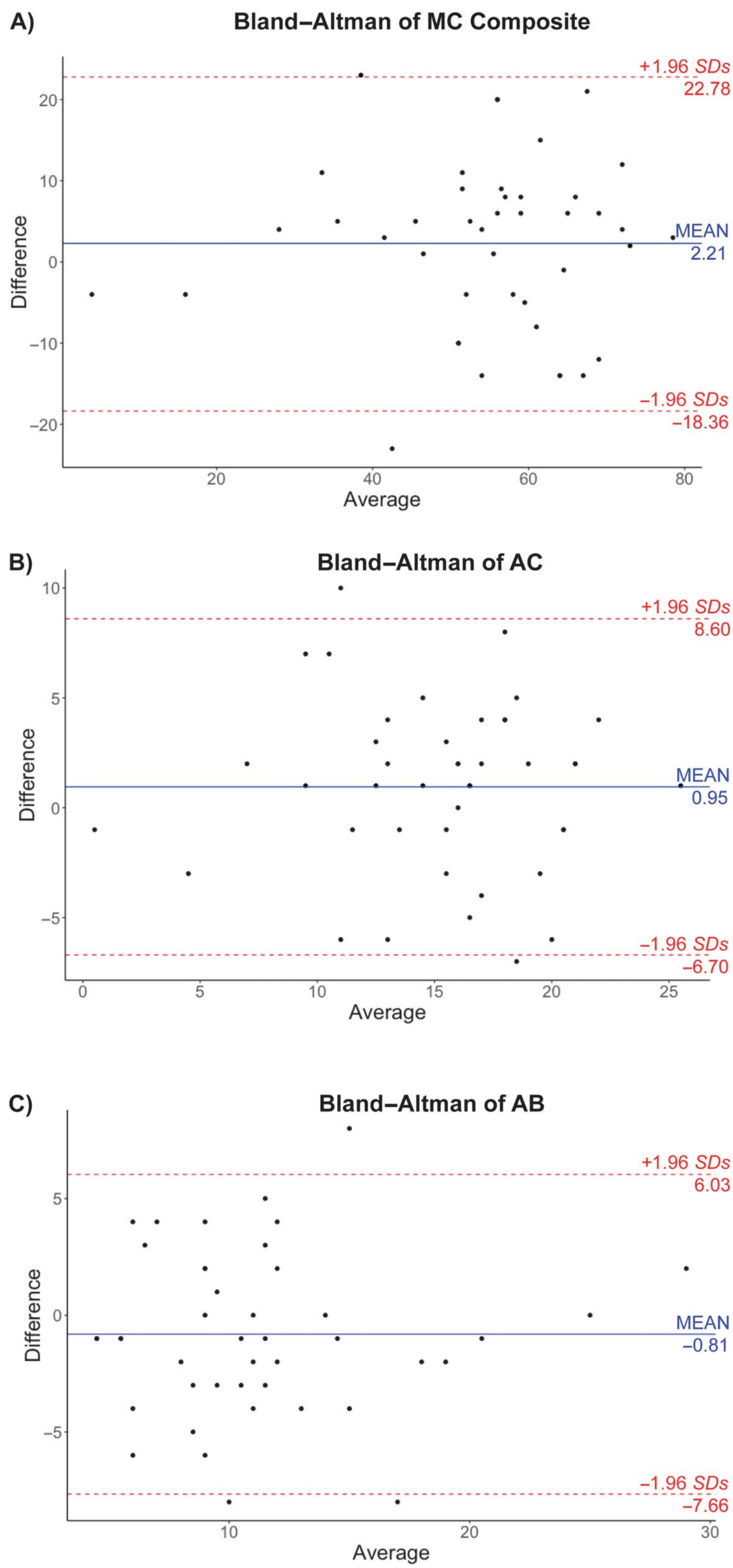
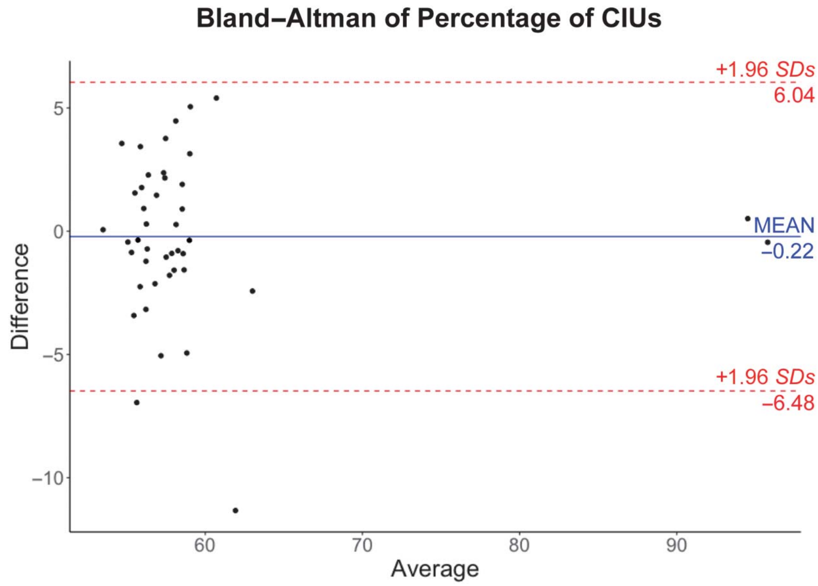


