



Published in final edited form as:

Int J Speech Lang Pathol. 2018 August ; 20(4): 406–421. doi:10.1080/17549507.2017.1293158.

Measuring discourse coherence in anomic aphasia using Rhetorical Structure Theory

Anthony Pak-Hin Kong¹, Anastasia Linnik², Sam-Po Law³, and Waisa Wai-Man Shum³

¹Department of Communication Sciences and Disorders, University of Central Florida, Orlando, FL, USA

²International Doctorate for Experimental Approaches to Language and Brain (IDEALAB), University of Potsdam, Germany; University of Groningen, Netherlands; Macquarie University, Australia; University of Trento, Italy; University of Newcastle, UK

³Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong SAR

Abstract

Purpose—The existing body of work regarding discourse coherence in aphasia has provided mixed results, leaving the question of coherence being impaired or intact as a result of brain injury unanswered. In this study, discourse coherence in non-brain-damaged (NBD) speakers and speakers with anomic aphasia was investigated quantitatively and qualitatively.

Method—Fifteen native speakers of Cantonese with anomic aphasia and 15 NBD participants produced 60 language samples. Elicitation tasks included story-telling induced by a picture series and a procedural description. The samples were annotated for discourse structure in the framework of Rhetorical Structure Theory (RST) in order to analyse a number of structural parameters. After that 20 naïve listeners rated coherence of each sample.

Result—Disordered discourse was rated as significantly less coherent. The NBD group demonstrated a higher production fluency than the participants with aphasia and used a richer set of semantic relations to create discourse, particularly in the description of settings, expression of causality, and extent of elaboration. People with aphasia also tended to omit essential information content.

Conclusion—Reduced essential information content, lower degree of elaboration, and a larger amount of structural disruptions may have contributed to the reduced overall discourse coherence in speakers with anomic aphasia.

Keywords

discourse analysis; aphasia; speech-language pathology

Corresponding author: Anthony Pak-Hin Kong, Department of Communication Sciences and Disorders, University of Central Florida, HPA-2 106, PO Box 162215, Orlando, FL 32816-2215; antkong@ucf.edu; Phone: (407) 823-4791; Fax: (407) 823-4816.

Declaration of conflicting interests

The authors report no conflict of interests.

Introduction

Oral narratives are a crucial form of discourse used in everyday life. Their production involves both micro- and macro-linguistic processes. The micro-linguistic level includes lexical processing, organisation of phonological information into morphological strings and words, and then into syntactic constructions (Marini & Fabbro, 2007). On the other hand, macro-linguistic abilities are associated with discourse processing at the supra-sentential level. For example, cohesion, or semantic relations between contiguous utterances, established through the use of lexical and grammatical devices, such as conjunctions, co-reference, and ellipsis (Halliday & Hason, 1976), is a macro-linguistic property. These discourse properties also mediate local connections between sentences or utterances with the global relations among propositions, minimal semantic units within a text, to integrate linguistic and conceptual features (Kintsch, 1994)

Coherence is a complex phenomenon (e.g. Givón, 1995), and the problem of defining it has been extensively addressed by a number of researchers (Foltz, 2007; Gernsbacher & Givon, 1995; Kehler, 2001, 2004; Kintsch & Van Dijk, 1978; Sanders & Spooren, 2001; van Dijk 1977). Coherence is a crucial property which transforms a sequence of sentences into a discourse (e.g. Kehler, 2001; Ultowska et al., 2003). It refers to the semantic connectedness or 'hanging together' of speech, or the semantic connectedness of discourse at the propositional level (Van Dijk, 1980). Discourse coherence can be further divided into local and global. Local coherence reflects a speaker's ability to establish connection between currently processed information with the immediately preceding context (O'Brien & Albrecht, 1993), whereas global coherence refers to a speaker's ability to semantically relate remote utterances to the theme, topic, or gist of a discourse (Kintsch & Van Dijk, 1978).

Discourse produced by speakers with aphasia is often perceived as vague and lacking clarity (Early & VanDemark, 1985; Gleason et al., 1980; Ulatowska, North, & Macaluso-Haynes, 1981). On the other hand, it has been noted that PWA preserve remarkably good functional communication skills (Holland, 1982; Huber, 1990; Olness & Ulatowska, 2010; Ulatowska, Allard, & Chapman, 1990). Studies of aphasic discourse to date have provided inconsistent findings. It has been hypothesised that the decreased understandability of discourse in aphasia reflects difficulties with establishing coherence (Van Dijk, 1980; Ulatowska et al., 1981).

Research on macro-linguistic abilities of persons with aphasia (PWA) has offered a variety of methods to evaluate local and global coherence defined through the concepts of topic maintenance and thematic unity, as well as coherence error analysis (e.g. Christiansen, 1995; Coelho & Flewellyn, 2003; Olness, 2006; Olness & Ulatowska, 2011; Ulatowska, North, & Macaluso-Haynes, 1981; Ulatowska et al., 1990; Wright, Capilouto, & Koutsoftas, 2013). They can be roughly divided into perceptual (involving raters' interpretation and evaluation) and factual (or data-driven) ones. The first group is represented by a number of rating scales used for assessing discourse coherence in aphasiology. Ratings are intended to assess the overall level of coherence of a discourse either through the assessment of every utterance's relatedness to the semantic unity of the discourse, or through a single rating of the degree to which it can be considered a unity. Neither of these case ratings provide any insights into the

linguistic factors contributing to the maintenance of coherence. On the other hand, the advantages of this approach are, firstly, that the perceptual nature of ratings reflects the fact that coherence is co-established by both speaker and listener (e.g. Gernsbacher & Givon, 1995; Wright & Capilouto, 2012), and secondly, high reliability and validity scores have been reported for some of the existing scales (e.g. Coelho et al., 2013; Glosser & Deser, 1991; Wright & Capilouto, 2012a).

Data-driven methods included story grammar analysis (Coeho et al., 1994), in which the number of complete episodes consisting of an initiating event, an action, and a direct consequence was calculated, similar to analysis of narrative superstructure (Labov 1972), i.e. setting, initiating event, complicating action, resolution, coda (Olness, 2006; Olness & Ulatowska, 2011), and analyses of propositional content and coherence violations (Andreetta, Cantagallo, & Marini, 2012; Christiansen, 1995). In addition, Wright and Capilouto (2012a) studied the effect of micro-linguistic impairments on PWA's maintenance of global coherence in a story-telling task. Their results suggested that reduced information content and lexical diversity had an effect on coherence in the stories of PWA, which is consistent with the findings in a number of earlier studies (e.g. Christiansen 1995, Coelho & Flewellyn 2003). Finally, Andreetta et al. (2012) investigated the effect of lexical retrieval difficulties on macro-linguistic processing during the construction of a narrative. They concluded that impaired word finding reduced the levels of sentence completeness and the overall degree of cohesion across utterances, whereas lexical fillers and repetitions lowered the overall level of global coherence in spoken discourse.

