Article publié dans l'International Journal of Language and Communication Disorders et qui peut être retracé à l'adresse suivante: http://doi.org/10.1111/1460-6984.12844

The longitudinal trajectory of discourse from the hyperacute to the chronic phase in mild to moderate poststroke aphasia recovery: A case series study

Amélie Brisebois[a,b], Simona Maria Brambati[c,d], Elizabeth Rochon[e,f,g,h], Carol Leonard[e,g,i], and Karine Marcotte[a,b]

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

[b]*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.*

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

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

[e] *Department of Speech-Language Pathology, Faculty of Medicine, University of Toronto, Toronto, Canada*

[f] *KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, Canada*

[g] *Heart and Stroke Foundation, Canadian Partnership for Stroke Recovery, Ontario, Canada*

[h] *Rehabilitation Sciences Institute, University of Toronto*

[i] *School of Rehabilitation Sciences, University of Ottawa*

Corresponding author:
Karine Marcotte, Ph.D.
Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal
5400 Gouin Ouest, Montréal, Québec, Canada, H4J 1C5
Phone number: 514-338-2222 extension 7710
Fax number: 514-340-2115 E-mail : karine.marcotte@umontreal.ca

What this paper adds

Multi-level discourse analysis allows for in-depth analysis of underlying discourse processes. To date, very little is known on the longitudinal discourse changes from aphasia onset through to the chronic stage of recovery. This study documents multi-level discourse features in four people with mild to moderate aphasia in the hyperacute, acute, subacute and chronic stage of post-stroke aphasia recovery.

The study found that most discourse variables demonstrated improvement throughout time. Macrostructural variables of coherence and thematic units improved throughout the continuum whereas microstructural variables demonstrated greater gains in the early compared to the late period of recovery.

This study suggests that multilevel discourse analysis will allow a better understanding of post-stroke aphasia recovery, although more research is needed to determine the clinical utility of these findings. Future research may wish to investigate longitudinal discourse recovery in a larger sample of people with aphasia with heterogenous aphasia profiles and severities.

Abstract

Background: Discourse analysis has recently received much attention in aphasia literature. Even if post-stroke language recovery occurs throughout the longitudinal continuum of recovery, very few studies have documented discourse changes from the *hyperacute* to the *chronic* phase of recovery.

Aims: The present study aims to document a multilevel analysis of discourse changes from the *hyperacut*e phase to the *chronic* phase of post-stroke recovery using a series of single cases study design.

Methods and Procedures: Four people with mild to moderate post-stroke aphasia underwent four assessments (*hyperacute*: 0-24 hours; *acute*: 24 to 72 hours; *subacute*: 7 to 14 days; *chronic*: 6 to 12 months post-onset). Three discourse tasks were performed at each time point: a picture description, a personal narrative, and a story retelling. Multilevel changes in terms of macrostructural and microstructural aspects were analyzed. Results of each discourse task were combined for each time point. Individual effect sizes were computed to evaluate the relative strength of changes in an *early* and a *late* recovery time frame.

Outcomes and Results: Macrostructural results revealed improvements throughout the recovery continuum in terms of coherence and thematic efficiency. Also, the microstructural results demonstrated linguistic output improvement for 3 out of 4 participants. Namely, lexical diversity and the number of correct information units per minute showed a greater gain in the *early* compared to the *late* recovery phase.

Conclusions: This study highlights the importance of investigating all discourse processing levels as the longitudinal changes in discourse operate differently at each phase of recovery. Overall results support future longitudinal discourse investigation in people with post-stroke aphasia.

Keywords: discourse, multilevel analysis, longitudinal study, aphasia, case series design

Introduction

Discourse is the bedrock of human oral communication, and its performance includes the simplest utterances produced and the most complex combination of ideas. Classically, discourse refers to language beyond the sentence level (Armstrong, 2000). Spoken discourse assessment and analysis are gaining widespread importance in clinical and research practices in all phases of aphasia recovery (Stark, Dutta, Murray, Fromm, et al., 2021). Also, linguistic discourse analysis is advantageous as it assesses the relationships between discourse and language abilities usually assessed by discrete language measures (Prins & Bastiaanse, 2004). However, the lack of standardized practices is a critical challenge for advancing discourse analysis practices (Stark, Dutta, Murray, Bryant, et al., 2021).

Even if natural and treatment-related post-stroke language recovery occurs over time (e.g., Bernhardt et al., 2017), only a few studies have explored discourse changes longitudinally (Brisebois et al., 2020, 2021; Ellis et al., 2005; Stockbridge et al., 2019). None of these studies have assessed multilevel discourse changes with multiple tasks from the hyperacute until the chronic phase of recovery. We propose to contribute to this topic and fill this gap in the literature. Single-case studies in individuals with brain damage, including people with aphasia, have contributed to cognitive models and a superior understanding of cognitive architecture (Medina & Fischer-Baum, 2017). Indeed, many single-case studies in aphasia have led to a more refined insight into treatment effects (Thompson, 2006). Hence, our longitudinal case series study will allow us to generate more specific hypotheses on what should be investigated in a larger and future study. Studying the language properties associated with each discourse processing level in a multilevel analysis is fundamental to adequately represent discourse function (Sherratt, 2007). Such analysis has many advantages and can capture quantitative and functional language information associated with discourse performance. Indeed, it allows for in-depth analysis of underlying discourse processes in non-brain damaged individuals (e.g., Sherratt, 2007) and people with aphasia (e.g., Marini, Andreetta, et al., 2011). The present study aims to document the multilevel facets of discourse throughout the continuum of aphasia recovery, from the *hyperacute* to the *chronic* recovery phase.

The need to measure longitudinal changes in discourse production

Discourse production is typically assessed in the very early period of the stroke care continuum (Stark, Dutta, Murray, Fromm, et al., 2021), but little is known about how discourse changes from stroke onset until the chronic phase of recovery. Current evidence in stroke recovery supports that the larger part of brain repair and behavioral recovery occurs in the first few weeks to months post-stroke for most people (e.g., Pedersen et al., 2004). The Stroke Recovery and Rehabilitation Roundtable task force proposed a framework (Bernhardt et al., 2017) that identified five critical phases of post-stroke recovery that incorporate the latest evidence in biology and knowledge about behavioral recovery. The hyperacute phase occurs between 0 and 24 hours after stroke onset; the acute phase occurs between one and seven days after stroke onset; the early subacute phase occurs between seven days and three months after stroke onset; the late subacute phase occurs between three and six months after stroke onset, and the chronic phase occurs after six months post-stroke onset. We expect post-stroke discourse recovery will follow a similar trend to that of overall post-stroke language recovery. However, to the best of our knowledge, only four studies conducted to date (including two from our group) have focused on discourse abilities across two or more of these phases (Brisebois et al., 2020, 2021; Ellis et al., 2005; Stockbridge et al., 2019).

Among these longitudinal discourse studies, Ellis et al. (2005) documented cohesive abilities in personal narratives at 1-, 6-, and 12-months post-stroke in 12 people who suffered a left hemisphere stroke but who had not been diagnosed with aphasia. The total number of cohesive ties remained stable during the first year following a stroke. Cohesion markers are linguistic units supporting structural and semantic relations between contiguous utterances and include reference, conjunction, and lexical ties (Halliday & Hasan, 1976). However, the proportion of adequate cohesive ties significantly increased during the same period. Thus, these results support the natural recovery over time of subtle language disruptions following a stroke. More recently, changes in the production of cohesive ties in a single-picture description task have been investigated in a group of participants with right (n=76) and left (n=145) hemisphere ischemic strokes at two-time points -- <1 week and 6-12 months post-stroke (Stockbridge et al., 2019). This study demonstrated that deficits in cohesion occur whether the stroke is located in the right or left hemisphere and that cohesion strategies vary across groups. However, the authors reported minimal changes in the production of cohesion markers over time, possibly due to very high inter-individual variability. Also, relatively few participants contributed data at both time points (i.e., less than 25% of the left-hemisphere sample). Our research group has also documented the recovery of discourse production, using a picture description task, in seventeen people with post-stroke aphasia at three-time points: 0 to 72 hours; 7 to 14 days; 6 to 12 months post-onset (Brisebois et al., 2021). We found changes in the macrostructural measure of thematic informativeness over time; however, there were no changes at the microstructural level (terms to be described in detail below). One of the limitations of these studies is the use of only one type of discourse task which limits, as mentioned above, the representativeness of the data collected. There is a clear need to report changes over time from the very early hyperacute phase up to the chronic phase to have a better understanding of how discourse abilities progress over time.

Multilevel discourse analysis

Theoretical framework of discourse production

The theorical framework of this study is based on Frederiksen's model of discourse (Frederiksen & Stemmer, 1993) which has been also used by Sherrat (2007) and Barker, Young and Robinson (2017). This model divides processes into three distinct stages: 1) conceptual preparation (i.e., idea generation and macrostructural processes), 2) linguistic formulation (i.e., microstructural processes which relate to sentence processing) and 3) articulation and monitoring of the verbal message. It includes both receptive and expressive processing. Figure 1 illustrates this discourse processing model, including top-down and bottom-up processes between the conceptual preparation and the linguistic formulation stages. The following paragraph describes the theoretical framework of discourse production in relation to a multilevel discourse analysis, which includes both macro-and micro-structural levels. This framework supports the subsequent choice of variables for the current study, with each level of processing and language aspect being represented by at least one variable. The correspondence between the levels of discourse processing, language aspects and discourse measures are reported in Appendix 3 of Supplementary Material 1.