Figure 2. Bland–Altman plot for the percentage of correct information units (CIUs).



Dalton (2016a) was constructed to reflect speakers of Canadian French. We also reported microstructural measures and provided a normative reference for PNBIs. Similar to the adaptation of The Pyramid and Palm Trees Test for Canadian French speakers (Callahan et al., 2010), our adaptation of the MC list (Richardson & Dalton, 2016b) for the Cinderella story retell task, which led to the removal of three infrequent items, is now freely available (see Supplemental Materials S3 and S4). Interrater reliability results ranged from good to excellent for MC Composite, AC, AI, IC, and AB but were poor for II. Analyses of systematic differences, evaluation of the strength of the relationship, and ICCs confirmed test–retest reliability for the MC variables MC Composite, AC, and AB as well as the microstructural variables duration, number of tokens, number of words per minute, and %CIU. Conversely, the MC codes AI, IC, and II as well as the microstructural variables MLU, verbs per utterance, propositional density, noun/verb ratio, open/closed class ratio, CIUs per minute, and MATTR demonstrated poor-to-moderate test–retest reliability. MDC90 is reported for all variables, thus providing guidelines that are culturally and linguistically adapted to Canadian French speakers for the Cinderella story retell task. As a result, this discourse assessment has the potential to detect preclinical language and/or cognitive deficits.

MCA Test–Retest Reliability

Concerning test–retest reliability, all MC codes except II demonstrated no significant differences between test and retest, thus supporting our hypothesis of the stability of the coding. The relationship between test and retest sessions for the MC codes AC, IC, AB, and MC Composite was moderate. For the AI and II codes, a weak test–retest association was found. These results are, in fact, in line with previous reports of incorrect MC codes being less reliable than others in persons with aphasia (Boyle, 2014; Kong, 2011). Our study demonstrated a lower strength of relationship between test and retest sessions compared to studies conducted with participants with aphasia (Boyle, 2014; Kong, 2011). This is also consistent with expectations of higher performance variability in PNBIs (Stark, Alexander, et al., 2022).