There is a lack of consensus on whether discourse coherence is impaired for people with aphasia. For example, a number of studies demonstrated a significantly reduced degree of discourse coherence and/or pathological coherence in aphasia (e.g. Andreetta et al., 2012; Andreetta & Marini, 2015; Christiansen, 1995; Coelho & Flewellyn, 2003; Wright & Capilouto, 2012), whereas other researchers have provided evidence to it being within normal limits (Glosser & Deser, 1991; Ulatowska et al., 1981; Ulatowska et al., 2013). Further disparities are related to the more global question about the origins of coherence. It has been suggested that impaired micro-linguistic skills lead to macro-linguistic processing difficulties. For example, poor cohesion, often caused by morphosyntactic deficits, was found to be correlated with global coherence (e.g. Armstrong, 1987). Contrary to this idea, several studies demonstrated that thematic coherence in oral discourse can remain relatively intact despite PWA's micro-linguistic deficits (e.g. Coelho & Flewellyn, 2003; Glosser & Deser, 1991; Ulatowska et al., 1981, Ulatowska, Weiss-Doyell, Freedman-Stern, & Macaluso-Haynes, 1983). This indicates that micro- and macro-level of discourse may be organised independently. The techniques discussed above are valuable for assessment and therapy outcomes evaluation as they reliably estimate the level of coherence in discourse; nevertheless, they do not provide an insight on the reasons behind coherence impairment or preservation. As a solution to this problem, a number of multi-level approaches have been developed and implemented to explore the relationship between the micro- and macro-linguistic abilities of PWA (e.g. Marini, Andreetta, & Carlomagno, 2011; Sherratt, 2007; Wright & Capilouto, 2012).

Several factors have limited the generalisability of the findings to date from a group of study participants to the wider aphasic population. One of the issues is the variability in discourse genres used in previous studies. Elicitation tasks included picture description (e.g. Andreetta et al., 2012; Marini, Boewe, Caltagirone, & Carlomagno, 2005; Saffran et al., 1989), personal narratives (e.g. Glosser & Deser, 1991; Olness & Ulatowska, 2011), and procedural discourse (e.g. Ulatowska et al., 1981, Ulatowska et al., 1990). Longacre (1996) emphasised that different genres are associated with specific patterns of linguistic features, as well as logical and thematic organisation. Different elicitation tasks might also impose different cognitive and linguistic demands on PWA (Bliss & McCabe, 2006). Specifically, story-telling using pictures can be considered cognitively less demanding than personal narrative or expository discourse because these two tasks require organisation of several ideas without visual support. Marini et al. (2005) investigated discourse of a large group of healthy adults and found a higher degree of coherence for narratives elicited using a sequential picture description task, as compared to a single picture. They attributed this difference to the higher level of inter-relationship among characters in a series of pictorial stimuli with multiple themes within the description. Olness (2006) examined the difference between narratives elicited using sequences of pictures and single pictures. She concluded that the latter have limitations for discourse studies, specifically “what is traditionally elicited by single pictures may not be discourse” (p.185) as it does not require connectivity, or coherence. Wright and Capilouto (2012a, 2012b), in turn, reported significantly lower coherence scores for personal recounts, as compared to picture-elicited stories, in PWA. These findings are clinically relevant to the selection of assessment and treatment materials targeting discourse production. Based on these considerations, two types of narrative were chosen for the current study, namely procedural description elicited with a single picture and story-telling with a series of pictures. One of the goals of this study was to test the applicability of the structural discourse analysis using Rhetorical Structure Theory to aphasic data. Following this consideration, the choice of elicitation tasks was motivated by the comparability of resulting samples. Future extensions of this study should include more complex narrative types, such as personal recounts.

Apart from different elicitation techniques used in previous research on discourse in aphasia, the variety of methods developed to analyse coherence is another factor possibly contributing to the disparities in results. Comparing the results of studies using different rating scales and narrative structure investigation methods is not always straightforward due to the differences in scoring procedures, including the object being assessed (Linnik, Höhle, & Bastiaanse, 2016). Different aspects of coherence evaluated in previous studies include, for example, appropriateness and completeness of thematic content (e.g. Glosser & Deser, 1991; Wright, Koutsoftas, Fergadiotis, & Capilouto, 2010), coherence violations (Christiansen, 1995; Marini et al., 2011; Andreetta et al., 2012), and semantic-pragmatic unity of discourse as a whole (Ulatowska, Olness, & Williams, 2004; Olness, 2006). The current study followed the latter path and addressed coherence as an intrinsic trait of discourse, specifically, its overall unity and connectedness.

In this study we suggest exploring macro-linguistic impairments in aphasia and the linguistic and cognitive processes underlying discourse production using a combination of both perceptual and structural approaches. Specifically, we apply Rhetorical Structure Theory

(Mann & Thomson, 1988), a relational approach to the investigation of discourse structure widely used in healthy discourse analysis, together with a global coherence rating scale. Through the integration of data-driven and perceptual perspectives we expect to be able to grasp different aspects of coherence as a phenomenon, and shed light on the disparate results of previous studies.

Rhetorical Structure Theory

It has been argued that coherent discourse has an internal structure and that this structure is hierarchical, rather than linear (e.g. Fox, 1987; Grosz & Sidner, 1986; Hobbs, 1985). One of the frameworks formalising this hypothesis is Rhetorical Structure Theory (RST; Mann & Thompson, 1988; Taboada & Mann, 2006). According to RST, discourse consists of elementary discourse units (EDUs), minimal building blocks of discourse structure which roughly correspond to clauses (Mann & Thompson, 1988). These basic segments are connected to each other with *coherence* (also called “semantic”, or “rhetorical”) *relations* (‘Consequence’, ‘Cause’, ‘Evaluation’, ‘Elaboration’, etc.) forming a complete tree-like structure (Figure 1). Coherence is thus achieved through the establishment of discourse structure.

