Aphasiology

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: http://www.tandfonline.com/loi/paph20

Comparing multi-modality and constraint-induced treatment for aphasia: a preliminary investigation of generalisation to discourse

Miranda L. Rose, Zaneta Mok, Marcella Carragher, Sarah Katthagen & Michelle Attard

To cite this article: Miranda L. Rose, Zaneta Mok, Marcella Carragher, Sarah Katthagen & Michelle Attard (2016) Comparing multi-modality and constraint-induced treatment for aphasia: a preliminary investigation of generalisation to discourse, Aphasiology, 30:6, 678-698, DOI: 10.1080/02687038.2015.1100706

To link to this article: https://doi.org/10.1080/02687038.2015.1100706

Aphasiology, 2016
Vol. 30, No. 6, 678–698, http://dx.doi.org/10.1080/02687038.2015.1100706

Routledge
Taylor & Francis Group

Comparing multi-modality and constraint-induced treatment for aphasia: a preliminary investigation of generalisation to discourse

Miranda L. Rose[a,b]*, Zaneta Mok[b,c], Marcella Carragher[a,b], Sarah Katthagen[a,b] and Michelle Attard[a,b]

[a]*La Trobe University, Melbourne, Australia;* [b]*Centre for Clinical Research Excellence in Aphasia Rehabilitation, Brisbane, Australia;* [c]*Australian Catholic University, Melbourne, Australia*

(*Received 7 May 2015; accepted 22 September 2015*)

Background: At present, there is no consensus regarding the comparative efficacy of constraint and non-constraint treatments. Moreover, studies to date have measured the effects of treatment using single-word confrontation naming tasks and omnibus aphasia batteries.
Aims: We applied reliable communicative measures of lexical and communicative behaviours to investigate the comparative impact of multi-modality aphasia therapy (M-MAT) and constraint-induced aphasia therapy plus (CIAT Plus) on three discourse genres.
Methods & Procedures: Thirteen people with chronic aphasia each participated in 2 weeks of intensive CIAT Plus and 2 weeks of intensive M-MAT (30 hours over 2 weeks). Seven participants undertook CIAT Plus first, while six undertook M-MAT first. Discourse samples from three conditions were collected (picture description, Cinderella story retell and semi-structured conversation) at three time points (pre-treatment, post-CIAT Plus and post-M-MAT). Primary outcome measures included lexical measures (total words, words per minute) and communication measures (total correct information units (CIUs), CIUs per minute).
Outcomes & Results: Results varied considerably within and across participants and also across the three discourse conditions. Furthermore, changes in both positive and negative directions were evident throughout the data set. There was a slight trend for better outcomes from CIAT Plus for individuals with mild aphasia and from M-MAT for individuals with moderate aphasia.
Conclusions: In order to inform clinical practice and facilitate treatment planning and evaluation, further research is essential to continue to develop reliable discourse measures; to reach consensus on what constitutes "meaningful change" within discourse data and to take steps to mitigate against the variability inherent within discourse.

Keywords: aphasia; therapy; CIAT; M-MAT; discourse

Introduction

Clinicians face the difficult task of selecting between numerous and varied aphasia treatment approaches. Yet despite the recent Cochrane review highlighting the pressing need for comparative research (Brady, Kelly, Godwin, & Enderby, 2012), there has been very little research comparing one aphasia treatment to another. Since the development of constraint-induced aphasia therapy (CIAT; Pulvermüller et al., 2001), there has been debate in the literature regarding the advantages of constraint-induced treatments over

*Corresponding author. Email: M.Rose@latrobe.edu.au

multi-modality approaches for the treatment of anomia in chronic aphasia (Rose, 2013). CIAT has typically been administered by speech-language pathologists as an intensive treatment (30 hours over 2 weeks) in groups of two to five participants, with an emphasis on socially driven language and communication, massed practice involving card-based interactive games, and shaped verbal instruction and reinforcement. Additionally, there is strong encouragement to speak rather than use non-verbal communication—with some trials placing visual barriers between interactants to minimise non-verbal communication, and other trials restricting gesture use behind the barriers (e.g., Maher et al., 2006). In contrast, multi-modality aphasia therapy (M-MAT) mirrors the administration of CIAT with the following modifications: the instructions and cues are provided in multiple modalities (gesture, drawing, writing and reading); there are no barriers between the interactants and while the goal is for participants to speak, both non-verbal and verbal means are actively encouraged for facilitating speech (Rose & Attard, 2011).

Positive changes have been demonstrated following both approaches and at present there is no consensus regarding the comparative efficacy of constraint and non-constraint treatments (Barthel, Meinzer, Djundja, & Rockstroh, 2008; Rose, 2013; Sickert, Anders, Munte, & Sailers, 2014). While several studies have shown greater and/or longer-lasting language gains following CIAT as compared to multi-modality treatments (Kurland, Baldwin, & Tauer, 2010; Kurland, Pulvermüller, Silva, Burke, & Andrianopoulos, 2012; Maher et al., 2006), equal evidence exists to suggest that multi-modality treatments such as model-oriented aphasia therapy (MOAT; Barthel et al., 2008) and M-MAT (Rose & Attard, 2011) produce gains comparable to or greater than constraint-induced approaches (Attard, Rose, & Lanyon, 2013; Barthel et al., 2008; Rose, Attard, Mok, Lanyon, & Foster, 2013). The majority of these studies, however, have measured the treatment efficacy using single-word confrontation naming tasks and omnibus aphasia batteries (Barthel et al., 2008; Kurland et al., 2010, 2012); only a handful of intensity studies to date have measured the effects of intervention at the level of discourse (Attard et al., 2013; Maher et al., 2006; Rose et al., 2013). Although anomia is generally considered a disorder of linguistic microstructure, its effects on overall communicative effectiveness are well recognised, particularly with regard to the creation of "information gaps" in discourse (Christiansen, 1995). Everyday conversations, the most common communicative activity for people with aphasia (Davidson, Worrall, & Hickson, 2003), are not restricted to single word production and CIAT and M-MAT can involve rehearsal at phrase and sentence level. It is therefore logical to evaluate the treatment efficacy in contexts that at least approximate the "real-life" discourse-level communication of people with aphasia.

Variability within discourse

It is widely acknowledged that intra-participant variability exists within discourse samples. For the purpose of this study, we defined "discourse" as language produced above the level of sentence, collected through task-based activities (e.g., picture description) as well as naturally occurring (e.g., conversation). Variability within discourse may relate to the type of stimuli used to elicit output—that is, picture-supported, conversation or monologue (Shadden, 1998; Shadden, Burnette, Eikenberry, & DiBrezzo, 1991). In addition, various samples of discourse genres impose different roles and expectations on speakers and these are likely to impact language production. For example, in a task-based activity (such as a picture description, a procedural narrative, etc.), the person with aphasia (PWA) is asked to provide information, is unlikely to be interrupted by an

interlocutor, and it is not expected that the interlocutor will facilitate co-construction of the target message. In contrast, in naturally occurring conversation there may be more opportunities for the PWA to opt out of lengthy contributions—for example, in order to reduce the potential for errorful output and to save face (Ramsberger & Menn, 2003)—and the interlocutor is less constrained with regards to collaborating in the construction of the message. Thus the interlocutor is an important potential source of variability—for example, word finding difficulties may be perceived as less severe in conversation where communication partners utilise a "hint-and-guess" strategy (Lubinski, Duchan, & Weitzner-Lin, 1980) but more disruptive in monologic tasks where this resource is not provided. Cognitive demand is a further variable, with different genres placing different demands on short-term memory and on the resources required for sequencing and organisation (Shadden, 1998). Finally, personal relevance is likely to affect output within discourse samples, with tasks requiring the PWA to speak about familiar and/or personally relevant topics likely to elicit more content and a higher quality of language (Shadden, 1998). Unsurprisingly, there is a lack of consensus regarding not only the type of data researchers and clinicians should collect in order to obtain a representative sample of "real life" communication, but also how best to capture evidence of the effects of treatment within these data. Whilst conversation may be an obvious and certainly an ecologically valid context in which to investigate the effects of treatment, it has been argued that conversation data possess poor reliability (Boyle, 2014). In any case, the field lacks a standardised, quantitative measure of conversation (Beeke, Maxim, Best, & Cooper, 2011). In an effort to address the issue of reliability, elicitation procedures of a more constrained nature are often employed (such as procedural tasks, story retell and picture description) in which there are a single topic and a number of pre-determined targets. As such, there is a trade-off between ecological validity—the extent to which an elicitation procedure reflects the complexity of everyday interactions—and reliability as a result of methodological control.