Prior research indicates that the MC codes AC, AI, and AB are reliable for research in discourse processing (Boyle, 2014; Fitzpatrick et al., 1998). The MC codes AC, AB, and MC Composite also obtained sufficient stability over time for clinical decisions concerning persons with aphasia (Kong, 2011). In our study with PNBIs, the MC codes MC Composite, AC, and AB were sufficiently stable, thus supporting the use of these codes to conduct group research studies (i.e., ICC > .70) in Canadian French speakers. This confirms and extends previous findings with samples of English speakers. However, the II code evidenced quite different psychometric properties than the other MC codes, with poor-to-excellent interrater reliability, a significant systematic difference, and a weak relationship between test and retest. Similar to the present results, Boyle (2014) found that the IN code in persons with aphasia obtained poor test–retest correlations across three sessions. Boyle (2014) suggested that the restricted range of IN responses may have influenced such low correlations; this is also a plausible explanation concerning our data set. However, it is important to note that, contrary to this finding, Nicholas and Brookshire (1995) found high test–retest correlations for the coding of IN responses in three discourse tasks across three sessions in persons with aphasia. The inclusion of more than one task in the calculation of stability may have affected the results (Boyle, 2014). Overall, in this study, the MC variables MC Composite, AC, and AB demonstrated the best psychometric properties with no systematic difference between test and retest, moderate associations between sessions, good ICC quality, and more than 90% of data within limits of agreement of ±1.96 *SD*s. These results suggest that MC Composite, AC, and AB are the most reliable codes to assess discourse production in Canadian French–speaking PNBIs.

Microstructural Test–Retest Reliability

Regarding the microstructural variables assessed, duration, number of tokens, and number of words per minute obtained good test–retest reliability, and %CIU obtained excellent test–retest reliability. Notably, duration, number of tokens, and number of words per minute reached the criteria for inclusion in research studies (ICC > .70), and %CIU attained the criterion for clinical use (ICC > .90). However, conflicting evidence has been found for some microlinguistic variables. For instance, lexical diversity demonstrated moderate-to-good test–retest reliability in persons with aphasia (Boyle, 2014) and moderate test–retest reliability in PNBIs (Stark, Alexander, et al., 2022); however, in this study, the test–retest reliability of this measure was poor. The nature of the metrics themselves may help explain these conflicting results. Specifically, we chose the MATTR variable to assess lexical diversity because its calculation considers the variation in the length of samples, whereas the type–token ratio used by Stark, Alexander, et al. (2022) and vocabulary diversity used by Boyle (2014) do not consider this potential confound. Another point to consider is that the test–retest reliability of microstructural variables has been mainly reported for the combination of discourse tasks, as per clinical guidelines for people with aphasia (Boyle, 2015). This practice is based on the view that the assessment of multiple discourse tasks is necessary to provide a comprehensive picture of an individual's discourse abilities. With respect to word retrieval measures, including MC codes, a combination of discourse tasks has been reported to improve the test–retest reliability of measures in persons with aphasia (Boyle, 2014). This method also increases the sample size, with a minimum of 300–400 words recommended to improve test–retest reliability (Brookshire & Nicholas, 1994a). Although this study included only one task, we collected mean samples of 758 words at test and 738 words at retest, which are well above the recommended minimum length of samples to investigate test–retest reliability. Assessing discourse performance on multiple tasks was beyond the scope of this study. Nonetheless, sample sizes and the discourse task combination are considerations to keep in mind when assessing the test–retest reliability of discourse measures and in future investigations.

MDC

Our report of expected variability and MDC (MDC90) allows future studies, including subclinical or clinical population comparisons, to provide reference data for speakers of Canadian French. In literature reviews of discourse measures in people with neurocognitive diseases, microstructural variables were identified to be different in people with mild cognitive impairment compared to PNBIs in picture description tasks (Filiou et al., 2020; Slegers et al., 2018). Indeed, number of words per minute, MLU, propositional density, lexical informativeness, and lexical diversity were variables that differentiated people with mild cognitive impairment or mild Alzheimer's disease from controls. To our knowledge, MCA has not been studied in people with subjective cognitive impairment, which is the subjective presence of cognitive decline without evidence of objective cognitive impairment (Jessen et al., 2020). Subtle cognitive decline is usually not detected by standard cognitive testing, and its identification requires highly sensitive measures with robust psychometrical features (Jessen et al., 2014). We would expect MCA to be able to detect early signs of cognitive decline because it demonstrated good diagnostic sensitivity with latent (Adams, 2021) and subclinical (Fromm et al., 2017) aphasia, as well as in healthy aging individuals (Richardson & Dalton, 2016b).