RST has been widely used in written and spoken discourse analysis, as well as in computational linguistics (see Taboada & Mann, 2006a, 2006b for a review). A number of other approaches were based on similar theoretical considerations about discourse structure, but proposed methodologically different solutions or frameworks (Cristea, Ide, & Romary, 1998; Grosz & Sidner, 1986; Lascarides & Asher, 2007; Moore & Polack, 1992; Moser & Moore, 1996; Prasad et al., 2008; Walker 1996; Wolf & Gibson, 2005). None of them has been as extensively tested as RST. Our choice of RST for the analysis of discourse in aphasia is motivated primarily by its plausibility with respect to certain cognitive processes involved in discourse production. Specifically, RST is based on the view that coherence relations are cognitive entities responsible for holding spans of discourse together and playing an important role in the interpretation of discourse. Different structures result in different interpretations of a discourse (e.g. Sanders, 1993; Taboada & Mann, 2006). Hence, the ability to establish coherence relations or a discourse structure of some sort is crucial for communication. Psychological validity and technical adequacy of RST has been questioned, due to the constraints it imposes on discourse, such as the requirement of a single inference when linking discourse segments or the lack of a unified opinion on the number and classification of coherence relations among the researchers using relational approaches. These points of criticism, among a number of others, have been addressed at length in previous literature (Marcu, 2000, 2003; Taboada 2004; and Taboada & Mann, 2006), and, despite constituting an important and interesting discussion, will not be reiterated in this paper. Although RST analysis may be more laborious and complicated than the methods previously used in aphasic discourse structure investigations, it is the first one that addresses aphasic speakers’ ability to establish connectivity between discourse segments using semantic and pragmatic relations. It delivers a more fine-grained result, thus granting an opportunity to look deeper into the processes of discourse organisation. Crucially, as compared to other methods, it provides information on relations (or a lack of thereof)

between discourse spans in addition to the quantitative measure of propositional content, which is particularly useful for clinical assessment of narrative deficits in PWA.

RST has a number of other advantages. RST partially models relative salience distribution during discourse production (Marcu, 1999; Stede, 2008). Namely, when linking two EDUs, one has to decide on the so-called *nuclearity* status of the constituents. In a *mononuclear* relation (see Figure 1a, EDUs No. 1-4), a nucleus is the constituent containing the more salient piece of information, while a satellite is the constituent, which could be eliminated without a substantial loss of essential information. In case of two EDUs being equally important for the discourse, a *multinuclear* relation will be assigned. Apart from the organisation of information in the discourse flow, semantic (also called “discourse” or “rhetorical”) relations in RST reflect some of the cognitive processes underlying discourse production, such as subjective evaluation, internal reasoning (e.g. ‘Explanation’, ‘Cause’, ‘Reason’), or motivations (‘Purpose’).

Chinese, including Cantonese, is generally characterised by the lack of inflectional morphology (Packard, 2000; Wang & Sun, 2015) and frequent use of elliptical sentences (Chung, Code, & Ball, 2004). The common absence of number and gender agreement between pronouns and nouns and the usual omission of topic and grammatical subjects in sentences can present difficulties for the theories relying on signalling of relations between parts of discourse, such as through discourse markers. Although some suggestions about connectors signalling certain relations have been made for English (Carlson & Marcu, 2001), RST does not imply the surface signalling of relations (see Taboada, 2006), which makes it an appropriate choice for Cantonese (and Chinese) discourse analysis. Since RST is not language-specific, it can also be used for cross-linguistic comparisons at the discourse level (see Iruskieta et al., to appear for discussion). The findings of this work are thus valuable for the understanding of the general principles of language production.

Aim

In this study we aimed to examine the differences in the way discourse coherence is established by people with aphasia (PWA) and non-brain-damaged (NBD) speakers. We hypothesised that PWA would demonstrate a significant impairment in the macro-linguistic organisation and, consequently, a reduced level of coherence (Andreetta & Marini, 2014; Christiansen, 1995). RST analysis was implemented in order to capture the interruptions in the process of building a discourse structure either as a result of a purely macro-linguistic deficit, or following word-finding deficits, paragrammatic errors, omission or repetition of important propositions normally exhibited by PWA.

In light of the hypothesis that different levels of cognitive demand are involved in different discourse elicitation tasks, our second aim was to examine the effect of genre on discourse structure and coherence in PWA. Several parameters, such as the availability of visual cues, presence of a thematic structure and story characters, have been suggested to have an effect on cognitive demand and as a result on the linguistic performance of PWA (e.g. Ulatowska et al., 2003; Olness, 2006; Fergadiotis, Wright, & Capilouto, 2011). Based on previous findings, we expected to find differences in structural properties in discourse of different genres, such as lower overall complexity expressed in the amount of elaboration and shorter

output in procedural discourse, and higher degree of structure depth, evaluation or commentary in story-telling, as well as a wider range of structure types and relations between spans of discourse in story-telling (Longacre, 1996; Olness, 2007; Pritchard, Dipper, Morgan, & Cocks, 2015; Ulatowska et al., 1981, 1983).

Lastly, this study also intended to validate a quantitative structural approach to the investigation of discourse coherence based on RST through the analysis of correlations between its outcomes and naïve listeners' ratings.

Method

Corpus and participants

The data used in the current study are a part of the corpus of Cantonese discourse of PWA (Kong & Law, 2016), for which data collection methods and stimuli were adapted for Cantonese from the AphasiaBank protocol (MacWhinney, Fromm, Forbes, & Holland, 2011). All language samples collected were transcribed in the Codes for the Human Analysis of Transcripts format (CHAT; MacWhinney, 2000). At the time of this study, the database contained transcripts of 149 unimpaired native speakers of Cantonese recruited from Hong Kong and 76 PWA (resulting from a single stroke and with a post onset time of at least six months) who completed nine different language tasks. At the time of intake, based on the subjects' performance on the initial part of the protocol, the personal data questionnaire, and the information about demographic, social, family, and medical history, none of the participants showed significantly impaired cognitive status that would impact their responses in the narrative tasks.

A total of 60 transcripts were extracted from the database representing 13 male and two female native Cantonese participants diagnosed with anomic aphasia, according to the Cantonese version of the Western Aphasia Battery (CAB; Yiu, 1992). An equal number of non-brain-damaged (NBD) participants were selected to match the participants with aphasia in gender, age (± 5 years), and education level (± 1 year). The age of the group with aphasia ranged between 43 and 72 years (mean = 55.2 years; SD = 9.70 years) and the CAB aphasia quotients ranged from 77.1/100 to 99/100 (mean = 89.6; SD = 7.09). The age of NBD subjects ranged between 44 and 71 years (mean = 55.8 years; SD = 8.08 years); none of them had any previous history of psychiatric or neurologic illness, learning disabilities, hearing and/or visual impairments that would affect their use of language. The education levels for both speaker groups ranged between six and 13 years.