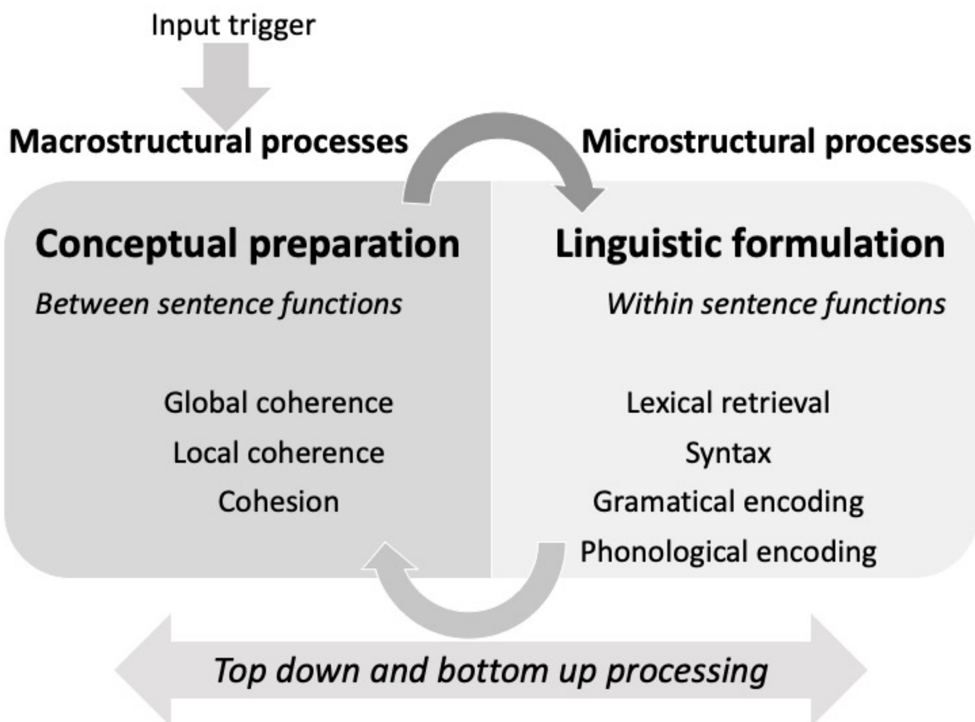

Figure 1 Discourse processing model based on Frederiksen and Stemmer (1993) and Sherrat (2007) and inspired by graphical representation of Barker, Young and Robinson (2017).

All discourse processes follow an input trigger, which is a response to an internal or external stimulus. At the conceptual stage, the speaker generates an abstract idea of the message and retrieves the appropriate conceptual frame (i.e., discourse type) following the input trigger. Thus, semantic information (e.g., context, participants, theme) is integrated into the frame. Also, the prelinguistic message is monitored by the linguistic content (i.e., what has been said by the speakers) and the extralinguistic context (i.e., the prevailing circumstance and the place and time of communication). This monitoring includes pragmatic adjustment, regulating the quantity and relevance of the communication intent in a specific context. In this framework, executive processes have a monitoring role, predominantly at the conceptual level (e.g., Barker et al., 2017). Conceptual information prompts propositional creation through macrostructural coherence and cohesion processes, which are mechanisms that connect sentences, and thematic informativeness. Coherence refers to the ability to maintain overall semantic and pragmatic unity and is represented at the local and the global levels (Kintsch & van Dijk, 1978). Complementarily, cohesion is achieved through cohesion ties, which are linguistic units, and refers to relationships within or between sentences (Halliday & Hasan, 1976). Also, at the macrostructural level and conceptual stage, thematic informativeness is the ability to name themes specific to a stimulus (Brisebois et al., 2020). Related to the conceptual processes, the preverbal message is then converted into a discourse plan at the linguistic formulation level. This level supports lexical association in the mental lexicon including lexical access and selection but also relates to semantic intention at the conceptual level. Therefore, the activation of lexical targets allows access to morphosyntactic (lemma) and phonologic characteristics of words. At the first step of the linguistic formulation level, morphosyntactical information supports thematic role attribution and sentence generation. The second step includes positional organization within and between sentences and appropriate syntactic/grammatical representation. Syllabic and phonological pieces of information are then selected and linked to the output system of speech sound production. The linguistic formulation level is represented in discourse analysis by microstructural features. For instance, syntactical complexity can be represented by the mean length of utterances and lexical diversity by the type-token ratio (Stark, 2019). The current study focuses on discourse production, more precisely on both conceptual preparation and linguistic formulation.

Multilevel discourse evidence in aphasia

Marini and Andreetta et al. (2011) elaborated a multilevel procedure grounded in linguistic and psychological theories to comprehensively study two picture description tasks (single picture and sequential story) in two people with fluent aphasia. Their analysis captured aspects of linguistic processing and provided relevant evidence about the communicative skills and the extent of thematic informativeness of two people with aphasia. Interestingly, their in-depth multilevel discourse analysis identified language disparities compared to 105 controls and demonstrated different discourse patterns for the two participants. The person with fluent aphasia following a traumatic brain injury demonstrated recovery at the lexical and syntactic levels, whereas residual impairments in coherence and cohesion were evident. However, the person with post-stroke non-fluent aphasia improved on lexical informativeness, which was associated with a decrease in global coherence errors (Marini, Andreetta, et al., 2011). The different etiologies of the language and communication deficits of these two individuals might explain these different discourse patterns. Namely, a more diffuse post-traumatic brain injury aphasia might explain the more persistent impairments at the conceptual/macrostructure level, whereas a more focal lesion explains the more specific linguistic deficits in the person with post-stroke aphasia. In this case, lexical improvement may have positively influenced coherence in discourse, showing interactions between the conceptual and the linguistic formulation theorical levels.

Using a similar approach, another study of 15 people with post-stroke aphasia and 15 healthy controls demonstrated that lexical diversity and the percentage of information units contributed to maintaining coherence in a story-telling task supported by wordless picture books, whereas syntactic complexity did not (Wright & Capilouto, 2012). These results support a relationship between the lexical and coherence levels of discourse processing and the importance of using a multilevel approach to better understand the relationships between the various levels of discourse. Similarly, Andreetta and Marini (2015) demonstrated that lexical impairment is associated with lower cohesion in the discourse of 20 people with chronic fluent aphasia in one picture description task and two cartoon story tasks. The authors suggested that lexical impairment impacted sentence production and subsequently directly affected cohesion mechanisms. A cross-methodological investigation of ten people with aphasia and ten healthy controls found that people with aphasia were generally perceived as less coherent than the healthy controls (Linnik et al., 2021). The authors also claimed that both micro-and macro-structural features contributed to the perception of discourse coherence during the story retelling. Overall results support a growing interest in the multilevel approach to describe and better understand discourse processing in people with aphasia. However, a significant drawback of the above-mentioned studies is that conclusions are based on only one or two discourse tasks.

Assessing multiple tasks in discourse analysis

The impact of discourse tasks is a critical factor in discourse analysis. Referring to Figure 1, the input trigger (e.g., direct request of narrative production or a stimulus such as a picture) shapes the conceptual preparation stage of discourse processing (Sherrat, 2007). Collecting multiple discourse tasks (e.g., personal narrative, story retell or picture description) is essential to reveal discourse competencies in post-stroke patients because it allows for a better representation of actual language use (Brookshire & Nicholas, 1994). Still, most studies investigating discourse, including ours (Brisebois et al., 2020, 2021), report only one type of task (Bryant et al., 2016) which may provide an incomplete profile of the individuals' discourse processing skills (Sherratt & Bryan, 2019).

Also, combining tasks and samples to study performance in multiple sessions allows a better grasp of the multifaceted aspects of discourse and supports superior stability of discourse measures. In a study of discourse production in 20 patients with chronic aphasia and 20 healthy controls, better test-retest stability of the number of words per minute was observed and the correct information units (CIUs) as the sample size increased (Brookshire & Nicholas, 1994). Indeed according to (Brookshire & Nicholas, 1994), the best balance between high test-retest stability and data processing efforts is achieved with combined samples of about 300 to 400 words for individuals with aphasia. In addition, combining multiple discourse tasks and samples has been recommended to improve test-retest stability in word-retrieval measures (Boyle, 2014). Various combinations of discourse tasks and samples have been reported to date (e.g., Edmonds & Babb, 2011; Whitworth et al., 2015; Zhang et al., 2020), but no precise combination of discourse tasks has been recommended in post-stroke discourse research to date.

Purpose

The present study aims to conduct a longitudinal multilevel analysis of discourse production throughout the continuum of stroke recovery in four people with post-stroke aphasia. Based on previous longitudinal studies (e.g., Brisebois et al., 2021) and the fact that language recovery lasts up to several months after stroke (Bernhardt et al., 2017; Pedersen et al., 1995), the central hypothesis of this case series study is that there will be positive changes in discourse production at both the macrostructural and microstructural levels throughout the recovery continuum. Based on Pedersen *et al.* (1995), a more specific hypothesis is that greater change will occur in the *early* phase of recovery, from onset to the *subacute* phase, with more subtle improvements in the *later* phase, from the *subacute* to the *chronic* phase.

Method

The study was approved by the ethics review board of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Ile-de Montréal (CIUSSS-NÎM; #2020-1900) and written informed consent was obtained from all participants.

Study design

Similar to Bernhardt et al.'s (2017) stroke recovery framework (Bernhardt et al., 2017)we conducted a longitudinal case series study during which we collected data at four time points. For the present study, the *hyperacute* phase is defined as between 0 and 24 hours after stroke; the *acute* phase as between 24 hours and three days after stroke; the *subacute* phase as between 7 and 10 days after stroke; and the *chronic* phase as at least six months after stroke. We measured multilevel discourse changes from the *hyperacute* to the *chronic* phase of post-stroke aphasia recovery using three discourse tasks: a picture description, a personal narrative, and a story retelling. Figure 2 illustrates the study design.