Clinical Implications

The present psychometric data in Canadian French will allow future studies to test the potential use of MCA in identifying subtle language changes and subjective cognitive decline. As mentioned previously, MCA demonstrated good diagnostic sensitivity (Adams, 2021; Fromm et al., 2016; Richardson & Dalton, 2016b), which suggests that it could be a sensitive measure, with robust psychometrical features, to detect subtle cognitive decline in older adults. The cultural and linguistic adaptation of any test or list is critical to avoid any potential bias when analyzing the results. Accordingly, three MCs were removed from the original list because they were used infrequently in our group of people speaking Canadian French. Another important reason for the adaptation of the MC list for the Cinderella story retell task was that, as a measure, it is relatively easy and quick to implement in language assessments, including both PNBIs and people with aphasia. Microstructural analyses typically rely on long transcriptions, which are less used in clinical settings

(Bryant et al., 2017). Similar to our thematic units list (Brisebois et al., 2020) developed for the picnic scene of the Western Aphasia Battery–Revised (Kertesz, 2006), the MC scoring of the Cinderella story retell task is based on a finite set of themes that are more easily quantified and, thus, more suitable for clinical settings. In addition to providing reference data regarding the longitudinal changes in discourse of PNBIs for MCA, our study also enriches the data available on the microstructure of discourse for the Cinderella story retell task (Stark, Alexander, et al., 2022; Stark & Fukuyama, 2021).

Study Limitations

This study is not without some limitations. First, concerning interrater analyses of transcriptions, we conducted the analysis at only one time point. We agree that, like others, samples from both test and retest could have been included in the analysis (Stark, Alexander, et al., 2022). Nonetheless, interrater reliability was calculated in 22% of the total samples, which is consistent with previous studies (e.g., Kong, 2011; Stark, Alexander, et al., 2022). Second, the sample size is relatively small. However, this sample size is comparable to other similar studies (e.g., Richardson & Dalton, 2016a), considering the population of reference. Third, the sample may not be representative of the older population since we included only speakers from 55 to 79 years of age. Fourth, in contrast to previous studies (e.g., Stark, Alexander, et al., 2022), we chose a longer period between testing sessions, ranging from 162 to 373 days, which may better reflect changes associated with typical aging (Mueller et al., 2018). Our sample's age range does fall well within the age range whereby the first signs of some degenerative disorders appear, such as primary progressive aphasia (Mouton et al., 2022) and subjective cognitive impairment (Jessen et al., 2014). We did not administer a second cognitive screening because the MoCA was conducted at the follow-up session. Finally, no vision or hearing screenings were conducted to ensure all participants had intact and sufficient vision and hearing abilities.

Conclusions

To conclude, the assessment of discourse abilities is considered an essential part of a comprehensive language and communication evaluation for people with acquired language difficulties (Bryant et al., 2017). Studying language abilities beyond the level of utterance may be particularly useful in identifying performance differences in people with more covert language impairments (Kong, 2011). This study focused on the development of a linguistically and culturally adapted, psychometrically sound discourse measure—that of the Cinderella story retell task—for speakers of Canadian French. The scarcity of discourse protocols and normative data in Canadian French, a nondominant language in North America, is a barrier to discourse assessment for both research and clinical purposes, as reported for other nondominant languages (Stark et al., 2021). The Cinderella story is well known to speakers of Canadian French (as it is to Canadian speakers of English). Thus, the cultural adaptation of the MC list of the Cinderella story retell task (Richardson & Dalton, 2016b) was well suited for the present cultural and linguistic adaptation. Detailed information on MCA is available on the AphasiaBank website (https://aphasia.talkbank.org/discourse/MainConcepts/); however, no such data yet exist for Canadian French. The overall results provide insight into typical performance and variation, which is crucial in order to differentiate language changes due to pathology (Boyle, 2014).