The language production and communicative success of PWA may be influenced by aphasia severity. Amongst other factors (such as lesion site and size), the severity of aphasia on initial presentation has been demonstrated to influence post-stroke recovery (Paulucci et al., 2001). More broadly, aphasia type has been shown to have prognostic significance for physical outcomes at 1-year post-stroke (Paulucci et al., 2001). However, initial aphasia severity does not rule out the possibility for recovery; in a group of individuals presenting with acute aphasia (i.e., within 6 months post-stroke), Nouwens et al. (2014) investigated patterns of recovery in relation to aphasia severity for individuals who were randomised to either cognitive-linguistic therapy or communicative therapy. The authors found that individuals with very severe aphasia improved significantly from baseline to 3-month and from 3-month to 6-month assessments, regardless of the intervention they received (although there was a trend favouring cognitive–linguistic therapy). Therefore, aphasia severity is likely to be an important variable to control for in comparative treatment studies.

Word retrieval and discourse measures

The lack of consensus surrounding discourse as an outcome measure also extends to the methods of measuring word retrieval within discourse. Previously used measures include the amount of information a participant can convey and how efficiently—for example, Nicholas and Brookshire's (1993) number of words and number of correct information units (CIUs) per minute and the percentage of successfully retrieved nouns and verbs in

discourse (Mayer & Murray, 2003). Measures of lexical diversity (Fergadiotis & Wright, 2011; MacWhinney, Fromm, Forbes, & Holland, 2011) have also been used to calculate the range of different words produced in a text. Lexical diversity represents the speaker's ability to "...access and retrieve target words from a relatively intact knowledge base (i.e., lexicon) for the construction of higher linguistic units" (Fergadiotis & Wright, 2011, p. 1415). Historically, a common measure of lexical diversity was to calculate the ratio of unique words in a text to the total number of words (type-token ratio), where a score closer to 0 indicates a lower range of vocabulary. Although commonly used in research and clinical settings, the problems associated with type-token ratio are well-documented—namely its sensitivity to text length: larger samples reduce opportunities to produce unique words and therefore yield lower type-token ratio values compared to smaller samples (e.g., Hess, Sefton, & Landry, 1986). One proposal to mitigate against this limitation has been to cap and standardise the text length at 300 tokens (e.g., Prins & Bastiaanse, 2004). However, this has proved difficult, not least because individuals with aphasia do not always produce samples of this length (Fergadiotis, Wright, & West, 2013). Therefore, there has been continued interest in pursuing a measure of lexical diversity that might be immune to sample length.

The moving average type-token ratio (MATTR, Covington & McFall, 2010; Covington, 2007) is a measure of lexical diversity from the field of computational linguistics. MATTR estimates the type-token ratio for progressive frames of a specified number of words—that is, it provides an estimated type-token ratio for words 1–100, then 2–101, 3–102, etc. These type-token ratio values are then averaged to produce the MATTR score. In a comparative study of lexical diversity measures within discourse samples of people with aphasia, Fergadiotis and colleagues (2013) investigated the validity of four measures: MATTR, the measure of textual lexical diversity (MTLD; McCarthy, 2005), *D* (McKee, Malvern, & Richards, 2000) and the hypergeometric distribution (HD-D; McCarthy & Jarvis, 2010). MATTR (alongside MTLD) was found to be a valid measure of lexical diversity and was unbiased to text length effects (Fergadiotis et al., 2013).

Reliability of lexical retrieval measures in discourse

Due to the variable nature of discourse, the stability of behaviours of interest and what constitutes meaningful change have been questioned. In conversation, verb retrieval has been found to be stable at a group level across multiple samples of naturally occurring conversation between people with aphasia and a close other (Carragher, Sage, & Conroy, 2013). For task-based discourse samples with 12 individuals with aphasia, Boyle (2014) examined the test-retest reliability of CIUs, percent words retrieved nouns and verbs, %T-units containing word-finding behaviours (German, 1990) and *D* (an index of lexical diversity) in picture description narrative discourse tasks drawn from Nicholas and Brookshire (1993). The number of words, words per minute (WPM), CIUs and CIUs per minute were found to be sufficiently stable for use in group research, with %CIUs approaching good stability once results from an outlier were removed. WPM and CIUs per minute were sufficiently stable for use in examining changes in individuals across time. Similarly, *D* was found to be sufficiently stable across time for use in group research (Boyle, 2014). Yet the basic premise of outcome measurement in this context is not straightforward; even in the face of a stable measure with proven test-retest reliability, the operationalisation of meaningful change has yet to be agreed upon. One aspect of this ongoing debate relates to what constitutes clinical, functional and real-life change in

communication. Another aspect relates to determining what constitutes minimum change so that this change is not in fact due to error or chance. In Boyle's (2014) sample, changes greater than 47 total words and 19 total CIUs were found to reliably indicate change at the recommended 70% confidence interval for group study; changes greater than 9 WPM and 12 CIUs per minute were found to reliably indicate change at a 90% confidence interval (MDC_{90}) for individual clinical decisions. In contrast, percent words retrieved nouns and verbs were not found to be sufficiently stable to establish a minimal detectable change value. See Appendix for operational definitions of relevant measures.

Current gap in evidence base

To date, the discourse measures that have demonstrated reliability have not been applied to compare the therapy effects of CIAT and M-MAT. If outcomes of both the treatment types are comparable across measures such as confrontation naming and participant preference (Attard et al., 2013; Barthel et al., 2008; Rose et al., 2013), evidence of effects at the level of discourse may have important clinical implications, particularly if one treatment is found to demonstrate increased potential for generalisation to discourse over the other.

Aims

We were interested in exploring the potential generalisation of two aphasia treatments—with strong emphasis on word retrieval tasks—to discourse. Formal hypotheses were not generated given the preliminary nature of this exploratory work. We aimed to apply discourse measures that have demonstrated reliability (Boyle, 2014) to compare the effects of M-MAT (Rose & Attard, 2011) and constraint-induced aphasia therapy plus (CIAT Plus) (Meinzer, Djundja, Barthel, Elbert, & Rockstroh, 2005). Data from three discourse genres (picture description, narrative retell and semi-structured conversation) were collected and analysed. For outcomes related to lexical items trained within the interventions, see Rose et al. (2013) and Attard et al. (2013). In anticipation of a heterogeneous response from individual participants, we were interested to conduct preliminary investigations of the impact of individual factors (i.e., aphasic severity) on responsiveness to the treatment. We did not include time post-onset as a factor in our result based on evidence that people with aphasia can make gains following aphasia intervention regardless of time post-onset (Moss & Nicholas, 2006).

Method

Participants

Thirteen people with chronic aphasia participated in the treatment study previously reported (Attard et al., 2013; Rose et al., 2013). Table 1 describes the participant demographics and characteristics.