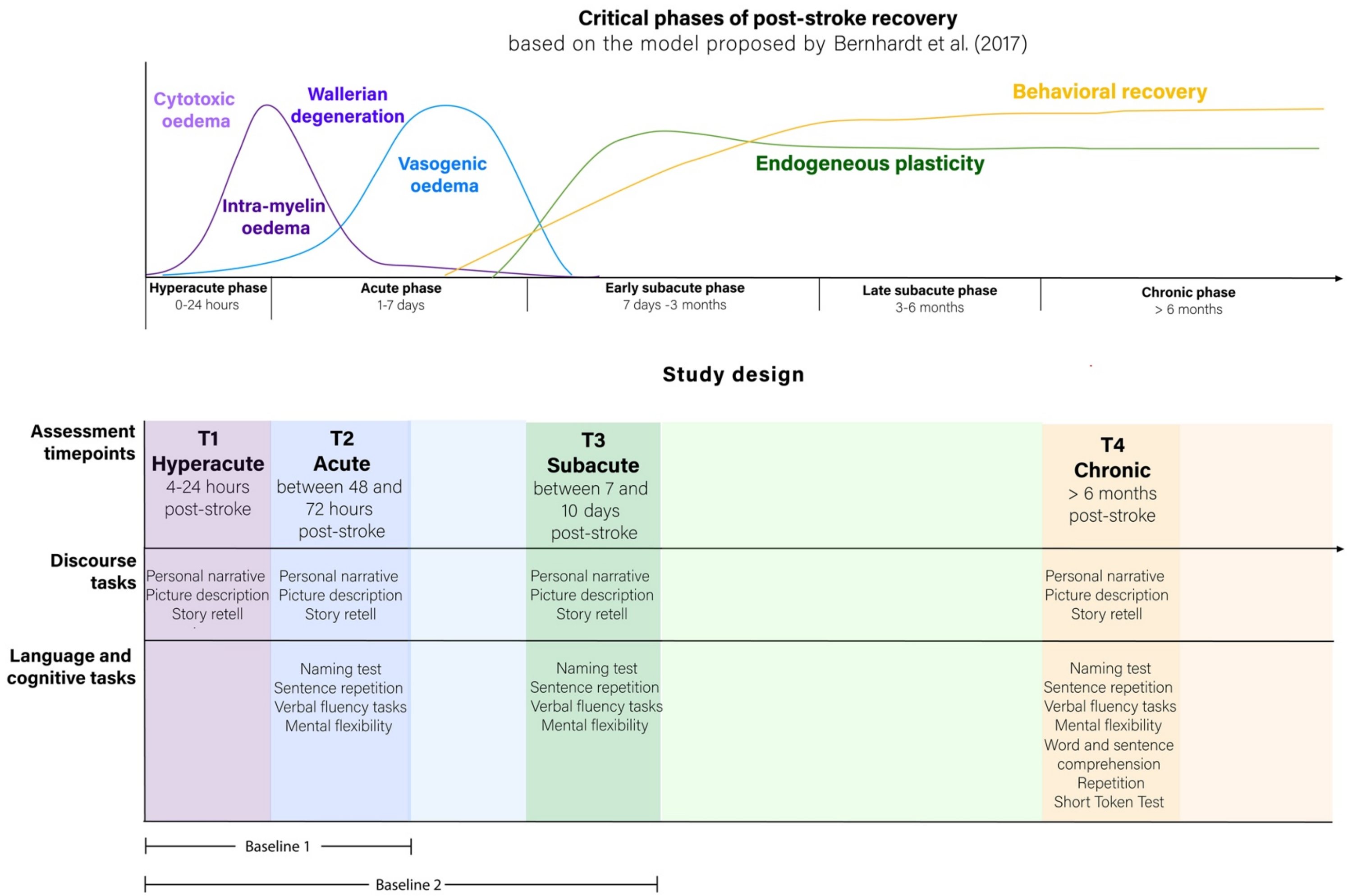

Figure 2 Critical phases of post-stroke recovery and study design

Participants

Four participants were recruited from the stroke unit of a hospital in Montreal, Canada. All participants were male and were native French-Canadian speakers. They were all right-handed and had sustained a very recent (<24 hours) ischemic stroke located in the left middle cerebral artery territory. No criteria concerning initial aphasia severity or lesion size were applied. Exclusion criteria were a history of major psychiatric disorders, learning disabilities, severe perceptual deficits (as identified by the on-call neurologist), left-handedness, and additional neurological diagnoses. Recruitment for this project occurred between May 2020 and August 2021 and is still ongoing for the larger research study. It was interrupted for some periods of time due to the COVID19 pandemic. As of August 2021, 11 participants were recruited and four had completed the four testing sessions. This is a convenience sample since the timing of recruitment was critical to participation and we selected only participants that completed the four testing sessions in August 2021.

Table 1 presents participants' sociodemographic characteristics. Assessments were conducted by a certified speech-language pathologist with more than ten years' experience. Severity scoring and aphasia type were based on the results obtained on the assessment tasks, clinical judgement and overall rating on the BDAE severity scale (Goodglass et al., 2001). Participant 1 is a 74-year-old retired professional with 17 years of education. He has lived independently with his partner before and after the stroke. He is bilingual but has not used his second language (English) frequently since retirement. Initially, he had mild to moderate anomic aphasia; no residual language impairment was evident at the last assessment. Participant 2 was 59 years old at stroke onset, has 11 years of education, and works in a factory. He lived independently with his partner before and after the stroke. He is also bilingual and uses both languages equally on an everyday basis. He initially had mild to moderate anomic aphasia, which evolved to mild anomic aphasia at the last assessment. Participant 3 was 88 years old at stroke onset, had ten years of education, and was retired. He lived with his partner before and after the stroke and is also the primary caregiver for his partner. He is monolingual. Initially, he had moderate Broca's aphasia, which evolved to mild to moderate Broca's aphasia at the last assessment. Participant 4 was 64 years old at stroke onset and had 16 years of education. He is now retired and runs a local business. He lived independently with his partner and their children before and after the stroke. He is bilingual (English is a second language) but mainly speaks French-Canadian on an everyday basis. Initially, he had mild anomic aphasia and no detectable language impairment at the last assessment. Participant 4 is the only one who received thrombolysis intravenous therapy. All participants were Caucasian.

Table 1 Participant characteristics

	Participant 1	Participant 2	Participant 3	Participant 4
Age at stroke onset (years)	74	59	88	64
Sex	Male	Male	Male	Male
Education	17	11	10	16
Data collection time points (days) and modality				
Hyperacute	1 - bedside assessment	1 - bedside assessment	1 - bedside assessment	1 - bedside assessment
Acute	2 - bedside assessment	2 - bedside assessment	2 - bedside assessment	2 - bedside assessment
Subacute	10 - in person / home	9 - virtual	10 - in person / home	9 - virtual
Chronic	249 - in person / home	259 - virtual	206 - in person / home	204 - virtual
Initial aphasia type and severity[1]	Transcortical motor Mild to moderate	Anomic Mild to moderate	Broca Moderate	Anomic Mild
Chronic aphasia type and severity[1]	None	Anomic Mild	Broca Mild to moderate	None

[1]Severity ratings based on the Boston Diagnostic Aphasia Examination Scale (BDAE).

Procedure

Assessments

Participants underwent four assessments longitudinally. Discourse tasks were administrated at every time point, and language and cognitive testing varied across testing points. All the assessments were conducted in French-Canadian. The first assessment (T1; *hyperacute* phase) occurred within the first 48 hours post-onset, the second assessment (T2; *acute* phase) took place between 48- and 72-hours post-stroke, the third assessment (T3; *subacute* phase) took place approximately seven days post-stroke, and the fourth assessment (T4; *chronic* phase) took place about eight months post-stroke. Testing took place in person during hospitalization and after discharge, either virtually using the Zoom platform because of the pandemic or at home. Each participant's specific data collection time points and place of assessment appear in Table 1. By the last assessment, the participants were medically stable. Only participant 2 received speech and language therapy; the rehabilitation center could not provide detailed information about treatment (intensity and duration). The patient reported that it mainly targeted word retrieval and swallowing. In the subacute phase, participants 1 and 3 considered that their language and communication were very similar to their pre-stroke status, and they chose not to receive speech-language therapy even though it was recommended by the Speech-Language Pathologist at hospital discharge (as per the Canadian Stroke Guidelines (Boulanger et al., 2018)). Participant 4 did not have any detectable language impairment at hospital discharge; thus, speech-language therapy was not recommended.

Discourse tasks

Three discourse monologic tasks were chosen for this study. Each assessment included the three spoken discourse tasks and was administered at every time point in random order, and language and cognitive testing varied across testing points (see Figure 2). Oral discourse tasks were: (1) the single picture description of the Picnic scene of the Western Aphasia Battery –Revised (WAB-R; Kertesz, 2006); (2) the '*Discours conversationnel*' [Conversational discourse] task of the '*Protocole Montréal Évaluation de la Communication*' which we will call 'personal narrative' because it involves a telling of a personal story with minimal to no interaction between speakers (MEC; Joanette et al., 2004); and (3) the story retell task of Cinderella (Greenslade et al., 2020). Instruction for the WAB-R oral picture description was: '*Racontez-moi ce qui se passe sur cette image*' [Tell me what is happening in the picture]. The personal narrative task was performed according to the instructions of the '*Discours conversationnel*' task from the MEC (Joanette et al., 2004), similar to the first part of the 'Connected speech' item of the Quick Aphasia Battery (Wilson et al., 2018). The instruction was: '*Parlez-moi de quelque chose à propos de vous, comme votre travail, votre famille ou vos loisirs, etc.*' [Speak to me about a personal event or interest (e.g., work, vacation, family)]. Participants were free to choose the theme they wanted. Appendix 1 lists the themes chosen by each participant at each time point for the personal narrative task. The examiner was instructed not to interrupt the participant and encourage elaboration if needed. The four participants did not need any prompts to complete the task in all assessments. The story retell task of Cinderella was administrated similarly to the AphasiaBank protocol (e.g., MacWhinney et al., 2011). Participants were shown wordless images in the Cinderella book and asked to remember the story as they went on. Then, the book was removed, and participants were asked to retell the story. The instruction was: '*Racontez-moi l'histoire de Cendrillon du mieux que vous pouvez*' [Tell me the Cinderella story as well as you can].

Language and cognitive tasks

Assessment duration was kept to a minimum at the *hyperacute* phase (T1) so that all patients could perform the entire protocol for this session. During the *hyperacute* and *acute* phases, the assessment is limited in time as many examinations and medical consultations are conducted in individuals with a stroke in order to identify the cause of the stroke and to get them medically stable. Additional language and cognitive tasks were performed at T2 and T3. The naming task '*Test de dénomination de Québec - 30*' (TDQ-30; Macoir et al., 2021), the sentence repetition test '*TEst Français de RÉPétition de Phrases*' (TEFREP; Bourgeois-Marcotte et al., 2015), the three verbal fluency tasks (i.e., semantic, orthographic and free) of the '*Protocole Montréal Évaluation de la Communication*' (MEC; Joanette et al., 2004), and the mental reactive flexibility test Alphaflex (Grotz et al., 2018) were administered at T2, T3 and T4. In the chronic phase (T4), when all participants can tolerate longer assessments, the Word and sentence comprehension task and the Word and sentence repetition task of the Montreal-Toulouse Aphasia Examination (MT-86; Nespoulous et al., 1992) as well as the picture-picture association task of the Pyramid and Palm Trees Test (Howard & Patterson, 1992) were also conducted. Detailed language results are provided in Appendix 2.