Transcripts from two discourse tasks were chosen to study the effect of genre on discourse macrostructure, namely, (1) a description of the procedure of making an egg and ham sandwich, elicited using a single picture with photos of the ingredients, and (2) a narrative elicited using a series of six pictures (black and white line drawings), depicting a boy who refused to take an umbrella on a rainy day. Specifically, the pictorial stimuli were first presented to the participants, who were then instructed to describe the procedure in the first task, and to tell a story with a beginning, a middle, and an end in the second task looking at the stimuli. These two genres have been used in a large number of previous studies, which allows a greater comparability of the results with previous findings. These tasks were

selected instead of personal recount (also available in the corpus) due to the greater control imposed on the content of the elicited discourse samples.

Data analysis

Discourse segmentation—Each language sample was first segmented into elementary discourse units (EDUs), the minimal semantic building blocks of a discourse (Mann & Thompson, 1988). Discourse segmentation is known to be challenging for annotators, especially when it comes to spoken language (e.g. Artstein & Poesio, 2008). RST was initially devised for written language, and EDU were defined primarily based on syntactic criteria, that is, the presence of a predicate. However, segment boundaries in spontaneous discourse are often not as clearly defined as they are in written discourse or discourse, even more so in PWA's speech. In this study a combination of phonological (e.g. prosodic contours and pauses), syntactic, and semantic (e.g. semantic completeness) criteria was used. Specifically, three sets of detailed guidelines the segmentation process. They included the (1) RST annotation manual, developed and tested in the process of creating the RST Discourse Treebank, a corpus of newspaper articles annotated with RST (Carlson & Marcu, 2001), (2) CHAT transcription format manual (MacWhinney 2000, Electronic edition of 2011), and (3) guidelines for segmenting spoken language developed by Kibrik and Podlesskaya (2009), focusing on dysfluencies such as repetitions, reformulations, and self-interruptions. The resulting segmentation procedure was similar to the one used and described by Marini, Andreetta, del Tin, and Carlomagno (2011). First, syntactic clauses were identified where possible, syntactically incomplete clauses were segmented based on phonological boundaries and semantic completeness. The hard-to-define term “semantic completeness” refers to the relative understandability of a piece of discourse and does not necessarily imply well-formedness of the EDU. PWA's speech often lacks syntactic and prosodic indicators of segment boundaries, along with “semantic completeness” in the common sense. For these cases, we took speaker's perspective into account and relaxed the aforementioned criteria. Specifically, two types of malformed EDUs were introduced, and semantic understandability along with phonological indicators were attributed more weight than syntactic criteria in the process of segmentation. For example, a participant trying to describe a scene from the stimuli where mother gives her son an umbrella can only be able to utter “Mother (...) umbrella”. Such cases were marked as “incomplete” EDUs, as they indicate attempts to transfer a proposition. A further category of “failed EDU” was introduced if one or more obligatory syntactic constituents were omitted and if the general sense was lost; this included the span from the ending boundary of the previous segment to the beginning of the next one. These two kinds of EDU were considered to be disruptive for discourse structures. An “incomplete” EDU was defined as a content-wise comprehensible clause with an omission of a critical syntactic component, such as an object of a transitive verb (e.g. as a result of a word-finding difficulty), whereas incomprehensible collections of words were defined as “failed” EDUs (e.g. an output containing jargons or neologisms). Additionally, syntactically well-formed discourse units that were semantically out of place, or empty information-wise, were marked with a technical relation denoted with a question mark when linked to the rest of the structure.

Spontaneous story-telling is occasionally characterised by deviations from the main story line, for example, comments and embedded discourse units used to elaborate on parts of other EDUs. This type of constructions was labelled “structural expansions” in the analysis. Another group of phenomena intrinsic to spontaneous speech consists of reformulations, self-corrections, repetitions, and retracing, that is, returning to an earlier part of discourse to add information that could help a listener to understand the story and its details. These dysfluencies and repair strategies are more frequent in discourse of people with aphasia (e.g. MacWhinney et al., 2011; Marshall & Tompkins, 1982).

Discourse structure annotation—After the segmentation, all EDUs were incrementally linked to each other using a set of rhetorical relations following the guidelines of Carlson and Marcu (2001). Marcu’s modification of the RST Tool (O’Donnell, 1997; Marcu, Amorrortu, & Romera, 1999) was used to perform the annotations. Twenty-six different semantic relations (out of possible 78), such as background, consequence, condition, explanation, and means, were used to annotate the discourse samples in the current study. Examples for each relation are given in the Appendix A. Further details about the relation definitions and assignment can be found in RST Discourse Treebank annotation manual (Carlson & Marcu, 2001).

Reliability of EDU segmentation and discourse structure (RST) annotation

Discourse, like other smaller elements of natural language, such as words and sentences, is open to more than a single interpretation. Despite the detailed guidelines, the RST annotations bear a certain degree of subjectivity due to the possibility of multiple analyses. Analysis of the same discourse sample is expected to sometimes yield multiple resultant structures (Mann & Thompson, 1987; Taboada & Mann, 2006, Stede, 2008). Comparing RST annotations is not a straightforward procedure (Marcu, 2000; Irukieta, daCunha, & Taboada, 2014). To estimate the reliability of applying this RST technique to quantify oral discourse across raters, the segmentation was first verified. As the decisions were not independent, percent agreement on all the annotation decisions (instead of the F-measure, i.e. precision and recall) was then obtained. In the current study, all the 60 samples were first annotated by author WS. The annotations were then checked by author AL, with an agreement reaching 95% for the segmentation and 85% for the discourse structure annotation. The remaining 15% of structure annotations were discussed by all the authors and an agreement was reached on the analysis.

Ten percent of the samples from each group (equal percentage of story-telling and procedural transcripts) were randomly selected to establish the intra-rater reliability for EDU segmentation and annotation. These samples were re-segmented and re-annotated by author WS. The point-to-point agreement, calculated by the formula $[\text{total agreements}/(\text{total agreements} + \text{total disagreements}) \times 100]$, revealed a 95.7% agreement for EDU segmentation and 89.7% for RST relation annotations.