Standardised assessments

Participants completed a battery of linguistic assessments at three time points: once at baseline and once following each treatment. Participants were assessed on the western aphasia battery–revised aphasia quotient (WAB-R AQ; Kertesz, 2007); the Boston naming test (BNT; Kaplan, Goodglass, Weintraub, Segal, & Van Loon-Vervoorn, 2001); the

Table 1. Summary of participant characteristics.

Participant	Age	Gender	Education (years)	Lesion side, stroke type	MPO	Pre-Tx WAB AQ	Pre-Tx WAB severity	Aphasia type	Limb apraxia*	Apraxia of speech[+]	Hemi-paresis	Handed-ness
RW	49	F	15	L ischemic	77	92.8	Mild	Anomic	Absent	Absent	Right	Right
SS	59	F	16	L (type NA)	25	91.2	Mild	Anomic	Absent	Mild	None	Right
LV	69	M	15	L (type NA)	34	85.6	Mild	Anomic	None	Mild-moderate	None	Right
JP	64	F	13	L haemorrhagic	22	77.2	Mild	Anomic	Moderate	Very mild	Right	Right
BH	39	M	15	L ischemic	88	63.8	Moderate	Broca's	Mild	Mild	Right	Right
ST	46	M	16	L SAH	22	61.5	Moderate	Broca's	Mild	Mild-moderate	Right	Right
AC	64	F	17	L ischemic	40	57.4	Moderate	Conduction	Mild	Mod-severe	None	Right
JB	53	M	15	L ischemic	17	56.8	Moderate	Broca's	Mild-moderate	Mild-moderate	Right	Right
LM	74	F	15	L ischemic	79	51.9	Moderate	Broca's	Moderate	Moderate	None	Right
PD	56	M	19	L ischemic	22	50.6	Severe	Broca's	Moderate	Mild	Right	Right
MT	58	F	15	L haemorrhagic	117	47.3	Severe	Broca's	Moderate	Moderate	Right	Right
PK	66	M	10	L ischemic	58	36.2	Severe	Broca's	None	Mod-severe	None	Right
CH	55	F	12	L ischemic	82	33.5	Severe	Broca's	Mild	Mild-moderate	Right	Right

Notes: MPO = months post-onset; Pre-Tx WAB-R AQ = pre-treatment western aphasia battery-revised aphasia quotient; SAH = sub-arachnoid haemorrhage; (type NA) = type not available; L = left.
*Test of oral and limb apraxia (Helm-Estabrooks, 1992). [+]Apraxia battery for adults (Dabul, 2000).

scenario test (Van Der Meulen, Van De Sandt-Koenderman, Duivenvoorden, & Ribbers, 2010) and the communicative effectiveness index (CETI) (Lomas et al., 1989). Details of the participants' performance on the full battery of linguistic assessment have been previously reported (Attard et al., 2013; Rose et al., 2013).

Intervention

Following previously published treatment protocols (Difrancesco, Pulvermüller, & Mohr, 2012; Rose & Attard, 2011), each participant underwent 2 weeks of intensive CIAT Plus treatment and 2 weeks of intensive M-MAT treatment (30 hours each over 2-week blocks), with a 1-week break between the two treatment blocks. Treatment occurred in groups of two to five participants, with two qualified speech-language pathologists. Treatment stimuli were 160 nouns and verbs, 80 per treatment, depicted as line drawings on cards. Seven participants undertook CIAT Plus first, while six undertook M-MAT first (for full details, see Attard et al., 2013; Rose et al., 2013).

Discourse data collection

Elicitation techniques were chosen primarily in line with our previous work (Attard et al., 2013; Rose et al., 2013), incorporating methodologies which were published later which offered a stronger evidence base (Boyle, 2014; Fergadiotis et al., 2013). Discourse samples from three different conditions were collected from each participant before the treatment, after CIAT Plus and after M-MAT. These consisted of:

(1) the picture description task from the WAB-R AQ,
(2) the Cinderella story retell, and
(3) a semi-structured conversation with one of the investigators.

All samples were video-recorded and orthographically transcribed according to the systematic analysis of language transcript conventions (Miller & Iglesias, 2010). For the picture description, participants were asked to describe the picture as fully as possible. For the Cinderella story retell, participants first viewed a wordless version of the picture book and were then asked to recall the story without the picture book present. As necessary, encouragement was used by the researcher to stimulate further language output. A less-constrained discourse genre consisted of a semi-structured conversation between participants and a member of the research team. The researcher had access to a list of topics to discuss (e.g., "Tell me about your family", "Tell me about your stroke") in order to standardise topics across participants and across time points. Each conversation lasted for approximately 20 minutes, of which the initial 16 minutes were used for the purposes of analysis. This was based on the minimum length of conversation achieved by most participants; exceptions to this are detailed later.[1]

Missing data

Story retell samples were not available for two participants (RW after M-MAT; ST after CIAT Plus). Due to the severity of his expressive impairment, PD did not produce any output during the story retell condition at any time point. Data from the semi-structured conversation condition were not available for BH and PD after CIAT Plus.

Data analysis: lexical measures

Lexical measures were chosen as the primary outcome measure given that the interventions specifically targeted noun and verb retrieval. Therefore, an indicator of success could reasonably be defined as an increase in the number of lexical items produced following the treatment. Primary outcome measures included those measures, which have demonstrated stability (Boyle, 2014), that is, total number of words produced and WPM. Also included in analyses were the number of substantive nouns, the number of substantive verbs, and a measure of lexical diversity—MATTR. Although the issue of test-retest reliability and meaningful change is still questionable with measures of noun and verb retrieval, these were still of interest in the current study as they directly reflected the focus on the interventions. Point-to-point inter-rater agreement was calculated for a random selection of 20% of the transcripts for tokens of substantive nouns (96.5% achieved) and tokens of substantive verbs (95.7% achieved). MATTR was used as a measure of lexical diversity based on evidence of robust validity within discourse data (Fergadiotis et al., 2013). MATTR was calculated for the picture description and the Cinderella retell data using the computer analysis of speech for psychological research (CASPR, version 2.0; Covington, 2007). For the picture description data, the window size was set low in order to include samples from all participants. The lowest number of words produced was 20; therefore, the MATTR window size was set at 20. For the Cinderella retell data, the window size was lowered to include the wide range of total words produced across participants. However, as a result it was necessary to exclude seven data points where participants produced fewer than 100 words.[2] MATTR scores are presented as numeric data due to lack of availability of data regarding what constitutes meaningful change on this measure.

Data analysis: communication measures

Communication measures were employed to capture the amount of information conveyed by the participants, namely CIUs (following Nicholas & Brookshire's 1993 guidelines) and CIUs per minute. CIUs were not calculated for semi-structured interview data given evidence of the lack of inter-rater reliability of CIUs in this context (Oelschlaeger & Thorne, 1999). Point-to-point inter-rater agreement was calculated for a random selection of 20% of the transcripts for coding of CIUs in picture description data (91.9% achieved) and Cinderella retell data (79.6% achieved).

Results

All participants data are displayed in the tables; the figures display data for those participants with a full data set. Analysis was focused on those behaviours for which we have access to stability data (i.e., Boyle, 2014): namely, measures found to be sufficiently stable for group studies (number of words produced; number of CIUs produced) and measures found to be sufficiently stable for individual participants (WPM; CIUs per minute). As such, both group and individual results are reported later. For the purposes of data analysis, two different groups exist: Group 1 who received CIAT Plus first (n = 7) and Group 2 who received M-MAT first (n = 6). Due to the considerable heterogeneity inherent within participants' responses to treatment and across discourse measures, analyses of individual severity are also presented.