Data collection and sample analysis

Samples were recorded using an H1n Zoom Handy Recorder for interviews in person and using Zoom for virtual assessments. Audio files of each discourse sample were imported and transcribed in the EUDICO Language Annotator (ELAN; Sloetjes & Wittenburg, 2008). Recordings were fully transcribed orthographically and included all verbal behaviors such as fillers. Two experienced speech-language pathologists (A.B. and a research assistant) transcribed the samples using Code for the Human Analysis of Transcripts (CHAT) conventions (MacWhinney, 2000) with additional guidelines for French users (Colin & Le Meur, 2016). Utterance segmentation was performed using CHAT conventions (MacWhinney, 2000) and was based on a combination of phonological, syntactic, and semantic criteria (see also Marini, Andreetta, et al., 2011).

Macrostructural analysis

Coherence refers to the ability to maintain overall semantic and pragmatic unity and is represented at the local and the global levels (Kintsch & van Dijk, 1978). Coherence was scored according to Marini and Andreetta et al. (2011) along two dimensions: (1) errors in local coherence or the semantic relatedness between contiguous utterances; (2) errors in global coherence or the relatedness of remote utterances with the overall theme or topic. Appendix 3 lists the discourse outcome measures of the present study in relation to the discourse processing levels.

Appendix 3 provides examples of local and global coherence scoring. Local coherence errors include topic switches and missing referents (as per Marini, Galetto, et al., 2011). Global coherence errors include utterances that are tangential, conceptually incongruent with the story, propositional repetitions, or filler sentences (Marini, Andreetta, et al., 2011).Global coherence scoring was: (1) tangential when it contained a derailment in the flow of discourse concerning the information already provided in a preceding utterance; (2) conceptually incongruent when it included ideas not directly addressed by the task; (3) a propositional repetition where the speaker repeated ideas, showing a lack of novelty, or directly restated utterances, reflecting perseveration; and (4) a filler utterance when it was an empty phrase that did not provide any additional information to the overall task or was a direct comment about the nature of the task. Two coherence outcome measures were then computed: local coherence errors (LCE) and global coherence errors (GCE) (a composite measure: propositional repetitions + tangential utterances + conceptually incongruent utterance + filler sentences). Both LCE and GCE are expressed as a percentage of the total number of utterances. Coherence scoring and examples are provided in Appendix 4.

Cohesion refers to the structural and semantic relations among contiguous utterances (Halliday & Hasan, 1976). Cohesion was scored similarly to Barker et al. (2017) using the classification elaborated by Halliday and Hasan (1976) including three types of cohesive ties: reference, conjunction, and lexical. The total number of cohesive ties for each category was computed, and each was judged according to its adequacy (adapted from Liles, 1985). Briefly, adequacy was scored as either 'complete' or 'incomplete.' A cohesion marker was scored as complete if the information referred to it was quickly found and unambiguously defined. A cohesion marker was scored incomplete when the information referred to it was not provided or if the listener was guided to ambiguous information elsewhere. Three cohesion outcome measures were selected from Barker et al. (2017): total complete cohesion (TCC; a composite score consisting of the sum of complete reference and complete conjunction), total complete lexical ties (TCLT), and total cohesive errors (TCohE) (a composite score consisting of the sum of reference errors and conjunction errors). Each measure of cohesion was expressed as a percentage of the total number of words uttered. Cohesion scoring and examples are provided in Appendix 4.

Thematic Units (TUs) consisted of a finite list of items specific to a stimulus and were developed for the picture description task of the WAB (Brisebois et al., 2020). More specifically, we identified 16 TUs produced by at least 75% of a group of 45 healthy French-Canadian speakers who completed the picture description of the WAB. A similar selection procedure has also been employed in previous studies (e.g., (Marini, Galetto, et al., 2011). TUs were scored only for the picture description task for the present study. This measure refers to the number of specific units the participants produced, where a maximum of 16 TUs could be obtained. Following the same rules, the 16 TUs were included in an analysis grid and given a score of one point. Macrostructural efficiency was also measured by calculating the number of TUs per minute.

Microstructural analysis

The initial selection of microstructural variables was made based on Stark (2019). These are: the mean length of utterance (MLU), the propositional density, the number of words per minute, the number of verbs per utterances, the type-token ratio, the open-closed class ratio, the noun to verb ratio and the number of tokens. Correct Information Units (CIUs; Nicholas & Brookshire, 1993) were also included to measure lexical informativeness in collected samples. Moving Average Token-Type Ratio (MATTR; Covington, 2007) was also calculated to compare lexical diversity across samples of varying lengths.

Transcriptions underwent detailed linguistic and textual analysis focusing on micro-structural measures known to be affected in aphasia (Stark, 2019). All micro-structural variables were extracted using the program EVAL of the Computerized Language ANalysis (CLAN) program (MacWhinney, 2000). Each variable was computed for all tasks in the *hyperacute*, *acute*, *subacute*, and *chronic* phases. Extracted productivity measures were defined as the total number of completed words and the number of words per minute. Syntactical complexity was measured by the MLU, the number of verbs/utterances, open/closed ratio, and noun-verb ratio. Propositional density measured content richness (Brown et al., 2008), and MATTR measured lexical diversity. The CIUs that represent a lexical informativeness score was also automatically extracted with the CLAN software after all non-CIU words were excluded from the transcripts as suggested by the

AphasiaBank procedure[1]. Total CIU and the rate of CIUs per minute were included in this study.

Macro- and micro-structural variables were all computed for all samples except for the TUs, which were only scored for the picture description task. All variables were collected for the four time points. Table 2 summarizes the commands used to extract macro- and micro-structural variables.

Table 2 CLAN commands used to extract discourse variables in the transcripts

CLAN Commands	Results
mor	• Tag parts of speech automatically using mor script
eval +t*PAR: +u	Evaluate transcripts to derive primary linguistic outcome variables • eval: evaluate microlinguistic information using the mor tier • +t*PAR: evaluate only the participant tier • +u: consolidate all files to single output
freq +t*par +d2	Evaluate the occurrence of each word on the participant tier • freq: extract all the words used in the file. along with their frequency counts. and calculate all the types and tokens • +t*par: evaluate only the participant tier
freq +s\[\] +t*PAR +d2**	Evaluate the occurrence of cohesion and coherence scores • freq: get a frequency count • +s\[**\]: extract all coding that have been marked with '*' (e.g.. local cohesion errors are marked [* LoC]) • +t*PAR: evaluate only the participant tier • +d2: consolidate all files to single output
freq +tPAR +b10 +d3	Evaluate the occurrence of each word on the participant tier • freq: get a frequency count • +b10: calculate the lexical diversity using the Moving Average Type-Token Ratio (MATTR). This index is based on a moving window that computes TTRs for each successive window of fixed length (i.e.. 10 words). • +d3: consolidate all files to single output

[1] https://aphasia.talkbank.org/discourse/ consulted on December 16th, 2021.

Each variable included results for each task to obtain a global result for each variable at each time point. More specifically, raw macrostructural scores were summed and computed. As for the microstructural variables, scores were computed for all tasks combined at each time point. MATTR and density scores were calculated as means for all tasks at each time point.

Visual and statistical analysis

We compared time frames in terms of *early* and *late* recovery for each participant. To document the *early* phase, we constructed Baseline 1, which represents the mean results for a given individual at the *hyperacute* and *acute* phases compared to the results at the *subacute* phase. To assess changes in the *late* phase, we constructed Baseline 2, which represents the mean results for a given individual at the *hyperacute*, *acute*, and *subacute* phases compared to the results at the *chronic* phase.

Similar to the visual analyses performed by Lee & Cherney (2018), data were plotted, and graphs were analyzed visually. To document the *early* phase, visual analysis was performed for each variable and participant between Baseline 1 (i.e., the mean results for one participant of the *hyperacute* and the *acute* phases) and the result in the *subacute* phase.

$$\text{Baseline1} = \bar{x}\ (hyperacute\ and\ acute)$$

In addition, the *late* phase was documented by comparing Baseline 2 (i.e., the mean results obtained in the *hyperacute*, the *acute,* and the *subacute* phases) and the result in the *chronic* phase for each participant.

$$\text{Baseline 2} = \bar{x}\ (hyperacute, acute\ and\ subacute)$$

Then, to quantify the relative weight of changes, effect sizes were calculated for all variables using the *d*-statistic (Beeson & Robey, 2006). Individual effect sizes allow for the comparison of the relative strength of the effect of time effect within and between individuals. The *d*-statistic has been used in single case studies to detect treatment effects, to determine the relative potency of results and synthesize meaningful conclusions (Beeson & Robey, 2006). Although we did not conduct a treatment study, we calculated effect sizes nonetheless, since data were collected at multiple time points in our case series study.

First, to assess effect size (d) in the *early* recovery phase, the result in Baseline1 was subtracted from the *subacute* result and divided by the standard deviation of Baseline1.

$$d_{early} = \frac{(subacute - Baseline1)}{S_{Baseline1}}$$

Then, to evaluate the effect size in the *late* recovery phase, the result in Baseline2 was subtracted from the *chronic* result and divided by the standard deviation of Baseline2.

$$d_{late} = \frac{(chronic - Baseline2)}{S_{Baseline2}}$$

Reliability

Inter-rater reliability for macrostructural variables of coherence, cohesion, and thematic units were independently manually scored by two authors (A.B. and K.M.) for 25% of the discourse samples. For microstructural variables, a trained speech-language pathologist independently conducted a second transcription on 8% of all transcripts, with A.B. as the primary transcriber. Two-way random effects intra-class correlation coefficients (ICC) were calculated using SPSS® v27.0 on all variables to determine consistency between raters (as in Marcotte et al., 2017). The results of these psychometric analyses are reported in Table 3.

Table 3 Inter-rater reliability for all variables (two-way random effects intraclass correlation)

Variables	Cronbach's alpha *(α)*
Global coherence errors	.742
Local coherence errors	.388
Total complete cohesion ties	.827
Total complete lexical ties	.661
Total cohesion errors	.175
Thematic Units	.957
Total number of words	1.0
Total number of utterances	.997
Number of words per minute	.991
Mean length of utterance (MLU)	.994
Noun verb ratio	.999
Number of verbs per utterance	.995
Open to close word category ratio	.980
Density	1.0
Moving Average Type Token ratio	.971
Correct Information Units	.999

All microstructural variables and the TUs met the threshold of excellent reliability, ICC > .90 (Boyle, 2020; Koo & Li, 2016). The complete cohesion ties obtained good reliability, i.e., ICC between 0.75 and 0.90. The global coherence errors and the complete lexical ties obtained moderate reliability, i.e., between 0.5 and 0.75. The local coherence errors and the total cohesion errors obtained poor reliability. Consequently, these two variables were not considered in the analyses.