Analysing discourse organisation using RST

Each annotated discourse sample was analysed in terms of 13 parameters possibly contributing to coherence. The parameters can be divided into five groups related to certain

discourse properties. Specifically, Group I is related to speaker's efficiency in formulating complete discourse units, with respect to the possible effect of fatigue. They include (1) Fluency of production in the first half of the sample: total number of EDUs produced in the first half of the recording divided by the time elapsed in minutes, and (2) Fluency of production in the second half of the sample: same method of calculation in (1). Group II is related to the connectivity of semantic units within a discourse and discourse complexity. They include (3) Total number of EDUs, or total length of a discourse sample, (4) Relations set: the number of different types of rhetorical relations used in the sample, and (5) Relation type frequency, or the total number of rhetorical relations of different types used. Group III is related to the degree of elaboration and complexity of discourse. They include (6) Depth of the structure: the maximum distance on the graph between the top and the lowest node, or the number of levels the structure "branches" down. This parameter was used together with length as a correlate of structural complexity and elaboration, reflecting the degree of detail in the story, (7) Percentage of structural expansions: the number of embedded discourse units and comments divided by the total number of EDUs, and (8) Percentage of mononuclear relations: the number of mononuclear relations in the sample divided by the total number of mono- and multinuclear relations. Group IV is related to the degree of inadequacy, impairment, or structural disruption. They include (9) Percentage of incomplete EDUs, and (10) Percentage of failed EDUs. Finally, Group V included three final measures that capture the effect of micro-linguistic parameters on the overall discourse coherence. They include (11) Percentage of errors: number of all semantic and phonemic paraphasias, morphosyntactic errors, and neologisms divided by the total number of words, (12) Type-token ratio: This is the ratio of number of different words to total number of words, excluding repetition, retracing, self-correction, and false-start, and (13) Percentage of function words: It is the ratio of all closed-class words to the total number of words.

Coherence ratings

Twenty university students, all native speakers of Cantonese, were recruited to rate the 60 audio recordings as naïve listeners. The raters were divided into four groups and the sequence of presentation of the 60 audio files was randomised across groups. The raters were asked to indicate for each audio the following: (1) the level of understandability on a 9-point scale, with "1" representing the rater did not understand the content at all and "9" representing all content was understood, (2) the connectedness and completeness of the content with three options – 'complete', 'incomplete, but understandable', and 'incomplete and hard to follow', (3) if the events were presented in the correct order, and (4) whether there was a part of the story or procedure description that was hard to understand, and if so which part it was. Before the study, a 30-minute briefing and a short practice session were provided to each group of raters.

Analysis of information content

In the present study the proportion of main events was calculated to analyse informativeness of the samples. Thematic content was measured through the proportion of main events, or thematic units, that is, main ideas and details in a stimulus (Wright et al., 2005; Marini, Carlomagno, et al., 2005; Capilouto et al., 2006; Wright et al., 2011; Marini et al., 2011; Andreetta et al., 2012). In other words, a main event was defined as an independent piece of

content that was of sufficient degree of importance to the topic being conveyed. The procedure described in Capilouto et al. (2005) and Capilouto et al. (2006) was followed to perform the main events analysis, with the difference being that the NBD samples were used to determine the main events in the stimuli. The information – events for the narratives and steps for the procedures – present in at least 70% of the NBD transcripts was classified as essential. Other information included by participants was considered to be optional, providing additional elaborative components to the main story line or the main steps in the procedure. The original procedure by Capilouto et al. suggests evaluating speakers' ability to convey relationships between the events in addition to producing main concepts. However, speakers with aphasia often use syntactic structures of relatively reduced complexity (e.g. Edwards & Bastiaanse, 1998), which could affect their ability to explicitly express relations, but not necessarily be indicative of their poor understanding of these relations. In addition, the grammatical system of Cantonese differs from that of European languages, which makes the procedure less straightforward. Thus, in the present study, the focus of the main event analysis was on the information content, while the relational component was addressed in the discourse structure analysis. After the list of main events and elaborative components was formed based on the analysis of the NBD samples, and its final version was agreed upon by all the authors, their proportion in the PWA samples was calculated. Detailed information can be found in Appendix B. Point-to-point inter-rater agreement for the PWA samples reached at least 90%.

Statistical analysis

The normality of the residuals of the data obtained from the RST annotations was tested using Shapiro-Wilk test and through Q-Q plotting. Production fluency, number of types of relations, and type-token ratio (TTR) were found to be normally distributed ($p > .05$). In the NBD group a ceiling effect was observed on most of the measures; hence, non-parametric statistical analyses were implemented for these data.

To investigate the differences in performance of PWA and NBD speakers in the two genres, a set of two-way mixed ANOVAs was administered and subsequent t-tests were carried out for post-hoc analyses. For the non-normally distributed parameters, the Mann-Whitney test and the Wilcoxon signed-rank test were employed to study the group difference and the effect of elicitation task, respectively. An adjustment of significance level was done using Bonferroni's method due to multiple comparisons (0.05/8 or 0.00625).

In the rating task, outliers with a z-score of absolute value greater than two standard deviations from the group variable means (less than 5%) were removed from the data set. The Mann-Whitney test was used to compare the between-group difference for all the four ratings of coherence as the residuals were not normally distributed. Wilcoxon signed-rank test was used to examine the effect of genre on coherence ratings.

To study potential factors contributing to discourse coherence and its impairment in aphasia, the Kendall tau rank correlation coefficients between the coherence ratings and the 12 micro- and macro-linguistic parameters were calculated.

Result

RST analysis

Group differences—The descriptive statistics of linguistic measures for each speaker group across the two elicitation tasks is provided in Table I. Our results demonstrated that the NBD group performed better than the group with aphasia on all RST measures, except for the percentage of structural expansions, for both tasks (Table I).

The results of the two-way ANOVA and Mann-Whitney test revealed a main effect of speaker group on a number of quantitative measures (Table II). As indicated in Table I, the NBD group was significantly faster in producing EDUs (Fluency 1st and 2nd half) and used a greater variety of relations (Size of the relation set) to connect them than the PWA group. The distribution of RST relations of different types in the two speaker groups was comparable (Figure 2), suggesting that our group of PWA largely preserved the use of relations in a similar manner to their NBD counterparts.

Review of the raw data indicated that the NBD group used more relations of attribution (i.e. complex constructions with direct and indirect speech and cognitive predicates) as well as background, explanation, and elaboration relations than the PWA group. In both groups, but especially in the PWA group, in the story-telling, the most frequently used RST relation types were those expressing causality, while in procedural discourse, temporal relations predominated.