Impact of the treatment on picture description: lexical measures

Following CIAT Plus therapy, Group 2 increased the total number of words produced above Boyle's (2014) threshold of 47 words (mean = 59.3, St. Dev. = 58.3). Group 1 did not show change on the total number of words produced following CIAT Plus therapy. Neither Group 1 nor 2 showed change on this measure following M-MAT. For WPM, a number of participants showed increases beyond the MDC_{90} score of 9 words: six following CIAT Plus (JB, LM, RW, SS, LV, JP) and six following M-MAT (JB, ST, LM, AC, CH, RW). There were also negative changes that reached the MDC threshold following CIAT Plus (ST, AC, PD, CH, MT, PK) and M-MAT (PD, MT, LV, PK, JP). In terms of lexical diversity, 10 participants demonstrated increased MATTRs: six participants showed larger increases following M-MAT (BH, SS, PK, PD, MT, RW) and four following CIAT Plus (LM, AC, CH, JP) (see Table 2 and Figure 1).

Impact of the treatment on picture description: communication measures

For the number of CIUs produced, neither Group 1 nor 2 showed change that reached Boyle's (2014) threshold of 19 CIUs. Regarding communicative efficiency (i.e., CIUs/minute), two participants demonstrated change above the MDC_{90} score of 12 CIUs/minute: participants LM improved following CIAT Plus, while LM also improved following M-MAT (LM, RW). Negative changes that reached the MDC_{90} score were also observed following both CIAT Plus (AC, CH, JP) and M-MAT (JB, ST, CH, LV, PK) (see Figure 2).

Summary of the impact of treatment on picture description (lexical and communication outcomes)

Lexical and communication measures were collapsed across severity ranges (mild, moderate, severe) to assess the treatment responsiveness based on the baseline severity of aphasia; see Table 5 for changes that meet the thresholds reported by Boyle (2014). With respect to changes in outcome measures in response to the picture description stimuli, participants with mild aphasia demonstrated more positive changes following CIAT Plus (nine positive changes) compared to M-MAT (three positive changes). Participants with moderate aphasia demonstrated more positive changes in response to M-MAT (five positive changes) compared to CIAT Plus (three positive changes). Differences between the treatments were less marked for those participants with severe aphasia, with no positive changes following CIAT Plus and one positive change following M-MAT.

Impact of the treatment on story retell: lexical measures

Group 1 increased the total number of words produced (meeting Boyle's threshold of 47 words) following CIAT Plus (mean = 56.4, St. Dev. = 30.9) and also following M-MAT (mean = 47.0, St. Dev. = 74.9). For the measure of WPM, six participants showed changes beyond the MDC_{90} score of nine, four following CIAT Plus (CH, JB, PK, LV) and two following M-MAT (LM, BH). Six participants showed negative change on the measure of WPM, two following CIAT Plus (LM, JP) and four following M-MAT (JB, CH, LV, PK). In terms of lexical diversity, three participants demonstrated increased MATTRs: two participants showed larger increases following CIAT Plus (AC, LV) and one following M-MAT (SS) (see Table 3 and Figure 3).

Table 2. Changes on lexical measures (total number of words and words/min) in the picture description task following the treatment.

Participants	Total no. of words				Words/min				Substantive nouns				Substantive verbs				MATTR (100)			
CIAT-Plus first	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$
JB	116	25	139	23	199	11	238	39	25	15	20	−5	8	−3	13	5	.53	−.10	.52	−.01
ST	42	8	25	−17	115	−42	167	52	20	2	8	−12	4	1	1	−3	.85	.15	NA	NA
LM	20	−1	27	7	200	60	231	31	6	−1	11	−2	3	2	2	−1	.85	.10	.64	−.21
AC	159	45	147	−12	108	−21	126	18	18	5	13	−5	4	1	3	−1	.73	.13	.67	−0.06
PD	70	5	91	21	263	−37	228	−35	2	1	1	−1	0	0	0	0	.32	−.17	.67	.35
CH	21	0	22	1	79	−49	132	53	5	−2	4	−1	2	2	0	−2	.63	.03	.36	−.26
MT	116	−26	43	−73	178	−35	136	−42	2	−5	2	0	1	−1	2	1	.72	.00	.77	.05
Mean	*77.7*	*8.0*	*70.6*	*−7.1*																
St. Dev.	*53.9*	*22.2*	*54.8*	*32.7*																
M-MAT first	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$
BH	72	29	74	2	NA	NA	164	NA	14	0	15	1	7	1	11	4	.67	.02	.63	−.04
RW	121	−40	198	77	132	14	152	20	30	−8	52	22	11	3	15	4	.80	.02	.82	.01
SS	153	−45	190	37	101	−6	114	13	28	−1	31	3	10	−3	12	2	.86	.04	.84	−.02
LV	242	78	367	125	169	−77	187	18	27	11	35	8	12	−2	20	8	.66	.02	.68	.02
PK	59	25	51	−8	101	−12	71	−30	16	−1	12	−4	1	−1	2	1	.75	.08	.63	−.12
JP	164	−10	287	123	197	−101	207	10	21	8	27	6	12	−1	23	11	.65	−.05	.74	.08
Mean	*135.2*	*6.2*	*194.5*	***59.3***																
St. Dev.	*67.2*	*47.0*	*121.2*	*58.3*																

Notes: $Diff_1$ = change from baseline score; $Diff_2$ = change from previous phase; NA = data not available; St. Dev. = standard deviation; data in bold = reached the thresholds reported by Boyle (2014).

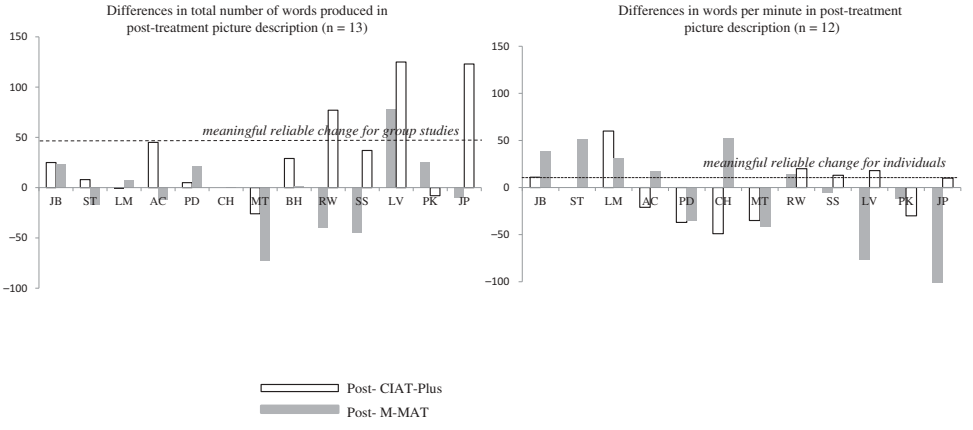

Figure 1. Changes on lexical measures (total number of words and words/min) in the picture description task following the treatment.

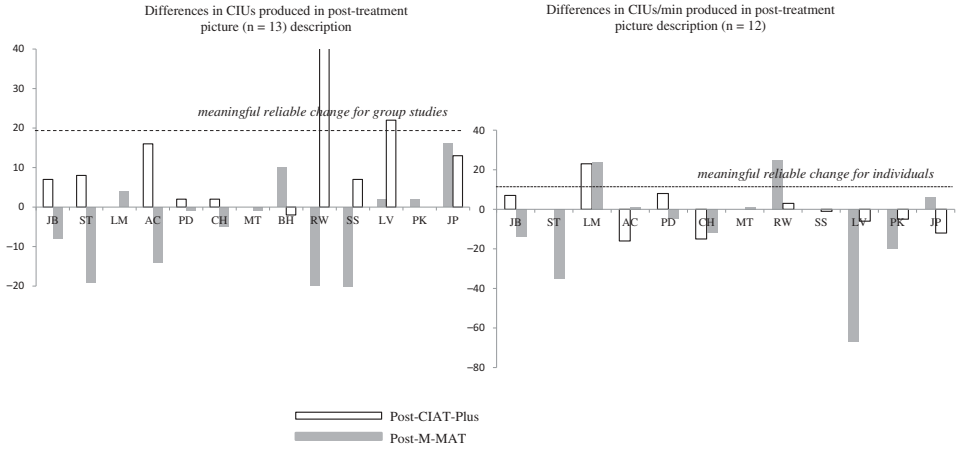

Figure 2. Changes on communication measures (CIUs and CIUs/min) in the picture description task following the treatment.