Supplementary Material 2 includes the Best Practice Guidelines checklist from Stark *et al.* (2022).

Results

The results reported below include only the effect sizes that were significant. The complete visual and quantitative analyses are reported in Tables 4 and 5 for the macrostructural and microstructural variables, respectively.

Table 4 Analyses of macrostructural discourse variables

		Early recovery		**Late recovery**	
		Visual analysis	Effect size	Visual analysis	Effect size
Variables	**Participants**	**Baseline1 < subacute**	**d_{early}**	**Baseline2 < chronic**	**d_{late}**
Global coherence errors (%)	P1	No	-1.58	No	-2.08
	P2	Yes	**4.07**	Yes	**6.46**
	P3	No	-1.59	No	-0.30
	P4	No	**-10.30**	No	-0.94
Total complete cohesion ties (%)	P1	No	-0.53	No	-0.45
	P2	No	-0.38	No	-0.15
	P3	No	-0.40	Yes	0.17
	P4	No	**-3.19**	No	-1.39
Total complete lexical ties (%)	P1	Yes	0.39	Yes	**7.35**
	P2	Yes	**5.13**	Yes	**4.47**
	P3	No	-	Yes	-
	P4	No	**-2.79**	No	-0.75
Thematic Units	P1	Yes	0.42	Yes	1.13
	P2	No	-	No	**-2.89**
	P3	Yes	0.71	Yes	2.31
	P4	Yes	0.71	Yes	2.31
Total Thematic Units per minute	P1	Yes	**4.02**	No	-0.11
	P2	Yes	0.37	Yes	0.84
	P3	Yes	**3.89**	Yes	**5.37**
	P4	No	-0.45	Yes	**4.63**

Note: Significant effect sizes are represented in **bold font**.

Table 5 Analyses of microstructural discourse variables

		Early recovery		Late recovery	
		Visual analysis	Effect size	Visual analysis	Effect size
Variables	**Participants**	**Baseline1 < subacute**	**d_{early}**	**Baseline2 < chronic**	**d_{late}**
Total words (FREQ_tokens)	P1	Yes	**3.67**	Yes	**4.62**
	P2	No	-0.01	No	-0.07
	P3	Yes	0.31	Yes	**2.96**
	P4	Yes	0.94	Yes	1.06
Number of words per minute	P1	Yes	**3.20**	No	2.21
	P2	Yes	0.15	Yes	-0.24
	P3	Yes	0.88	Yes	**4.39**
	P4	Yes	0.88	Yes	1.57
Mean length of utterances (words)	P1	Yes	2.15	Yes	1.14
	P2	Yes	**4.04**	Yes	1.42
	P3	Yes	**1147.34**	Yes	0.35
	P4	Yes	2.55	Yes	-0.19
Noun/verb ratio	P1	No	-1.13	No	-0.15
	P2	No	-0.34	Yes	0.86
	P3	No	-0.80	No	-0.33
	P4	No	-1.69	No	-0.89
Number of verbs per utterance	P1	Yes	2.56	Yes	0.77
	P2	Yes	203.84	Yes	0.85
	P3	Yes	**8.56**	No	-0.42
	P4	Yes	2.47	Yes	0.66
Open/closed ratio	P1	No	**-15.85**	No	-2.03
	P2	Yes	0.12	No	-0.22
	P3	No	-1.00	Yes	2.22
	P4	Yes	0.67	No	-0.45
Density	P1	Yes	2.13	Yes	1.76
	P2	Yes	1.11	Yes	0.93
	P3	Yes	2.48	No	-0.21
	P4	Yes	0.47	No	-0.27
Moving Average Type Token ratio (MATTR)	P1	Yes	2.55	Yes	0.79
	P2	Yes	**6.48**	Yes	1.08
	P3	Yes	**3.07**	Yes	0.47
	P4	Yes	**3.05**	Yes	1.15

		Early recovery		Late recovery	
		Visual analysis	Effect size	Visual analysis	Effect size
Variables	**Participants**	**Baseline1 < subacute**	$\mathbf{d_{early}}$	**Baseline2 < chronic**	$\mathbf{d_{late}}$
Correct Information Units (CIUs)	P1	Yes	**3.99**	Yes	**4.80**
	P2	Yes	0.33	Yes	0.28
	P3	Yes	0.41	Yes	**3.38**
	P4	Yes	0.91	Yes	1.25
Correct Information Units (CIUs) per minute	P1	Yes	**3.45**	Yes	2.27
	P2	Yes	**4.38**	Yes	0.64
	P3	Yes	1.44	Yes	**5.31**
	P4	Yes	0.85	Yes	1.77

Note: Significant effect sizes are represented in **bold font**.

Macrostructural variables

For the macrostructural variables, the dependent variables are the number of global coherence errors, the total number of complete cohesion ties, the total number of complete lexical ties, the total number of Thematic Units (TUs), and the number of Thematic Units per minute (TUs/minute).

The percentage of global coherence errors significantly decreased over time for P4, who demonstrated fewer errors per utterance with a large effect size of -10.30 (M=11.75, SD= .90) in the *early* phase. However, P2 demonstrated more global coherence errors throughout time, with a medium effect size of 4.07 (M=3.16, SD= .80) during the *early* phase and a large effect size of 6.46 (M=4.24, SD=1.96) during the *late* phase.

Regarding the percentage of complete cohesion ties, only P4 demonstrated a significant decrease during the *early* time frame with a small effect size for P4 of -3.19 (M=20.86, SD= 1.05). The percentage of complete lexical ties showed a significant increase during the *early* phase for P2, with a medium effect size of 5.13 (M= .31, SD= .44). This variable also demonstrated an increase during the *late* phase for P1 and P2, with a large effect size of 7.35 (M= 2.28, SD= 0.89) for P1 and a medium effect size of 4.47 (M= 1.06, SD= 1.34) for P2. P4 significantly produced fewer lexical ties with a small effect size of -2.79 (M= 3.97, SD= 1.12) in the *early* phase. The results for P3 for this variable were not considered since this participant produced too few lexical ties in most of the samples.

As for the total number of TUs (only for the picture description task), P2 experienced a decrease throughout the time continuum with a small effect size of -2.89 (M= 13.67, SD= 0.58) during the *late* phase. The number of TUs/minute increased during the *early* time frame for P1, P2, and P3, and it decreased for P4. P1 demonstrated a medium effect size of 4.02 (M= 11.68, SD= 1.54) and P3 a small effect size of 3.89 (M= 6.59, SD= .54). An increase of this variable occurred during the *late* phase of recovery for P2, P3, and P4, with P3 and P4 both demonstrating medium effect sizes of 5.37 (M= 7.30, SD= 1.28) and 4.63 (M= 12.45, SD= 1.14), respectively.

Microstructural variables

The microstructural variables reported are the number of total words, the number of words per minute, the mean length of utterances (MLU), the noun-verb ratio, the number of verbs per utterance, the open/closed ratio, the density, the moving average token-type ratio (MATTR), the number of correct information units (CIUs) and the number of correct information units per minute (CIUs/minute).

The total number of words significantly increased for P1 in the *early* time frame with a small effect size of 3.67 (M=595.50, SD= 75.66). As for the *late* phase, this variable also increased for P1 and P3. P1 showed a medium effect size of 4.62 (M= 688.0, SD= 168.91) and P3 a small effect size of 2.96 (M= 211.67, SD= 70.73).

Efficiency, in terms of the number of words per minute, significantly increased for P1 during the *early* time frame with a small effect size of 3.20 (M=102.22, SD= 12.78). This variable also significantly increased for P3 in the *late* phase with a medium effect size of 4.39 (M= 59.99, SD= 7.97).

The MLU in words demonstrated a significant increase in P2 and P3 during the *early* time frame with a medium effect size of 4.04 (M= 5.22, SD= .74) for P2. P3's results demonstrated a very large effect size (i.e., 1147.34) due to very low variability during the *early* phase (M= 4.16, SD= .00).

The number of verbs per utterance significantly increased for P1 and P3 during the *early* time frame, with a small effect size of 2.56 (M= .22, SD= .14) for P1 and a large effect size of 8.56 (M= .17, SD= .02) for P3. During the *late* time frame, low variability accounts for a very large effect size of 203.84 (M= .26, SD= .00) for P2.

The open/closed ratio increased significantly for P1 during the *early* time frame, with a large effect size of -15.85 (M= 1.27, SD= .01).

The MATTR significantly increased for three participants during the *early* time frame with a large effect size of 6.48 (M= .93, SD= .00) for P2, and small effect sizes of 3.07 (M= .74, SD= .06) for P3 and of 3.05 (M= .91, SD= .01) for P4.

The total number of CIUs significantly increased during the *early* phase for P1, with a medium effect size of 3.99 (M= 575.0, SD = 62.23). The same trend was observed during the *late* time frame for two participants, with a medium effect size of 4.80 (M= 657.67, SD= 149.79) for P1 and a small effect size of 3.38 (M= 207.33, SD= 63.88) for P3. The number of CIUs/minute also significantly increased for two participants during the *early* time frame -- P1 demonstrated a small effect size of 3.45 (M= 98.70, SD= 10.48) and P2 a medium effect size of 4.38 (M= 109.72, SD= 3.09). During the *late* phase, P3 obtained a medium effect size of 5.31 (M= 59.05, SD= 6.85).

Figures 3 and 4 visually represent the longitudinal time changes of macro- (Figure 3) and micro- (Figure 4) structural variables. Given the large number of variables, as seen in Tables 4 and 5, we selected a sample of variables that obtained the greater number of significant effect sizes for illustration in the figures. In these figures, discourse variables are represented in linear timeline plots to illustrate time changes for each participant. The four time points are represented on the *x-axis* by the log (number of days post-stroke), and the discourse variables are represented on the *y*-axis. A unique line represents each participant's results. These visual representations were used for the visual analysis.