With reference to the discourse samples, a qualitative observation of the data indicated that the PWA's transcripts contained more reformulations and corrections. The PWA group also produced significantly fewer EDUs and more errors than the NBD group in both narrative ($U = 55.00$, $z = -2.39$, $p < .01$ and $U = 19.00$, $z = -4.15$, $p < .001$, respectively) and procedural discourse ($U = 50.50$, $z = -2.58$, $p < .01$ and $U = 32.00$, $z = -3.77$, $p < .001$, respectively).

Genre differences—There was a main effect of genre on the coherence measures (Table II). The story-telling task yielded faster discourse production in EDUs ($F(1,58)=28.5$, $p = .0002$; $F(1,58)=7.23$, $p=.01$) and a larger set of RST relations than the procedural discourse task ($F(1,58)=48.56$, $p < .0001$). In addition, there was an effect of genre on the type of structures used, namely, mono- versus multi-nuclear relations ($W=383.5$, $Z=3.106$, $p=.009$, $r=.4$). Procedural discourse contained a higher proportion of mono-nuclear relations, reflecting a difference in the macro-linguistic organisation in the two genres. Participants with aphasia also produced significantly more EDUs ($T = 4$, $z = -3.19$, $p < .001$) and a greater depth of resultant discourse structure ($T = 2$, $z = -3.19$, $p < .001$) in story-telling task than in procedural description.

Interaction effects—There was an interaction effect of group and genre on fluency (Table II and Figure 3).

Independent t-tests revealed that the non-brain-damaged participants produced EDUs significantly faster in both the first [$t(28) = -3.64$, $p < .001$] and second half [$t(28) = -4.63$,

$p < .001$] of the story-telling task than the PWA group. The effect of genre on TTR was stronger in the PWA group ($t(14) = -4.378, p < .001$), with a significantly higher TTR in procedural discourse than in the narratives.

Analysis of coherence ratings

A summary of descriptive statistics of the naïve listeners' subjective ratings of coherence is displayed in Table III. The PWA group's discourse was rated by naïve listeners with a lower understandability score. The results of a Mann-Whitney test revealed that the NBD group's discourse obtained a significantly higher understandability and clarity ratings than the group with aphasia on both discourse tasks (Table IV). Wilcoxon signed-rank test demonstrated that the effect of genre on the ratings was not significant in either speaker group. The PWA speakers also made more mistakes with the order of events in the storytelling, but not in procedural discourse.

Correlations between the linguistic measures and coherence ratings

RST—Most of the correlations between measures based on the RST analysis and subjective coherence ratings were significant at the .05 level or higher (Table V). Measures of structural expansion did not show significant correlations with the coherence rating; this is in contrast to the significant correlation found in the degree of elaboration, the size of the relations set used, and the total number of EDUs produced. There was an inverse correlation between the amount of failed and incomplete elementary discourse units and all four subjective ratings.

Micro-linguistic measures—Percentage of word-level errors was significantly correlated with all four coherence measures, especially clarity/understandability in both tasks. No significant correlations were found between TTR and percentage of function words, and the coherence ratings.

Information content—Fourteen of the 15 speakers with anomic aphasia mentioned all the essential steps, but not the two optional elements (elaborative materials). The situation was not as straightforward with the story-telling task. Many elaborative materials were omitted in the PWA transcripts, only half of the total amount of essential information was mentioned by the majority (80%) of the anomic speakers. The results of Mann-Whitney tests suggested that there was group effect ($U = 119, p < 0.001$ for story-telling; $U = 24.5, p < 0.05$ for procedural discourse). To summarise, NBD's samples were more informative than the ones produced by PWA, and the difference was more obvious in the story-telling task.

Discussion

The present study pioneered the application of a formal discourse analysis framework of Rhetorical Structure Theory (RST) to the investigation of coherence in spoken discourse in aphasia. Macro-linguistic properties of connected speech in 15 speakers with anomic aphasia and an equal number of non-brain-damaged (NBD) participants were examined. The analysis contrasted a number of parameters extracted from the RST annotation of discourse structure of 60 transcribed spoken samples and subjective coherence ratings of these samples in order to explore the contribution of macro-structural properties to coherence. To examine

genre-related differences in discourse coherence, two tasks were used, namely story-telling elicited by a series of pictures, and procedural description.

According to RST, discourse consists of elementary meaningful information units (or EDUs). Our group of participants with aphasia produced significantly fewer EDUs and with a lower efficiency (in EDU/min) than the NBD group. Discourse structure was constructed with a smaller variety of semantic, or rhetorical, relations in the narratives of PWA, which corresponded to significantly lower coherence ratings as judged by naïve native speakers. Despite the lack of difference in the depth of discourse structures by the two groups, there was a strong correlation between this variable and coherence ratings. Thus, the complexity of discourse, expressed by its length, depth, and the number of relations used to build it, was shown to be a factor influencing the perception of discourse coherence. Importantly, in contrast to previous studies, language complexity was measured at the discourse level, rather than the syntactic level (e.g. Ulatowska et al., 1981, 1983; Korpijaakko-Huuhka & Lind, 2012). Our results also demonstrated a certain association between micro-linguistic impairments and reduced coherence. The percentage of errors was higher in the PWA group and had an effect on the perception of the degree of coherence. On the other hand, it was found that there was little to no correlation between coherence ratings and other micro-linguistic parameters, lexical diversity (TTR) and percentage of function words. Calculating TTR for nouns and verbs separately in the future studies could provide more information on this issue, mainly because verb and noun production deficits have been argued to be dissociated in PWA (e.g. Luzzatti et al., 2002).

Morphosyntactic impairments, such as word-level errors and omissions, often lead to incomplete discourse units, in which one or more of the obligatory syntactic components are missing and even failed, or completely incomprehensible EDUs. Initially we hypothesised that micro-linguistic impairments do not directly contribute to poor coherence per se, but rather through their negative impact on discourse structure. Specifically, coherence ratings were correlated with the percentage of failed and incomplete EDUs. As expected, the speech of PWA contained significantly more reformulations, corrections, false-starts, and retracing. It has been noted that excessive use of repair strategies can result in an overall lack of clarity of speech (Tavakoli & Skehan, 2005). A post-hoc qualitative review of the raw data suggested that these strategies were often used to repair the discourse structure after failed EDUs. The results of this study suggest that discourse-level impairment is indeed the result of micro-linguistic impairments rather than the problem with information organisation and logical structuring of discourse. However, we argue that the observed tendencies may indicate a relationship between the quality of discourse structure and micro-linguistic variables exist, which makes it difficult to disentangle the effects between micro- and macro-linguistic variables on coherence.