Impact of the treatment on story retell: communication measures

For the number of CIUs produced, neither Group 1 nor 2 showed change that reached Boyle's (2014) threshold of 19 CIUs. For CIUs/minute, three participants demonstrated change above the MDC_{90} score of 12 CIUs/minute, two following CIAT Plus (CH, LV) and one following M-MAT (JB). One participant (CH) showed negative change on this measure following M-MAT (see Figure 4).

Summary of the impact of treatment on story retell (lexical and communication outcomes)

Lexical and communication measures were collapsed across severity ranges (mild, moderate, severe) to assess the treatments responsiveness based on the baseline severity of aphasia; see Table 5 for changes that meet the thresholds reported by Boyle (2014). In the

Table 3. Changes on lexical measures (total number of words and words/min) in the story retell task following the treatment.

Participants	Total no. of words				Words/min				Substantive nouns				Substantive verbs				MATTR (100)			
CIAT-Plus first	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$	Post-M	$Diff_2$
JB	264	66	283	19	248	56	189	−59	21	8	30	9	2	1	10	8	.31	NA	NA	NA
ST	NA	NA	143	NA	NA	NA	156	NA	NA	NA	34	NA	NA	NA	8	NA	NA	NA	.44	NA
LM	130	55	100	−30	122	−45	167	45	8	3	6	−2	4	4	1	−3	.48	NA	NA	NA
AC	216	81	319	103	149	7	156	7	6	5	7	1	15	9	21	6	.52	.05	.48	−.04
PD	NA	NA	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0	0	0	NA	NA	NA	NA
CH	4	4	0	−4	80	80	0	−80	1	−1	0	−1	0	0	0	0	NA	NA	NA	NA
MT	170	76	317	147	159	NA	161	2	4	2	7	3	1	−3	10	9	.51	NA	.48	NA
Mean	*156.8*	***56.4***	*193.7*	***47.0***																
St. Dev.	*99.0*	*30.9*	*132.5*	*74.9*																
M-MAT first	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$	Post-C+	$Diff_2$
BH	86	51	142	56	166	16	158	−8	21	13	36	15	7	3	17	10	NA	NA	.53	NA
RW	NA	NA	919	NA	NA	NA	158	NA	NA	NA	117	NA	NA	NA	109	NA	.58	.00	NA	NA
SS	579	153	459	−120	NA	NA	132	NA	51	5	53	2	50	4	47	3	.57	.02	.55	−.02
LV	419	17	543	124	188	−27	209	21	25	13	27	2	19	7	21	2	.52	−.06	.54	.02
PK	153	37	212	59	115	−19	177	62	16	12	10	−6	0	0	0	0	.27	−.01	.24	−.02
JP	578	62	626	48	248	−8	228	−20	29	−14	32	3	67	19	30	−34	.48	−.03	.47	−.01
Mean	*363.0*	***64.0***	*483.5*	*33.4*																
St. Dev.	*232.8*	*52.5*	*284.4*	*91.0*																

Notes: $Diff_1$ = change from baseline score; $Diff_2$ = change from previous phase; NA = data not available; St. Dev. = standard deviation; data in bold = reached the thresholds reported by Boyle (2014).

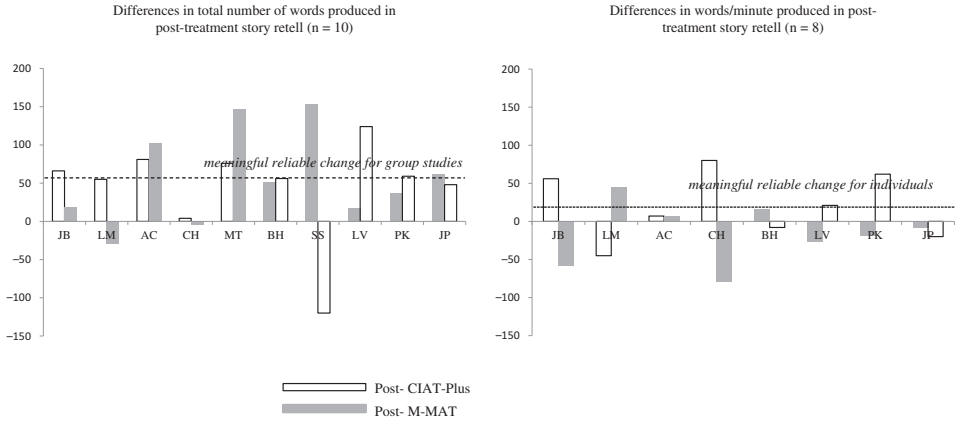

Figure 3. Changes on lexical measures (total number of words and words/min) in the story retell task following the treatment.

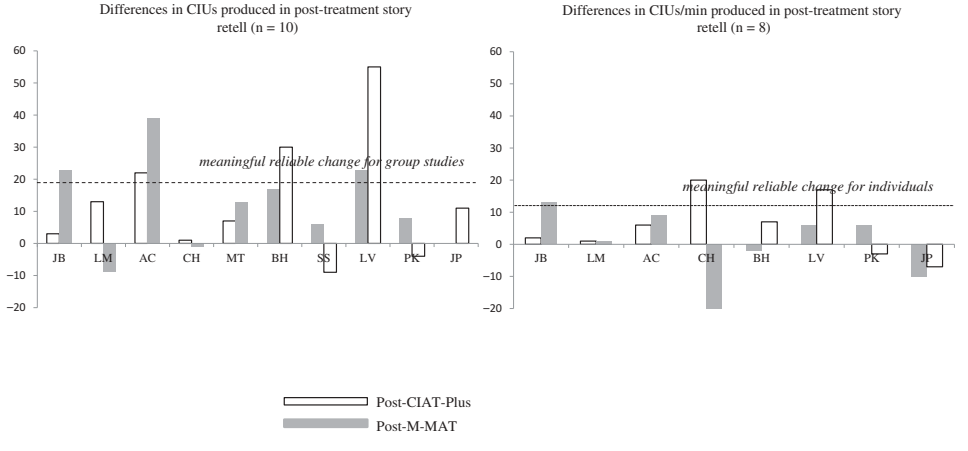

Figure 4. Changes on communication measures (CIUs and CIUS/min) in the story retell task following the treatment.

story retell task, CIAT Plus resulted in slightly more positive changes for those participants with mild aphasia (five positive changes) compared to M-MAT (three positive changes). For those participants with moderate aphasia, the treatments were equivocal (seven positive changes following both CIAT Plus and M-MAT). Participants with severe aphasia showed more positive changes following CIAT Plus (five positive changes) compared to M-MAT (one positive change).

Impact of the treatment on semi-structured interview: lexical measures

For the total number of words produced, neither Group 1 nor 2 showed change that reached Boyle's (2014) threshold of 19 CIUs. Following CIAT Plus, Group 2 demonstrated negative change that met Boyle's threshold. Ten participants showed an increased percentage of substantive nouns produced following one or both treatments (range of

increase: 1–13% of total words produced), five following CIAT Plus (JB, ST, LM, LV, PK) and six following M-MAT (LM, AC, CH, RW, PK, JP). Negative changes were seen following both treatments: six participants following CIAT Plus (AC, CH, MT, RW, SS, JP) and three following M-MAT (JB, ST, MT). For the percentage of substantive verbs produced, seven participants showed increases percentage of substantive verbs produced (range of increase: 1–4% of total words produced): three following CIAT Plus (ST, PK, JP) and four following M-MAT (ST, AC, MT, SS). Negative changes were observed following both treatments: three following CIAT Plus (JB, AC, SS) and six following M-MAT (LM, CH, BH, RW, LV, JP) (see Table 4).