Data Availability Statement

The raw data presented in this article are not readily available because of the sensitivity of the video materials. The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Acknowledgments

This project was funded by Heart and Stroke Foundation of Quebec Grants-in-Aid G-16-00014039 and G-19-0026212, awarded to Simona Maria Brambati and Karine Marcotte, respectively. Simona Maria Brambati holds a Career Award from the Fonds de recherche du Québec – Santé, and Amélie Brisebois holds a scholarship from the Fonds de recherche du Québec – Santé. We are very grateful to all the participants who contributed to this study. Also, we wish to thank Marianne Désilets-Barnabé and Noémie Desjardins for their help in data collection.

References

Adams, J. B. (2021). *Main Concept, Sequencing, and Story Grammar (MSSG) analyses of the Cinderella story in latent aphasia*. The University of New Mexico Digital Repository. https://digitalrepository.unm.edu/shs_etds/30

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society Series D: The Statistician, 32*(3), 307–317. https://doi.org/10.2307/2987937

AphasiaBank. (2022). [Homepage]. https://aphasia.talkbank.org/

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

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations, 9*. https://doi.org/10.1177/2059799116672875

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135–160. https://doi.org/10.1177/096228029900800204open_in_new

Boyle, M. (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research, 57*(3), 966–978. https://doi.org/10.1044/2014_JSLHR-L-13-0171

Boyle, M. (2015). Stability of word-retrieval errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology, 24*(4), S953–S960. https://doi.org/10.1044/2015_AJSLP-14-0152

Brisebois, A., Brambati, S. M., Désilets-Barnabé, M., Boucher, J., García, A. O., Rochon, E., Leonard, C., Desautels, A., & Marcotte, K. (2020). The importance of thematic informativeness in narrative discourse recovery in acute post-stroke aphasia. *Aphasiology, 34*(4), 472–491. https://doi.org/10.1080/02687038.2019.1705661

Brookshire, R. H., & Nicholas, L. E. (1994a). Speech sample size and test–retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research, 37*(2), 399–407. https://doi.org/10.1044/jshr.3702.399

Brookshire, R. H., & Nicholas, L. E. (1994b). *Test–retest stability of measures of connected speech in aphasia* [Clinical Aphasiology paper]. Pro-Ed. http://aphasiology.pitt.edu/163/

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. https://doi.org/10.1080/02687038.2016.1239013

Callahan, B. L., Macoir, J., Hudon, C., Bier, N., Chouinard, N., Cossette-Harvey, M., Daigle, N., Fradette, C., Gagnon, L., & 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. https://doi.org/10.1093/arclin/acq013

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

Cavanaugh, R., Richardson, J. D., & Dalton, S. G. (2021). MainConcept: An open-source web-app for scoring main concept analysis. *R package (Version 0.0.1.0000)*. https://github.com/aphasia-apps/mainConcept

Coelho, C. A. (2002). Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research, 45*(6), 1232–1248. https://doi.org/10.1044/1092-4388(2002/099)

Colin, C., & Le Meur, C. (2016). Adaptation du projet aphasiabank á la langue française: Contribution pour une évaluation informatisée du discours oral de patients aphasiques [Master's thesis, Université Paul Sabatier, Toulouse III]. Bibliothéques del'UT3. http://thesesante.ups-tlse.fr/1747/

Covington, M. A. (2007). *MATTR user manual (CASPR Research Report 2007–05)*. University of Georgia Institute for Artificial Intelligence.