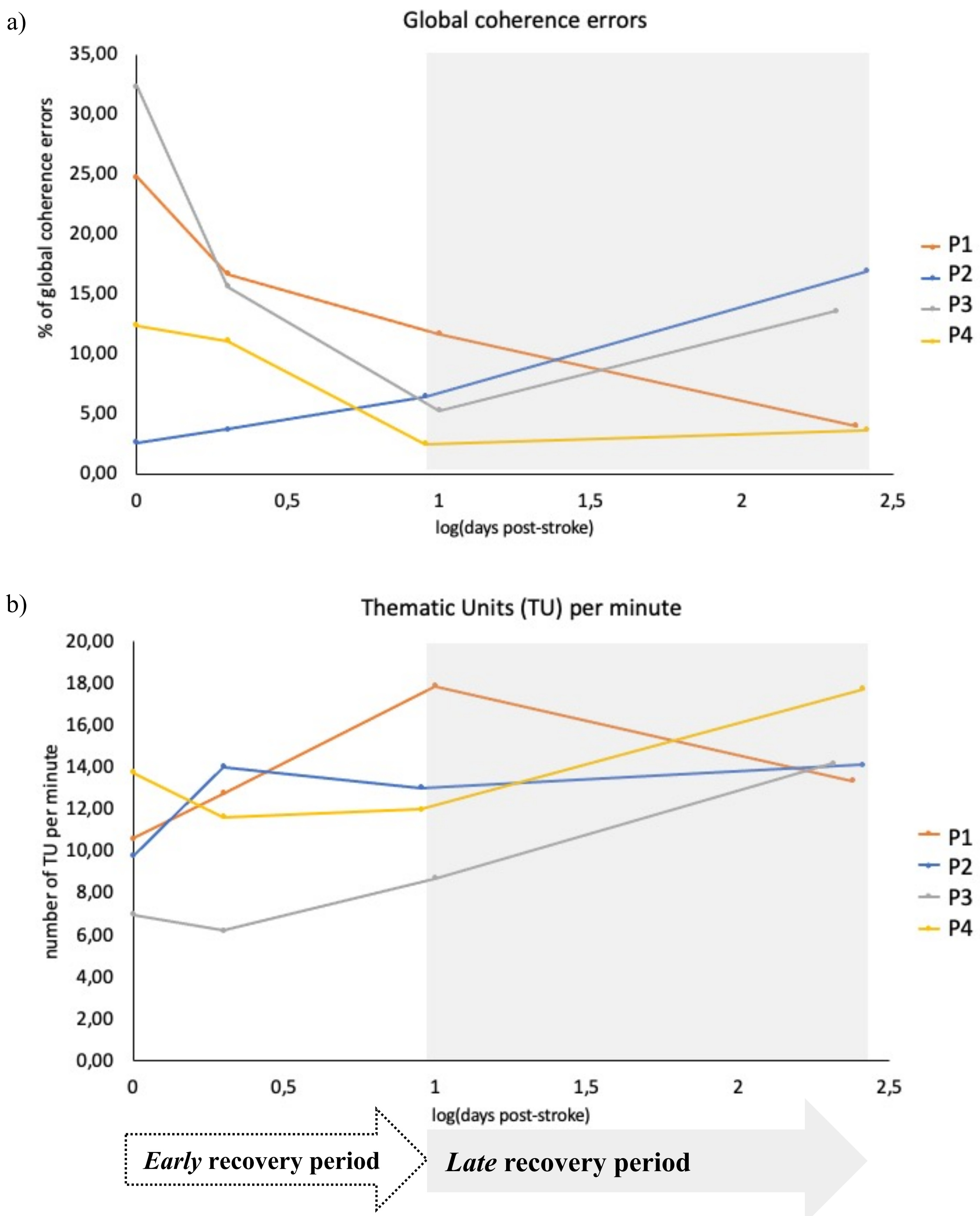

Figure 3 Macrostructural variables of (a) Global coherence and (b) Thematic Units per minute for each participant and each time point

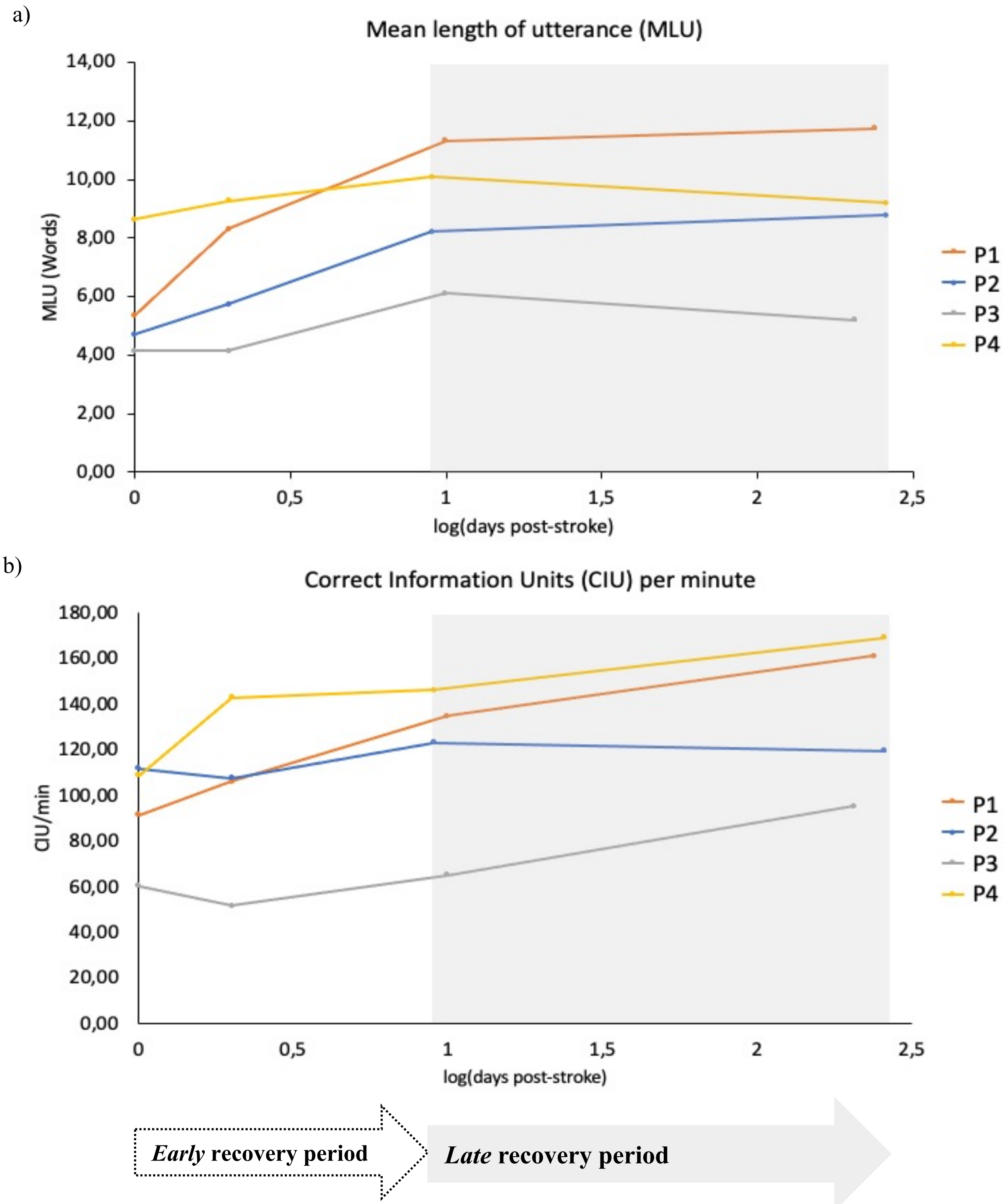

Figure 4 Microstructural variables of (a) Mean length of utterances and (b) Correct Information Units per minute for each participant and each time point

Discussion

This case series study aimed to document the multilevel facets of discourse throughout the continuum of aphasia recovery, from the *hyperacute* to the *chronic* phase in four people with a left hemisphere stroke. Overall, all participants showed discourse improvements during the *early* and/or *late* recovery phase. If we compare the *early* to the *late* phase, two microstructural variables (MATTR and CIUs/minute) demonstrated a greater number of significant effect sizes during the *early* compared to the *late* recovery phase.

Concerning the first aim, consistent with previous evidence (Brisebois *et al.*, 2021) and current knowledge of long-term post-stroke language recovery (Pedersen *et al.*, 1995; Bernhardt *et al.*, 2017), significant improvements in discourse production at both the macrostructural and microstructural levels were observed throughout the recovery continuum. At the macrostructural level, each participant improved significantly on one to four (out of five) macrostructural variables during at least one phase. All macrostructural variables included in the analyses yielded at least one significant effect size, but the percentage of lexical ties and the TUs/minute stood out with a greater number of significant effect sizes across participants. Three people with aphasia demonstrated a significant effect size in either the *early* or *late* recovery phase. Interestingly, the number of TUs/minute demonstrated more significant effect sizes ranging from small to large for two participants in each time frame. This variable represents the ability to express timely relevant information units during the picture description task (Brisebois *et al.*, 2020), and was collected only for one task. TUs and TUs/minute are inherently linked to a fixed stimulus, making it impossible to compute in unstructured tasks such as the personal narrative. Still, macrostructural informativeness efficiency has already been of particular interest in acute stroke, demonstrating diagnostic sensitivity in left and right brain-damaged patients (Agis *et al.*, 2016). The number of TUs/minute has been shown to improve significantly in a group of people with post-stroke aphasia, both in *acute* (Brisebois *et al.*, 2020) and *chronic* (Brisebois *et al.*, 2021) phases of recovery. Also, TUs are very easy and quick to score in the early phases of aphasia, which increases their applicability in clinical settings. The present results support the possibility of using this variable as a potential outcome measure of post-stroke aphasia, at least in the standard-of-care continuum of post-stroke recovery.