In summary, both treatments resulted in bidirectional changes in lexical and communicative measures in semi-structured interview data. In terms of specific lexical classes, more participants (n = 10) showed change in the percentage of nouns produced compared to verbs (n = 7). Changes in specific lexical classes were slight (range of change in nouns: 1–13%; range of change in verbs: 1–4%) and no standard metric exists as yet to guide researchers and clinicians on how to define meaningful change in this data.

Summary of comparison of discourse results by the treatment type

Table 5 presents a summary of the changes that exceeded the thresholds suggested by Boyle (2014) across severities (in the final column) and by outcome measures in specific tasks (final row). Data are presented for individual participants, although we acknowledge that some of the measures were recommended for group studies (number of words produced; number of CIUs produced). Looking across positive changes across discourse genres, it appears that participants with moderate aphasia responded best to the treatments with 24 positive changes (with a slight advantage for M-MAT over CIAT Plus). These participants are closely followed by individuals with mild aphasia who demonstrated 22 positive changes (with an advantage for CIAT Plus). Gains in lexical retrieval and communication were also possible for participants with severe aphasia, who responded to both treatment types.

Discussion

This study investigated discourse outcomes following two contrasting intensive aphasia treatments, CIAT Plus and M-MAT, in 13 individuals with chronic post-stroke aphasia who received both treatments. All participants demonstrated some positive change on at least one discourse task following the treatment; however, there was considerable variability regarding the particular measures they showed change on and the amount of change demonstrated across the different discourse tasks. Overall, for the 13 participants, CIAT and M-MAT appeared to be broadly equally efficacious. These results represent early work in investigating the generalisation of intensive treatments to discourse and much foundational work remains to be done—particularly regarding reaching a consensus as a field as to how to quantify the generalisation of treatment effects within discourse and defining meaningful change, most likely involving not only statistical significance but also patient and carer-reported outcomes.

The current study revealed considerable variability in individual outcomes across discourse genres. This is not a surprising finding but it underlines the challenges of working with discourse data, which are vulnerable to the effects of extraneous variables such as fatigue, cognitive processing and factors relating to the communication partner. If we are interested in investigating the effect of treatment within discourse, this variability is

Table 4. Changes on lexical measures (total number of words) in the semi-structured conversation following the treatment.

Participant	Total no. of words				Substantive nouns				Substantive verbs			
CIAT-Plus first	Post-C+	$Diff_1$	Post-M	$Diff_2$	Post-C+	$Diff_1$ (%)	Post-M	$Diff_2$ (%)	Post-C+	$Diff_1$ (%)	Post-M	$Diff_2$ (%)
JB	793	17	670	−123	54	2%	34	−2%	20	−1%	17	0%
ST	243	−118	306	63	61	8%	52	−8%	10	2%	20	2%
LM	356	−44	358	2	37	3%	48	3%	9	0%	3	−2%
AC	984	82	894	−90	52	−1%	63	2%	39	−1%	43	1%
PD	NA	NA	637	NA	NA	NA	38	NA	NA	NA	19	NA
CH	137	7	130	−7	10	−2%	26	13%	2	0%	1	−1%
MT	1569	180	1603	34	81	−1%	72	−1%	38	0%	51	1%
Mean	*680.3*	*20.7*	*656.9*	*−20.2*								
St. Dev.	*545.2*	*102.7*	*490.2*	*72.1*								
M-MAT first	Post-M	$Diff_1$	Post-C+	$Diff_2$	Post-M	$Diff_1$ (%)	Post-C+	$Diff_2$ (%)	Post-M	$Diff_1$ (%)	Post-C+	$Diff_2$ (%)
BH	416	−71	NA	NA	75	0%	NA	NA	11	−1%	NA	NA
RW	1351	135	1251	−100	177	1%	109	−4%	91	−1%	85	0%
SS	624	−61	601	−23	73	0%	46	−4%	61	4%	26	−5%
LV	1519	−234	1433	−86	107	0%	128	2%	87	−1%	77	0%
PK	516	−9	510	−6	26	1%	31	1%	3	0%	6	1%
JP	963	74	885	−78	159	4%	132	−2%	86	−1%	91	1%
Mean	*898.2*	*−27.7*	*936.0*	***−58.6***								
St. Dev.	*457.9*	*128.6*	*400.8*	*41.5*								

Notes: $Diff_1$ = change from baseline score; $Diff_2$ = change from previous phase; Diff (%) = percentage of total words produced; NA = data not available; St. Dev. = standard deviation; data in bold = reached the thresholds reported by Boyle (2014).

Table 5. Number of changes (not number of participants) that meet the MDC_{90} criteria across elicitation context, outcome measures and severity.

		Picture				Story				Semi-structured conversation	Total changes (across severities)	
		Lexical		Communication		Lexical		Communication		Lexical		
Participant	Severity	Total words	WPM	Total CIUs	CIUs/min	Total words	WPM	Total CIUs	CIUs/min	Total words	Positive	Negative
RW	Mild	C+	C + M+	C + M–	M+					M + C–	22	14
SS			C+	M–		M + C–				M–	(C: 14)	(C: 6)
LV		C + M+	C + M–	C+	M–	C+	C + M–	C + M+	C+	M– C–	(M: 8)	(M: 8)
JP		C+	C + M–		C–	C + M+	C–			M + C–		
BH	Moderate					C + M+	M+	C+		M–	24	12
ST			M + C–	M–	M–					M + C–	(C: 11)	(C: 5)
AC			M + C–		C–	C + M+		C + M+		C + M–	(M: 13)	(M: 7)
JB			C + M+		M–	C+	C + M–	M+	M+	M–		
LM			C + M+		C + M+	C+	M + C–					
PD	Severe		M– C–								8	14
MT		M–	M– C–			C + M+				C+	(C: 6)	(C: 5)
PK			M– C–		M–	C+	C + M–				(M: 2)	(M: 9)
CH			M + C–		C– M–		C + M–		C + M–			
Total changes (across measures)	Positive	4 (C: 3) (M: 1)	12 (C: 6) (M: 6)	2 (C: 2) (M: 0)	3 (C: 1) (M: 2)	13 (C: 8) (M: 5)	6 (C: 4) (M: 2)	6 (C: 3) (M: 3)	3 (C: 2) (M: 1)	5 (C: 2) (M: 3)		
	Negative	1 (C: 0) (M: 1)	11 (C: 6) (M: 5)	3 (C: 0) (M: 3)	8 (C: 3) (M: 5)	1 (C: 1) (M: 0)	6 (C: 2) (M: 4)	0 (C: 0) (M: 0)	1 (C: 0) (M: 1)	9 (C: 4) (M: 5)		

Notes: C = CIAT Plus; M = M-MAT; + or – denotes a change that exceeds or decreases by the thresholds reported by Boyle (2014).

unavoidable. However, it might be somewhat mitigated in future studies by the collection of multiple samples both at baseline and following the treatment so that inherent variability can be excluded from analyses of change related to the treatment (Carragher et al., 2013). Within the current study, one possible consequence of inherent variability within this data might be the negative changes reported: across discourse genres and the treatment types, this study revealed both positive and negative changes that met the thresholds described by Boyle (2014) as beyond measurement error. Such bidirectional changes most likely reflect natural variability within discourse data. As such, changes in either direction must be interpreted with caution.