The information content analysis demonstrated that essential information content in the story-telling was reduced in the PWA group. These results were not in line with the previous studies reporting preservation of essential information in PWA's narrative discourse (e.g. Ulatowska et al., 1983), but consistent with other findings reporting lower information content in narratives by people with aphasia (e.g. Capilouto, Wright, & Waganovich, 2006). Specifically, Andretta et al. (2012) reported that individuals with anomic aphasia had fewer

lexical information units, reduced utterance length, and more semantics errors; this reduction of sentence completeness as well as presence of lexical fillers and repetitions then led to impaired cohesion and coherence, respectively. Our present findings based on analyses involving RST confirms the view of Andreetta et al. (2012). Reduced informativeness corresponds to lower coherence ratings in the PWA group. Informativeness is thus potentially one of the factors influencing coherence perception. In this study, we had a homogeneous group of anomic participants, which means that the informativeness of their discourse is more or less similar, as are the underlying deficits. A group with more variability in terms of linguistic impairments, for example, different types of aphasia, would make it possible to tease apart the effect of informativeness and other factors. In other words, a group of participants with various aphasia types could be insightful in answering this question. Furthermore, our current findings were drawn from the use of only one single discourse sample for each genre. This inconsistency of producing information content across PWA can be attributed to the individual variation. Since it is reasonable to expect intra- and inter-participant variability in producing oral discourse, further extension of validating the application of RST to discourse analysis should expand the range and amount of genres.

As it was mentioned earlier, several factors had been shown to increase or decrease cognitive demand of an elicitation task (e.g. Coelho, Liles, & Duffy, 1995; Stine-Morrow, Miller, & Leno, 2001). The stimuli of the story-telling task visually depicted the temporal and logical sequences of the story event, whereas only a picture of the main ingredients for making a sandwich were provided in the procedural description task. The visual information about the temporal and logical relations present in the picture sequence could have contributed to the elicitation of a more structured and organised story (Wright & Capilouto, 2012a). On the contrary, the stimulus in the procedural description task did not necessarily facilitate discourse elicitation through a partially predetermined structure (Marini et al., 2005). Hence, a higher cognitive demand might have been expected of the procedural discourse task, which involved recalling of the order of steps, rather than describing a depicted sequence. Due to these considerations, telling a story in response to the picture sequence was expected to elicit a longer and a more elaborate narrative. Our results confirmed this tendency but further investigation in a separate study that carefully considers additional factors of relationships between (non-)animated characters or emotive reactions involved in time and/or cognitive load processing is warranted. In procedural discourse, participants with aphasia were almost equally as good as the NBD counterparts at preserving the correct order of events. This was potentially the result of a relatively flexible order of events in the chosen procedure. However, we also observed a qualitative difference in macro-linguistic organisation of discourse of the two genres. Specifically, procedural discourse is built with a larger proportion of mono-nuclear relations, in which every part of the relation pair contains equally important information. Our results quantitatively confirmed that there is a difference between macro-linguistic patterns of the two genres (e.g. Ulatowska et al., 1990). The RST analysis, as compared to a more traditional view of discourse organisation in terms of setting, complicating action, and resolution, delivered a more fine-grained representation of the macro-structural patterns of the two genres.

Another factor that could play a role in task performance is the effect of topic familiarity, as previous findings have suggested that discourse on familiar topics tends to be more elaborate

(Li, Williams, & Della Volpe, 1995). For example, according to Britton and Tesser's cognitive model (1982), prior knowledge can be conceptualised as a cognitive schema, which is activated when dealing with a more familiar topic. However, since the factor of content novelty to our speakers between the story-telling and procedural task was not controlled in this study. The results of the structural analysis that revealed a reduced ability to spontaneously formulate and elaborate on the story-telling output might, in fact, be more related to the more complex internal organisation of story-telling at the macro-structure level, as compared to procedural discourse. This observation seems to support previous reports demonstrating that performance on different discourse tasks may be influenced by the nature of the target material. Hence, the importance of careful selection of elicitation tasks should be highlighted in clinical or experimental settings.

The RST-based coherence analysis presented in this study has offered a novel, systematic and objective way of examining the macro-structural features of oral discourse in anomic aphasia. Specifically, it offers a hierarchical analysis of oral discourse, which is different from most reports of disordered or PWA discourse (see Linnik et al., 2016). This is evidenced by the strong correlations between the RST measures and the subjective coherence ratings of narrative production by speakers with anomic aphasia (Table V). The two discourse elicitation tasks investigated in the present study are also quite different from other published papers on PWA discourse production reviewed by Linnik et al. (2016). Future extension of the current investigation can involve a more diverse genre types elicited from a larger number of speakers with different types and severity of aphasia to determine whether coherence is or is not dependent on the micro-linguistic performance of a speaker. To our knowledge, apart from the work of Kibrik and Podlesskaya (2009) on spontaneous speech in children with neurosis, RST has not been applied to the investigation of clinical data before. Further RST-based analysis of discourse in other adult language-impaired populations, such as individuals with dementia, who have been reported to have problems with producing coherent discourse (e.g. Bourgeois & Hickey, 2009), can inform language production theory as well as therapy.

Limitations and future research

Two questions remained unanswered and will hopefully receive more attention in future work on coherence in aphasia, and in discourse in general. Firstly, our findings do not answer the question whether coherence is established through linguistic means only, or whether, for example, common ground (that is, the shared knowledge between the individual with aphasia and the person to whom they are telling the story) between participants and SLP as well as participants and raters are involved as well. Secondly, it is not entirely clear whether correlates such as understandability, completeness, and connectedness reflect the concept of coherence. Further studies building on the current work should address these issues, as well as extend the investigation to other spoken discourse genres. The limitations of this study include a relatively limited sample size restricted to a one type of aphasia. The choice of correlation analysis for this study, motivated by the ceiling effects and the number of observations, unlike regression analysis or similar methods, imposes certain constraints on the interpretation of our findings. In addition, variations in language production in different aphasia types could potentially provide additional information to complete the

picture drawn in this study. Thirdly, given RST is a new analysis within aphasia discourse measurement, it is important to establish the reliability or stability of the measures; however, the current study was based on an existing language corpus that does not contain participants' retest language samples across time. Further study should establish the test-retest reliability of employing RST to discourse analysis. This is particularly relevant when one attempts to make clinical decisions about PWA's discourse performance based on isolated evaluations (Boyle, 2014) because not all reported discourse measures are sufficiently stable across-time. Despite the above-mentioned limitations, the present findings provide a solid foundation for further studies investigating the contribution of linguistic parameters of discourse to its perceived coherence.