Lexical cohesion markers have also been of particular interest in the literature when studying discourse in stroke patients (e.g., Stockbridge *et al.*, 2019). Among the cohesion variables explored in the present study, lexical markers seem to be the most promising, with two persons with aphasia having significantly improved their use of lexical ties over time (Ellis *et al.*, 2005). In contrast, a group difference in terms of lexical ties between the acute (<7 days post-stroke) and chronic phases of recovery was not significant in Stockbridge et al. (2019). One potential reason that may account for the long-term changes observed by Ellis *et al.* (2005) and our study relates to the number and the nature of discourse tasks used. Namely, Stockbridge et al. (2019) extracted their measure from the sole single picture description of the Cookie theft from the Boston Diagnosis Aphasia Examination (Goodglass *et al.*, 2000). In contrast, Ellis et al. (2005) studied personal retell, and our study included a combination of three discourse samples (i.e., single picture description, personal narrative, and story retell). A recent study (Zhang *et al.*, 2020) reported on the multidimensional aspect of cohesion in post-stroke aphasia discourse performance using a combination of multiple discourse tasks (i.e., nine tasks from AphasiaBank). Results highlighted the complexity of cohesive use in people with chronic aphasia. Indeed, different profiles of lexical cohesion in non-fluent and fluent aphasia speakers compared to controls have been observed. More precisely, non-fluent speakers have demonstrated less diversity of lexical ties but were not significantly different from controls, whereas fluent aphasia speakers produced significantly fewer lexical ties than controls (Zhang *et al.*, 2020). In our study, when examining the raw data, lexical ties were less frequently used than reference and conjunction ties for all participants, which is consistent with Stockbridge et al.'s (2019) report. In sum, findings endorse the ongoing investigation of cohesion ties and macrostructural informativeness efficiency in post-stroke aphasia recovery since these variables have been shown to change over time in more people with aphasia in both time frames.

Microstructural results also support a general gain in linguistic discourse performance over time. Interestingly, the MLU, the MATTR, the number of CIUs, and the number of CIUs per minute improved for all participants in both time frames. Specifically, the MATTR, the CIUs, and the CIUs per minute yielded the greatest number of significant effect sizes. These measures all demonstrated three significant effect sizes in both phases, which shows improvement in lexical informativeness and efficiency, as well as lexical diversity. Indeed, the study of many discourse tasks, including a personal narrative and a story retell, might have created more opportunities for lexical diversity, measured by MATTR, hence more possibilities for improvement. Stark (2019) has demonstrated that the type-token ratio, a variable of lexical diversity, improves in the context of a Story Retell task. However, Stark's study did not include a personal narrative task, which we hypothesize would be a better context for lexical output than a structured task such as story retell. Leaman and Edmonds (2021) also recently identified language production variations between tasks, strengthening the importance of the inclusion of several relevant discourse tasks to assess aphasia comprehensively. Overall, multilevel analysis of discourse changes pointed out a different evolution of variables at each discourse level, which supports the inherent complexity of studying changes in discourse production longitudinally.

Regarding the secondary aim of the study to assess changes in the *early* vs. the *late* phase of recovery, our case series study did highlight *early* microstructural changes, whereas no clear trend emerged from the macrostructural analysis. Notably, all participants improved on more microstructural variables in the *early* than the *late* phase, and a greater number of significant small to large effect sizes were obtained in the *early* phase. In addition, most microstructural variables changed significantly on an individual basis, consistent with previous reports of interindividual variability in microstructural measures (Cameron, Wambaugh and Mauszycki, 2010; Boyle, 2015). If we look at the specific variables, lexical diversity (i.e., MATTR) and microstructural informativeness efficiency (CIUs/minute) demonstrated the greatest number of significant effect sizes, specifically during the *early* recovery phase. These findings are of special interest since our previous studies found no early microstructural changes in lexical diversity using a group analysis (Brisebois *et al.*, 2020, 2021) and they support the contribution of individual observations in the analysis of discourse in people with aphasia. A recent study on subacute aphasia recovery demonstrated that *subacute* language recovery is multidimensional in nature and is hardly represented by a unique factor (Stefaniak *et al.*, 2022). Indeed, individual observations of heterogenous recovery trends are concealed when combined in a group analysis. Individual analyses, such as case series studies, allow a more comprehensive observation of post-stroke aphasia recovery. Current knowledge in discourse suggests that micro- and macro-structural processes are intertwined (Marini, Andreetta *et al.*, 2011; Wright and Capilouto, 2012; Andreetta and Marini, 2015; Linnik *et al.*, 2021) but our results showed different trends during the recovery phases. Including an earlier time point in the *hyperacute* phase might also explain our observations. To our knowledge, no discourse study in aphasia has systematically included such an early time point in their experimental design, which could be explained by the challenge of early recruitment of stroke patients. Overall, this study and many others (e.g., Hillis and Heidler, 2002; Saur *et al.*, 2006) support the idea of including a larger number of persons and varied aphasia severities and profiles in the early phases post-stroke to better understand longitudinal discourse recovery processes.

Perhaps surprisingly, some variables demonstrated statistically significant deterioration for some, but no participant demonstrated significant changes suggestive of only deterioration. For instance, P2 demonstrated a global coherence deterioration combined with an improvement of lexical cohesion ties in both phases. This result contradicts previous results where cohesion ties have been considered essential to maintain coherence in discourse (Zhang *et al.*, 2020). A similar but opposite trade-off of lexical cohesion and global coherence changes was observed in P4, whose global coherence improved significantly while the percentage of complete cohesion ties and lexical ties decreased significantly during the early phase. Interestingly, both P2 and P4 improved on at least two microstructural variables even if some deterioration was noticed for one macrostructural variable. It is still unclear whether these changes led to a positive or negative functional outcome for these two participants with anomic aphasia given the current results. However, a combination of positive and negative changes could be interpreted as an attempt for speakers to balance discourse output in the presence of linguistic transformation following aphasia. Improvements of cohesion ties have been documented in case studies (e.g., Coelho *et al.*, 1994; Coelho & Flewellyn, 2003), but a more global measure of cohesive harmony showed stability in three aphasic speakers during a twelve-month period (Armstrong, 1997). More recently, subtle changes have been documented in cohesion recovery in a group of individuals with left hemisphere stroke (Stockbridge *et al.*, 2019). Altogether, the present results highlight the importance of investigating the relationship between the various macro- and microstructural variables at the individual level, as it will allow a more thorough understanding of how people with aphasia communicate in everyday discourse and may provide insight into the development of treatment for discourse difficulties.

Current results are supported by intraclass correlation reliability analyses, which correspond to Stark et al.'s (2022) discourse reporting standards. Only variables that met at least moderate reliability scores were kept for the analyses. Namely, as far as we know, it is the first time that intraclass correlations are conducted on the global coherence measure, even if it has been used extensively in many studies (e.g.; Christiansen, 1995; Sherratt and Bryan, 2012; Barker, Young and Robinson, 2017). On another note, local coherence errors obtained poor reliability in our study, which contradicts the previous report of good inter-rater reliability with picture description tasks in people with traumatic brain injury (Marini, Galetto, *et al.*, 2011). Raw scores indicated very few local coherence errors in our samples, which might have led to poor reliability scores. Also, the percentage of total complete cohesion obtained a good reliability score, congruent with previous reports of high inter-rater agreement (Liles, 1985; Coelho and Flewellyn, 2003) and reliability (i.e., $\kappa = 0.84$; Stockbridge *et al.*, 2019) in cohesion scores. To our knowledge, previous research did not report details on individual cohesion variables but rather an overall score of reliability on all cohesion variables, which might explain discrepancies. All microstructural variables studied in the present paper obtained excellent reliability, consistent with previous studies in structured (e.g., Nicholas and Brookshire, 1993b; Boucher *et al.*, 2021; Brisebois *et al.*, 2021) and unstructured tasks (Leaman and Edmonds, 2021).

The present study has some limitations. Firstly, the small number of participants precludes the generalizability of results. The participants also presented mild to moderate aphasia and two of them fully recovered 6 months post-onset. In a larger group study, it was found that post-stroke patients demonstrated no language impairments in a standard language assessment, and yet were differentiated from controls on most discourse measures (Fromm *et al.*, 2017). Documenting discourse features in people with no perceptible language impairment might help gain a deeper understanding of challenges associated with life participation and employment (Fromm *et al.*, 2017). Even if our study did not include healthy controls, the identification of discourse changes in individuals with mild to no perceptible language impairment is highly relevant and should be explored in a larger study. Also, these four individuals might not provide be a typical representation of chronic and persistent post-stroke aphasia, they are nonetheless representative of many individuals with milder impairments following a stroke. Another limitation is that since enrolment consisted of the first four consecutive people with aphasia, by chance all four participants were men. As such, future studies should include women to determine whether the discourse patterns noted here are also found in women. Another third limitation of this study relates to the potential impact of combining the results from each discourse task at each time point. Despite the above-noted advantages of combining data from different tasks, it remains possible that each task may elicit different types of information. For instance, Stark (2019) demonstrated that linguistic properties of discourse output (i.e., microstructural features) are task-dependent in a large sample, including controls and people with aphasia. However, similar to other authors (e.g., Edmonds and Babb, 2011;

Whitworth *et al.*, 2015; Zhang *et al.*, 2020), we chose to combine the results from three different tasks to enhance the psychometric stability of measures (Brookshire and Nicholas, 1994).

Finally, as a preliminary study, we presented a case series study. As already noted, this design inherently limits the generalization of our findings to a larger population. Group studies are necessary to understand prominent trends in discourse recovery. However, qualitative examination of multilevel discourse analysis supports a better understanding of complex language and communication behaviors (Sorin-Peters, 2010), such as was found in the present study with the inverted patterns of performance between cohesion and coherence observed in P2 and P4. Individual effect sizes using the *d*-statistic have been used to quantify the relative weight of changes in the present case series design, but in order to do so, the results of multiple time points were combined. The *d*-statistic is generally used to assess the efficacy of an intervention, and baselines are then measured in the chronic phase of recovery, where no changes are expected between the repeated baselines. In our design, changes were expected and observed in the early phases of post-stroke aphasia recovery, especially in the hours/days following a stroke. The calculation of a baseline using the mean of the results obtained when spontaneous changes were present might have led to an underestimation of changes between the baselines and subacute and chronic phases of recovery. That being said, our preliminary findings support the importance of identifying individual trends and provide a more refined understanding of longitudinal discourse changes in people with post-stroke aphasia. Interindividual variabilities in discourse are acknowledged in the field (Cameron, Wambaugh and Mauszycki, 2010; Boyle, 2014, 2015) and should be considered in order to grasp the full complexity of discourse performance in the clinic and in research.

Conclusion

This study aimed to investigate the longitudinal multilevel analysis of discourse in post-stroke individuals with aphasia. Results indicate that most variables showed improvement from the *hyperacute* to the *chronic* phase of recovery both at the macro-and micro-structural levels of discourse processing in French Canadian speakers with post-stroke aphasia. Our conclusions support the further longitudinal multilevel analysis of discourse in a larger group of persons to better understand discourse recovery dynamics in post-stroke patients.

Acknowledgments

We are very grateful to all the participants and their families for contributing to this study. Also, we wish to thank Anne-Marie Chouinard for her help in manuscript data processing and preparation.