The thresholds for reliable change utilised within the current study were taken from the recent Boyle (2014) study, which utilised the Nicholas and Brookshire (1993) complex pictures and picture sequencing tasks. Whilst these are the only current measures with proven stability for measuring change in discourse, it is entirely possible that the thresholds defined in the Boyle (2014) study may not directly translate to the WAB-R AQ picture description and Cinderella story retell tasks used in the current study. It is possible that we have underestimated meaningful change in our results, given the smaller samples elicited in our study and therefore the reduced opportunity to produce words and CIUs. There is a need to undertake future studies examining the test-retest reliability of the WAB-R AQ picture description task and Cinderella story retells and develop clear measures of meaningful change, if their use in investigating discourse outcomes is to continue.

Given the variability in discourse results demonstrated in the current study, further research is required to better understand the specific individual factors that lead to generalisation across the different discourse genres. At a broader level, individuals with different severity ratings at baseline differed in their responses to the treatment. Individuals with mild aphasia tended to obtain more positive changes following CIAT Plus and individuals with moderate aphasia tended to obtain more positive changes following M-MAT. The mechanisms underlying such potential differential responses are unclear. One possible explanation concerns the different cognitive processes and learning strategies inherent in each treatment. CIAT Plus involves pairing words and pictures in a repetitive drilled practice paradigm and might operate by strengthening semantic to phonologic links for words. Therefore, CIAT Plus' therapeutic mechanism may be best suited to individuals whose semantic-phonology links are poor but who have reasonably intact semantic and phonological processing, as is common in aphasia of mild severity. M-MAT focuses on cross-modal learning and thus may be better suited to individuals with more degraded linguistic systems, such as in moderate-severe aphasia, where simple repetition of verbal targets is insufficient to enable learning and the use of multi-modal strategies become necessary. Clearly, aphasia severity is not the sole factor predicting the treatment responsiveness on discourse measures. Further research is required to define relevant participant variables underpinning varying treatment responsiveness to CIAT Plus and M-MAT. Such work could also include a detailed investigation on the impacts on non-verbal communication, which was beyond the scope of the current study.

The findings of the current study underline a larger issue within aphasia efficacy studies. It remains unclear as to the best measure to represent meaningful change in discourse, even at the level of lexical retrieval. Clearly, when the research question concerns efficiency, a timed measure is required. However, further research is required to investigate how best to evaluate quantitative changes in the semi-structured conversation context and to manage the complexity of balancing ecological validity and

reliability (in contrast to the more constrained contexts of picture description and story retell).

A limitation in the current study concerns possible order effects in the data set. Although every participant experienced both treatments at the same dose, we were not able to balance participant aphasia severity across the groups in relation to who received which treatment first. Given there may be a relationship between aphasia severity and a) the amount of treatment required to produce change and/or b) the development of treatment fatigue causing reduced treatment responsiveness over time, it is possible that our treatment comparison results are impacted by the treatment order. We argue that a large between subjects group study is required in order to balance participant demographic variables across the groups and avoid possible treatment order effects. The rationale for including a gap of 1 week between the interventions was primarily a pragmatic one to reduce the likelihood of participant drop-out; it was also based on the hypothesis that we would not see continued improvement once each treatment had been completed. On reflection, for a within-group design, a longer gap would be preferential in order to monitor any continued change and reduce potential for contamination of effects from separate treatments. However, a large between subjects group design would eliminate the need for a gap between treatments.

Future research may also consider the relationship between standardised measures of language and communication, measures of lexical retrieval in discourse and other more functionally oriented measures of discourse—for example, frequency of communication breakdown or listener/observer ratings of global communication. The field is moving towards running a greater number of randomised controlled trials—usually requiring a single primary outcome measure. Thus, a robust discussion and consensus on optimal outcome measures (particularly those concerning discourse) is timely, in an attempt to capture change in everyday communication and participation for people living with aphasia.

Acknowledgement

The authors wish to thank Associate Professor Gerasimos Fergadiotis for his advice and insight regarding MATTR.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The first author was supported by an Australian Research Council Future Fellowship [FT100100446] and the first, second and third author by the NHMRC Centre for Clinical Research Excellence in Aphasia Rehabilitation [#56935].

Notes

1. At the pre-treatment testing point, SS and LM only managed to produce samples of 8:30 minutes and 12:00 minutes, respectively. Although they both managed to produce longer samples within the semi-structured conversation following CIAT Plus and M-MAT (ranging from 19:30 to 21:09 minutes), all analyses were limited to the 8:30 minute time point for SS and the 12:00 minute time point for LM to facilitate comparison between the samples at each testing phase. Two participants with severe aphasia were not able to achieve the 20-minute target; therefore, the time used for their semi-structured conversation analyses was 5 minutes (CH) and 11 minutes (MT).
2. Excluded from MATTR analysis: CH post-CIAT Plus; MT baseline; BH baseline, post-M-MAT; JB baseline; LM baseline, post-M-MAT.

References

Attard, M., Rose, M., & Lanyon, L. (2013). The comparative effects of multi-modality aphasia therapy and constraint-induced aphasia therapy-plus for severe chronic Broca's aphasia: An in-depth pilot study. *Aphasiology*, *27*, 80–111. doi:10.1080/02687038.2012.725242

Barthel, G., Meinzer, M., Djundja, D., & Rockstroh, B. (2008). Intensive language therapy in chronic aphasia: Which aspects contribute most? *Aphasiology*, *22*, 408–421. doi:10.1080/02687030701415880

Beeke, S., Maxim, J., Best, W., & Cooper, F. (2011). Redesigning therapy for agrammatism: Initial findings from the ongoing evaluation of a conversation-based intervention study. *Journal of Neurolinguistics*, *24*, 222–236. doi:10.1016/j.jneuroling.2010.03.002

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

Brady, M., Kelly, H., Godwin, J., & Enderby, P. (2012). *Speech and language therapy for aphasia following stroke*. The Cochrane Collaboration. John Wiley & Sons.

Carragher, M., Sage, K., & Conroy, P. (2013). The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation. *Neuropsychological Rehabilitation*, *23*, 846–887. doi:10.1080/09602011.2013.832335

Christiansen, J. A. (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*, *51*, 291–317. doi:10.1006/brln.1995.1062

Covington, M. A. (2007). *CASPR Research Report 2007-05: MATTR user manual*. Athens: The University of Georgia.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*, 94–100. doi:10.1080/09296171003643098

Dabul, B. L. (2000). *Apraxia battery for adults (ABA-2)* (2nd ed.). Austin, TX: Pro-Ed.

Davidson, B., Worrall, L., & Hickson, L. (2003). Identifying the communication activities of older people with aphasia: Evidence from naturalistic observation. *Aphasiology*, *17*, 243–264. doi:10.1080/729255457

Difrancesco, S., Pulvermüller, F., & Mohr, B. (2012). Intensive language-action therapy (ILAT): The methods. *Aphasiology*, *26*, 1317–1335. doi:10.1080/02687038.2012.705815

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

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech Language Pathology*, *22*, S397–408. doi:10.1044/1058-0360(2013/12-0083)

German, D. J. (1990). *Test of adolescent adult word finding*. Austin, TX: Pro-Ed.

German, D. J. (1991). *Test of word finding in discourse (TWFD): Administration, scoring, interpretation and technical manual*. Austin, TX: Pro-Ed.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M. A. K., & Matthiessen, C. (2006). *Construing experience through meaning: A language-based approach to cognition*. London: Continuum.

Helm-Estabrooks, N. (1992). *Test of oral and limb apraxia*. Chicago: The Riverside Publishing Group.

Hess, C. K., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research*, *29*, 129–134. doi:10.1044/jshr.2901.129

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of English*. Cambridge: Cambridge University Press.