Criel, Y., Deleu, M., De Groote, E., Bockstael, A., Kong, A. P.-H., & De Letter, M. (2021). The Dutch main concept analysis: Translation and establishment of normative data. *American Journal of Speech-Language Pathology, 30*(4), 1750–1766. https://doi.org/10.1044/2021_AJSLP-20-00285

Dalton, S. G. H., Hubbard, H. I., & Richardson, J. D. (2020). Moving toward non-transcription based discourse analysis in stable and progressive aphasia. *Seminars in Speech and Language, 41*(01), 032–044. https://doi.org/10.1055/s-0039-3400990

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology, 24*(4), S923–S938. https://doi.org/10.1044/2015_AJSLP-14-0161

DeepL Traduction – DeepL Translate. (2022). [Homepage]. https://www.DeepL.com/translator

Dipper, L. T., & Pritchard, M. (2017). Discourse: Assessment and therapy. In F. D. M. Fernandes (Ed.), *Advances in speech-language pathology*. Intech. https://doi.org/10.5772/intechopen.69894

Donoghue, D., Physiotherapy Research and Older People (PROP) group, & Stokes, E. K. (2009). How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *Journal of Rehabilitation Medicine, 41*(5), 343–346. https://doi.org/10.2340/16501977-0337

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Filiou, R.-P., Bier, N., Slegers, A., Houzé, B., Belchior, P., & Brambati, S. M. (2020). Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: A scoping review. *Aphasiology, 34*(6), 723–755. https://doi.org/10.1080/02687038.2019.1608502

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment, 2*(14). https://doi.org/10.3310/hta2140

Frederiksen, C. H., & Stemmer, B. (1993). Conceptual processing of discourse by a right hemisphere brain–damaged patient. In H. Brownell & Y. Joanette (Eds.), *Narrative discourse in neurologically impaired and normal aging adults* (pp. 239–278). Singular.

Fromm, D., Forbes, M., Holland, A., Dalton, S. G., Richardson, J., & MacWhinney, B. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Fromm, D., Greenhouse, J., Hou, K., Russell, G. A., Cai, X., Forbes, M., Holland, A., & MacWhinney, B. (2016). Automated proposition density analysis for discourse in aphasia. *Journal of Speech, Language, and Hearing Research, 59*(5), 1123–1132. https://doi.org/10.1044/2016_jslhr-l-15-0401

Gagnon, C., Olmand, M., Dupuy, E. G., Besnier, F., Vincent, T., Grégoire, C.-A., Lévesque, M., Payer, M., Bérubé, B., Breton, J., Lecchino, C., Bouabdallaoui, N., Iglesies-Grau, J., Gayda, M., Vitali, P., Nigam, A., Juneau, M., Hudon, C., & Bherer, L. (2022). Videoconference version of the Montreal Cognitive Assessment: Normative data for Quebec-French people aged 50 years and older. *Aging Clinical and Experimental Research, 34*(7), 1627–1633. https://doi.org/10.1007/s40520-022-02092-1

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

Greenslade, K. J., Stuart, J. E. B., Richardson, J. D., Dalton, S. G., & Ramage, A. E. (2020). Macrostructural analyses of Cinderella narratives in a large nonclinical sample. *American Journal of Speech-Language Pathology, 29*(4), 1923–1936. https://doi.org/10.1044/2020_AJSLP-19-00151

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

Jessen, F., Amariglio, R. E., Buckley, R. F., van der Flier, W. M., Han, Y., Molinuevo, J. L., Rabin, L., Rentz, D. M., Rodriguez-Gomez, O., Saykin, A. J., Sikkes, S. A. M., Smart, C. M., Wolfsgruber, S., & Wagner, M. (2020). The characterisation of subjective cognitive decline. *The Lancet Neurology, 19*(3), 271–278. https://doi.org/10.1016/S1474-4422(19)30368-0

Jessen, F., Amariglio, R. E., van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K. A., van der Flier, W. M., Glodzik, L., van Harten, A. C., de Leon, M. J., McHugh, P., Mielke, M. M., Molinuevo, J. L., Mosconi, L., Osorio, R. S., Perrotin, A., . . . Wagner, M. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia, 10*(6), 844–852. https://doi.org/10.1016/j.jalz.2014.01.001

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

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

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, 54*(1), 148–159. https://doi.org/10.1044/1092-4388(2010/09-0240)