Data availability statement

The ethics review board of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Ile-de Montréal that granted permission for this research project (CIUSSS-NÎM; #2020-1900) does not currently allow publication of datasets (i.e., language samples). Individual raw discourse results are provided in Supplementary Materials 1.

Funding statement

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". This work was supported by Heart and Stroke Foundation of Canada grant (G-19-0026212).

Conflict of interest disclosure

All authors reported no conflict of interest regarding this manuscript.

Ethics approval statement

The ethics review board of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Ile-de Montréal granted permission for this research project (CIUSSS-NÎM; #2020-1900).

Patient consent statement

All participants provided written informed consent to participate in the present study.

Permission to reproduce material from other sources

n/a

References

Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. Aphasiology, 29(6), 705–723. https://doi.org/10.1080/02687038.2014.979394

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

Barker, M. S., Young, B., & Robinson, G. A. (2017). Cohesive and coherent connected speech deficits in mild stroke. Brain and Language, 168, 23–36. https://doi.org/10.1016/j.bandl.2017.01.004

Beeson, P. M., & Robey, R. R. (2006). Evaluating Single-Subject Treatment Research: Lessons Learned from the Aphasia Literature. Neuropsychology Review, 16(4), 161. https://doi.org/10.1007/S11065-006-9013-7

Bernhardt, J., Hayward, K. S., Kwakkel, G., Ward, N. S., Wolf, S. L., Borschmann, K., Krakauer, J. W., Boyd, L. A., Carmichael, S. T., Corbett, D., & Cramer, S. C. (2017). Agreed definitions and a shared vision for new standards in stroke recovery research: The Stroke Recovery and Rehabilitation Roundtable taskforce. International Journal of Stroke, 12(5), 444–450. https://doi.org/10.1177/1747493017711816

Boulanger, J., Lindsay, M., Gubitz, G., Smith, E., Stotts, G., Foley, N., Bhogal, S., Boyle, K., Braun, L., Goddard, T., Heran, M., Kanya-Forster, N., Lang, E., Lavoie, P., McClelland, M., O'Kelly, C., Pageau, P., Pettersen, J., Purvis, H., … Butcher, K. (2018). Canadian Stroke Best Practice Recommendations for Acute Stroke Management: Prehospital, Emergency Department, and Acute Inpatient Stroke Care, 6th Edition, Update 2018. International Journal of Stroke, 13(9), 949–984. https://doi.org/10.1177/1747493018786616

Bourgeois-Marcotte, J., Wilson, M. A., Forest, M., & Monetta, L. (2015). TEFREP: Épreuve de répétition de phrases en franco-québécois. Développement, validation et normalisation. Canadian Journal on Aging / La Revue Canadienne Du Vieillissement, 34(3), 391–396. https://doi.org/10.1017/s0714980815000173

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. (2020). Choosing Discourse Outcome Measures to Assess Clinical Change. Seminars in Speech and Language, 41(1), 1–9. https://doi.org/10.1055/S-0039-3401029

Brisebois, A., Brambati, S. M., Boucher, J., Rochon, E., Leonard, C., Désilets-Barnabé, M., Desautels, A., & Marcotte, K. (2021). A longitudinal study of narrative discourse in post-stroke aphasia. Aphasiology, 1–26. https://doi.org/10.1080/02687038.2021.1907295

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. (1994). 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.

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

Colin, C., & Le Meur, C. (2016). ADAPTATION DU PROJET APHASIABANK À LA LANGUE FRANÇAISE: Contribution pour une évaluation informatisée du discours oral de patients aphasiques: Vol. Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophonie. Université Paul Sabatier, Toulouse III. Toulouse, France.

Covington, M. A. (2007). MATTR user manual (CASPR Research Report 2007–05).

Edmonds, L. A., & Babb, M. (2011). Effect of Verb Network Strengthening Treatment in Moderate-to-Severe Aphasia. American Journal of Speech-Language Pathology, 20(2), 131–145. https://doi.org/10.1044/1058-0360(2011/10-0036)

Ellis, C., Rosenbek, J. C., Rittman, M. R., & Boylstein, C. A. (2005). Recovery of cohesion in narrative discourse after left-hemisphere stroke. Journal of Rehabilitation Research and Development, 42(6), 737–746.

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

Goodglass, H., Kaplan, E., Barresi, B., Goodglass, H., Goodglass, H., Goodglass, H., Goodglass, H., Goodglass, H., Kaplan, E., & Kaplan, E. (2001). The Boston Diagnostic Aphasia Examination: BDAE-3 long form kit.

Greenslade, K. J., Stuart, J. E. B., Richardson, J. D., Dalton, S. G. H., & 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

Grotz, C., Seron, X., & Adam, S. (2018). Alphaflex: Un nouveau test exécutif. Données normatives chez des sujets âgés de 50 ans et plus. Revue de neuropsychologie, 10(2), 168–178. https://doi.org/10.1684/nrp.2018.0462

Halliday, M. A., & Hasan, Ruqaiya. (1976). Cohesion in english. In English language series. Longman.

Howard, D., & Patterson, K. (1992). Pyramids and Palm Trees: A test of semantic Access from Pictures and Words (B. S. Edmunds, Ed.). Thames Valley Test Company.

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

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

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

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

Lee, J. B., & Cherney, L. R. (2018). Tau-U: A Quantitative Approach for Analysis of Single-Case Experimental Data in Aphasia. American Journal of Speech-Language Pathology, 27(1S), 495–503. https://doi.org/10.1044/2017_AJSLP-16-0197

Liles, B. Z. (1985). Cohesion in the Narratives of Normal and Language-Disordered Children. Journal of Speech and Hearing Research, 28(1), 123–133. https://doi.org/10.1044/JSHR.2801.123

Linnik, A., Bastiaanse, R., Stede, M., & Khudyakova, M. (2021). Linguistic mechanisms of coherence in aphasic and non-aphasic discourse. Aphasiology. https://doi.org/10.1080/02687038.2020.1852527

Macoir, J., Chagnon, A., Hudon, C., Lavoie, M., & Wilson, M. A. (2021). TDQ-30-A New Color Picture-Naming Test for the Diagnostic of Mild Anomia: Validation and Normative Data in Quebec French Adults and Elderly. Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 36(2), 267–280. https://doi.org/10.1093/arclin/acz048

MacWhinney, B. (2000). The CHILDES Project: Tolls for Analzying Talk: Vol. 3rd Editio. 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

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. https://doi.org/10.1044/2017_JSLHR-L-17-0012

Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. Aphasiology, 25(11), 1372–1392. https://doi.org/10.1080/02687038.2011.584690

Marini, A., Galetto, V., Zampieri, E., Vorano, L., Zettin, M., & Carlomagno, S. (2011). Narrative language in traumatic brain injury. Neuropsychologia, 49(10), 2904–2910. https://doi.org/10.1016/j.neuropsychologia.2011.06.017

Medina, J., & Fischer-Baum, S. (2017). Single-case cognitive neuropsychology in the age of big data. Cognitive Neuropsychology, 34(7–8), 440–448. https://doi.org/10.1080/02643294.2017.1321537

Nespoulous, J. L., Lecours, A. R., Lafond, D., Lemay, M. A., Puel, M., Joanette, Y., Cot, F., & Rascol, A. (1992). Protocole Montréal-Toulouse d'examen linguistique de l'aphasie: MT-86 module standard initial, M1b(2e édition révisée par Renée Béland et Francine Giroux). 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 and Hearing Research, 36(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Pedersen, P. M., Stig Jørgensen, H., Nakayama, H., Raaschou, H. O., Olsen, T. S., Jorgensen, H. S., Nakayama, H., Raaschou, H. O., Olsen, T. S., Stig Jørgensen, H., Nakayama, H., Raaschou, H. O., & Olsen, T. S. (1995). Aphasia in acute stroke: Incidence, determinants, and recovery. Ann Neurol, 38(4), 659–666. https://doi.org/10.1002/ana.410380416

Pedersen, P. M., Vinter, K., & Olsen, T. S. (2004). Aphasia after stroke: Type, severity and prognosis. The Copenhagen aphasia study. Cerebrovascular Diseases (Basel, Switzerland), 17(1), 35–43. https://doi.org/10.1159/000073896

Prins, R., & Bastiaanse, R. (2004). Review. Aphasiology, 18(12), 1075–1091. https://doi.org/10.1080/02687030444000534

Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. Aphasiology, 21(3–4), 375–393. https://doi.org/10.1080/02687030600911435

Sherratt, S., & Bryan, K. (2019). Textual cohesion in oral narrative and procedural discourse: The effects of ageing and cognitive skills. International Journal of Language & Communication Disorders, 54(1), 95–109. https://doi.org/10.1111/1460-6984.12434

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., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., Ramage, A. E., Roberts, A., Den Ouden, D. B., Brock, K., McKinney-Bock, K., Paek, E. J., Harmon, T. G., Yoon, S. O., Themistocleous, C., Yoo, H., Aveni, K., Gutierrez, S., & Sharma, S. (2021). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. American Journal of Speech-Language Pathology, 30(1s), 491–502. https://doi.org/10.1044/2020_AJSLP-19-00093

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

Stockbridge, M. D., Berube, S., Goldberg, E., Suarez, A., Mace, R., Ubellacker, D., & Hillis, A. E. (2019). Differences in linguistic cohesion within the first year following right- and left-hemisphere lesions. Aphasiology, 1–15. https://doi.org/10.1080/02687038.2019.1693026

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. Journal of Communication Disorders, 39(4), 266–291. https://doi.org/10.1016/j.jcomdis.2006.02.003

Whitworth, A., Leitão, S., Cartwright, J., Webster, J., Hankey, G. J., Zach, J., Howard, D., & Wolz, V. (2015). NARNIA: a new twist to an old tale. A pilot RCT to evaluate a multilevel approach to improving discourse in aphasia. Aphasiology, 29(11), 1345–1382. https://doi.org/10.1080/02687038.2015.1081143

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. https://doi.org/10.1371/journal.pone.0192773

Wright, H. H., & Capilouto, G. J. (2012). Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. Aphasiology, 26(5), 656–672. https://doi.org/10.1080/02687038.2012.676855

Zhang, M., Geng, L., Yang, Y., & Ding, H. (2020). Cohesion in the discourse of people with post-stroke aphasia. Https://Doi.Org/10.1080/02699206.2020.1734864, 35(1), 2–18. https://doi.org/10.1080/02699206.2020.1734864