Kaplan, E., Goodglass, H., Weintraub, S., Segal, O., & Van Loon-Vervoorn, A. (2001). *Boston naming test*. Austin, TX: Pro-Ed.

Kertesz, A. (2007). *Western aphasia battery-revised examiner's manual*. USA: Pearson Education.

Kurland, J., Baldwin, K., & Tauer, C. (2010). Treatment-induced neuroplasticity following intensive naming therapy in a case of chronic Wernicke's aphasia. *Aphasiology*, *24*, 737–751. doi:10.1080/02687030903524711

Kurland, J., Pulvermüller, F., Silva, F., Burke, K., & Andrianopoulos, M. (2012). Constrained versus unconstrained intensive language therapy in two individuals with chronic, moderate-to-severe aphasia and apraxia of speech: Behavioral and fMRI outcomes. *American Journal of Speech-Language Pathology*, *21*, S65–S87A. doi:10.1044/1058-0360(2012/11-0113)

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Development and psychometric evaluation of a functional

communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*, 113–124. doi:10.1044/jshd.5401.113

Lubinski, R., Duchan, J., & Weitzner-Lin, B. (1980). Analysis of breakdowns and repairs in adult aphasic communication. In R. Brookshire (Ed.), *Clinical Aphasiology Conference Proceedings* (pp. 111–116). Minneapolis, MN: BRK Publishers.

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

Maher, L., Kendall, D., Swearengin, J., Rodriguez, A., Leon, S., Pingel, K., & Rothi, L. J. (2006). A pilot study of use-dependent learning in the context of constraint induced language therapy. *Journal of the International Neuropsychological Society*, *12*, 843–852. doi:10.1017/S1355617706061029

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology*, *17*, 481–497. doi:10.1080/02687030344000148

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity* (Doctoral dissertation). University of Memphis. Available from Proquest Dissertations and Theses (UNI no. 3199485).

McCarthy, P. M., & Jarvis, S. (2010). MTLD, voc-D and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behaviour Research Methods*, *42*, 381–392. doi:10.3758/BRM.42.2.381

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*, 323–337. doi:10.1093/llc/15.3.323

Meinzer, M., Djundja, D., Barthel, G., Elbert, T., & Rockstroh, B. (2005). Long-term stability of improved language functions in chronic aphasia after constraint-induced aphasia therapy. *Stroke*, *36*, 1462–1466. doi:10.1161/01.STR.0000169941.29831.2a

Miller, J., & Iglesias, A. (2010). *Systematic analysis of language transcripts (SALT)*. Madison, WI: University of Wisconsin-Madison.

Moss, A., & Nicholas, M. (2006). Language rehabilitation in chronic aphasia and time post-onset: A review of single-subject data. *Stroke*, *37*, 3043–3051. doi:10.1161/01.STR.0000249427.74970.15

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

Nouwens, F., De Jong-Hagelstein, M., De Lau, L. M. L., Dippel, D. W. J., Koudstaal, P. J., Van De Sandt-Koenderman, W. M., & Visch-Brink, E. G. (2014). Severity of aphasia and recovery after treatment in patients with stroke. *Aphasiology*, *28*, 1168–1177. doi:10.1080/02687038.2014.907865

Oelschlaeger, M., & Thorne, J. C. (1999). Application of the correct information unit analysis to the naturally occurring conversation of a person with aphasia. *Journal of Speech Language and Hearing Research*, *42*, 636–648. doi:10.1044/jslhr.4203.636

Paulucci, S., Grasso, M. G., Antonucci, G., Bragoni, M., Troisi, E., Morelli, D., & Rizzi, F. (2001). Mobility status after inpaitent stroke rehabilitation: 1-year follow-up and prognostic factors. *Archives of Physical Medicine and Rehabilitation*, *82*, 2–8. doi:10.1053/apmr.2001.18585

Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*, 1075–1091. doi:10.1080/02687030444000534

Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-induced therapy of chronic aphasia after stroke. *Stroke*, *32*, 1621–1626. doi:10.1161/01.STR.32.7.1621

Ramsberger, G., & Menn, M. (2003). Co-constructing Lucy: Adding a social perspective to the assessment of communicative success in aphasia. In G. Goodwin (Ed.), *Conversation and brain damage*. New York, NY: Oxford University Press.

Rose, M. (2013). Releasing the constraints on aphasia therapy: The positive impact of gesture and multimodality treatments. *American Journal of Speech Language Pathology*, *22*, S227–239. doi:10.1044/1058-0360(2012/12-0091)

Rose, M., & Attard, M. (2011). *Multi-modality aphasia therapy (M-MAT): A procedural manual*. Melbourne: La Trobe University.

Rose, M., Attard, M., Mok, Z., Lanyon, L., & Foster, A. (2013). Multi-modality aphasia therapy is as efficacious as a constraint-induced aphasia therapy for chronic aphasia: A phase 1 study. *Aphasiology*, *27*, 938–971. doi:10.1080/02687038.2013.810329

Shadden, B. B. (1998). Information analysis. In L. R. Cherney, B. B. Shadden, & C. A. Coelho (Eds.), *Analyzing discourse in communicatively impaired adults*. Gaithersburg, MD: Aspen Publishers.

Shadden, B. B., Burnette, R., Eikenberry, B., & DiBrezzo, R. (1991). All discourse tasks are not created equal. *Clinical Aphasiology*, *20*, 327–341.

Sickert, A., Anders, L., Munte, T., & Sailers, M. (2014). Constraint-induced aphasia therapy following sub-acute stroke: A single-blind, randomised clinical trial of a modified therapy schedule. *Journal of Neurology, Neurosurgery & Psychiatry*, *85*, 51–55. doi:10.1136/jnnp-2012-304297

Van Der Meulen, I., Van De Sandt-Koenderman, M., Duivenvoorden, H. J., & Ribbers, G. M. (2010). Measuring verbal and non-verbal communication in aphasia: Reliability, validity, and sensitivity to change of the Scenario Test. *International Journal of Language & Communication Disorders*, *45*, 424–435. doi:10.3109/13682820903111952

Wright, H. H., Silverman, S. W., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, *17*, 443–452. doi:10.1080/02687030344000166

Appendix. Operational definitions of relevant outcome measures

Term	Definition
Correct information units (CIUs)	"Words that are intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the content of the picture(s) or the topic" (Nicholas & Brookshire, 1993, p. 348). CIUs measure communicative informativeness and efficiency.
Percentage of correct information units (%CIUs)	The number of CIUs divided by the total number of words in the language sample (Nicholas & Brookshire, 1993).
Type-token ratio (TTR)	A measure of lexical diversity used to capture conversational vocabulary in adults with aphasia (Wright, Silverman, & Newhoff, 2003).
Moving average type-token ratio (MATTR)	A measure of lexical diversity based on progressive frames of a specified number of words (Covington, 2007; Covington & McFall, 2010).
T-units	A global index of word-finding behaviours. A T-unit is a grammatical structure, which is a complete thought, consisting minimally of a noun phrase plus verb phrase (Boyle, 2014; German, 1991).
Substantive nouns*	All proper and common nouns except the following general nouns: *thing, stuff, people, person, animal*, and *place* (Halliday & Hasan, 1976; Halliday & Matthiessen, 2006).
Substantive verbs*	Excludes auxiliaries; all forms of the verbs *be, have, do*; and any light uses of verbs, for example, *I gave a talk, I take a nap, I made a copy* (Huddleston & Pullum, 2002).

*Because our specific interest for this paper lies in propositional content productivity, we excluded from noun and verb counts: i) all comments made by the participants on the tasks (for example, *I can't say the word*); ii) all qualifiers and modifiers (e.g., *I think I have to go to the shower, I guess it was yesterday*) and iii) all mazes—that is, repetitions, revisions, false starts and filled pauses (Miller & Iglesias, 2010).