Kong, A. P.-H. (2021). Dialectally-sensitive norms of the Spanish version of main concept analysis (Span-MCA) for quantifying neurogenically impaired spoken discourse. *Revista de Investigación en Logopedia, 11*(2), Article e69932. https://doi.org/10.5209/rlog.69932

Kong, A. P.-H., Whiteside, J., & Bargmann, P. (2016). The main concept analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. *Logopedics Phoniatrics Vocology, 41*(3), 129–141. https://doi.org/10.3109/14015439.2015.1041551

Kong, A. P.-H., & Yeh, C.-C. (2015). A Taiwanese Mandarin Main Concept Analysis (TM-MCA) for quantification of aphasic oral discourse. *International Journal of Language & Communication Disorders, 50*(5), 580–592. https://doi.org/10.1111/1460-6984.12157

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/J.JCM.2016.02.012

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology, 64*(1), 96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002

MacWhinney, B. (2000). *The CHILDES Project: Tolls for Analyzing Talk, 3rd Edition*. Lawrence Erlbaum Associates.

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

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology, 24*(6–8), 856–868. https://doi.org/10.1080/02687030903452632

Marcotte, K., Lachance, A., Brisebois, A., Mazzocca, P., Désilets-Barnabé, M., Desjardins, N., & Brambati, S. M. (2022). Validation of videoconference administration of picture description from the Western Aphasia Battery–Revised in neurotypical Canadian French speakers. *American Journal of Speech-Language Pathology, 31*(6), 2825–2834. https://doi.org/10.1044/2022_AJSLP-22-00084

Mouton, A., Plonka, A., Fabre, R., Tran, T. M., Robert, P., Macoir, J., Manera, V., & Gros, A. (2022). The course of primary progressive aphasia diagnosis: A cross-sectional study. *Alzheimer's Research & Therapy, 14*(1), Article 64. https://doi.org/10.1186/s13195-022-01007-6

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

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MOCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

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

Nicholas, L. E., & Brookshire, R. H. (1993b). *A system for scoring main concepts in the discourse of non-brain-damaged and aphasic speakers* [Clinical Aphasiology paper]. Pro-Ed. http://aphasiology.pitt.edu/1440/

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

Richardson, J. D., & Dalton, S. G. (2016a). *AphasiaBank | Main concept analysis*. https://aphasia.talkbank.org/discourse/MainConcepts/

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

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

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

Slegers, A., Filiou, R.-P., Montembeault, M., & Brambati, S. M. (2018). Connected speech features from picture description in Alzheimer's disease: A systematic review. *Journal of Alzheimer's Disease, 65*(2), 519–542. https://doi.org/10.3233/JAD-170881

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category-ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. https://archive.mpi.nl/tla/elan

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology, 28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Stark, B. C., Alexander, J. M., Hittson, A., Doub, A., Igleheart, M., Streander, T., & Jewell, E. (2022). *Test–retest reliability of microlinguistic information derived from spoken discourse in persons with chronic aphasia.* OSF Preprints. https://doi.org/10.31219/osf.io/b3s27

Stark, B. C., Bryant, L., Themistocleous, C., den Ouden, D.-B., & Roberts, A. C. (2022). Best practice guidelines for reporting spoken discourse in aphasia and neurogenic communication disorders. *Aphasiology, 37*(5), 761–784. https://doi.org/10.1080/02687038.2022.2039372

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., & Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language, and Hearing Research, 64*(11), 4366–4389. https://doi.org/10.1044/2021_JSLHR-20-00708

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. https://doi.org/10.1080/23273798.2020.1862258

Yazu, H., Kong, A. P.-H., Yoshihata, H., & Okubo, K. (2022). Adaptation and validation of the main concept analysis of spoken discourse by native Japanese adults. *Clinical Linguistics & Phonetics, 36*(1), 17–33. https://doi.org/10.1080/02699206.2021.1915385

Copyright of American Journal of Speech-Language Pathology is the property of American Speech-Language-Hearing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.