Conclusion

Our study has supported the argument that factors other than discourse organisation are critical to the perception of coherence. Word-level errors and reduced information content negatively impact discourse structure, leading to lower coherence ratings. The clinical implication of this conclusion is that discourse-level treatment may not be efficient or complete without additional therapy targeting production at the word level. Theoretical conclusions that we would like to draw from this work are the following. Adding to previous studies examining discourse production using multi-level approaches (e.g. Marini, Andreetta, del Tin, & Carlomagno, 2011; Sherratt, 2007; Wright & Capilouto, 2012a), our study demonstrated that it is a combination of macro- and micro-linguistic properties that makes discourse a coherent whole. We identified a number of linguistic parameters potentially required for a discourse to be coherent; these parameters included fluency and number of EDU, number of relation sets, depth of discourse structure, TTR, as well as percentage of error and failed EDU. By using ratings, coherence has been addressed as a perceived rather than a solely linguistic concept. Considering a listener's perspectives is an important step in the understanding of how coherence works.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study is supported by a grant funded by the National Institutes of Health (NIH-R01-DC010398) to Anthony Pak-Hin Kong (PI) and Sam-Po Law (Co-I), and the Erasmus Mundus Joint Doctorate Fellowship IDEALAB to Anastasia Linnik. Special thanks to the staff members in the following organizations (in alphabetical order) for their help in subject recruitment: Christian Family Service Center (Kwun Tong Community Rehabilitation Day Center), Community Rehabilitation Network of The Hong Kong Society for Rehabilitation, Internal Aphasia Clinic at the University of Hong Kong, Hong Kong Stroke Association, Lee Quo Wei Day Rehabilitation and Care Centre of The Hong Kong Society for Rehabilitation, Queen Mary Hospital, and Self Help Group for the Brain Damage. The authors would also like to acknowledge our subjects who participated, and Christy Lai, Lina Wong, and Winsy Wong for their assistance in data collection and processing.

References

Andreetta S, Cantagallo A, Marini A. Narrative discourse in anomic aphasia. *Neuropsychologia*. 2012; 50(8):1787–1793. DOI: 10.1016/j.neuropsychologia.2012.04.003 [PubMed: 22564448]

- Andreetta S, Marini A. Narrative assessment in patients with communicative disorders. *Travaux neuchâtelois de linguistique*. 2014; 60:69–84.
- Artstein R, Poesio M. Inter-coder agreement for computational linguistics. *Computational Linguistics*. 2008; 34(4):555–596.
- Bliss LS, McCabe A. Comparison of discourse genres: Clinical implications. *Contemporary Issues in Communication Science and Disorders*. 2006; 33(2):126–137.
- Bloom RL. Hemispheric responsibility and discourse production: Contrasting patients with unilateral left and right hemisphere damage. In: Bloom RL, Obler LK, De Santi S, Ehrlich JS, editors *Discourse analysis and applications: Studies in adult clinical populations* Hillsdale, NJ: Erlbaum; 1994 8194
- Bourgeois MS, Hickey EM. *Dementia: From diagnosis to management – a functional approach* New York, NY: Psychology Press; 2009
- Boyle M. Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*. 2014; 57:966–978. DOI: 10.1044/2014_JSLHR-L-13-0171
- Britton BK, Tesser A. Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of Verbal Learning and Verbal Behavior*. 1982; 21(4):421–436. DOI: 10.1016/S0022-5371(82)90709-5
- Capilouto G, Wright HH, Wagovich SA. CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*. 2005; 38(6):431–444. DOI: 10.1016/j.jcomdis.2005.03.005 [PubMed: 16199238]
- Carlson L, Marcu D. *Discourse tagging reference manual*. ISI Technical Report ISI-TR-545 2001 Retrieved from <http://www.isi.edu/~marcu/discourse>
- Christiansen JA. Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*. 1995; 51(2):291–317. [PubMed: 8564473]
- Chung KKH, Code C, Ball MJ. Lexical and non-lexical speech automatisms in aphasic Cantonese speakers. *Journal of Multilingual Communication Disorders*. 2004; 2(1):32–42. DOI: 10.1080/1476967031000154213
- Coelho C, Flewellyn L. Longitudinal assessment of coherence in an adult with fluent aphasia: A follow-up study. *Aphasiology*. 2003; 17(2):173–182. DOI: 10.1080/729255216
- Coelho C, Lê K, Mozeiko J, Hamilton M, Tyler E, Krueger F, Grafman J. Characterizing discourse deficits following penetrating head injury: A preliminary model. *American Journal of Speech-Language Pathology*. 2013; 22:S438–S448. DOI: 10.1044/1058-0360(2013)12-0076 [PubMed: 23695915]
- Coelho CA, Liles BZ, Duffy RJ. Impairments of discourse abilities and executive functions in traumatically brain-injured adults. *Brain Injury*. 1995; 9(5):471–477. [PubMed: 7550218]
- Cristea D, Ide N, Romary L. *Veins Theory: A Model of Global Discourse Cohesion and Coherence*. Proceedings of the 17th International Conference on Computational Linguistics-Volume. 1998; 1:281–285.
- da Cunha I, Torres-Moreno JM, Sierra G. On the development of the RST Spanish Treebank. Proceedings of the 5th Linguistic Annotation Workshop; Association for Computational Linguistics; 2011 Jun. 110
- Early EA, VanDemark AA. Aphasic speakers' use of definite and indefinite articles to mark given and new information in discourse. In: Brookshire RH, editor *Clinical Aphasiology Proceedings Vol. 15*. Minneapolis, MN: BRK Publishers; 1985 248254
- Edwards S, Bastiaanse R. Diversity in the lexical and syntactic abilities of fluent aphasic speakers. *Aphasiology*. 1998; 12(2):99–117.
- Fergadiotis G, Wright HH, Capilouto GJ. Productive vocabulary across discourse types. *Aphasiology*. 2011; 25(10):1261–1278. DOI: 10.1080/02687038.2011.606974 [PubMed: 22904592]
- Fergadiotis G, Wright HH. Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*. 2011; 25(11):1414–1430. [PubMed: 23125474]
- Fox B. *Discourse Structure and Anaphora* Cambridge: Cambridge University Press; 1986
- Givón T. Coherence in text vs. coherence in mind. In: Gernsbacher MA, Givón T, editors *Coherence in spontaneous text* John Benjamins Publishing; 1995 59115

- Gleason JB, Goodglass H, Obler L, Green E, Hyde MR, Weintraub S. Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech and Hearing Research*. 1980; 23(2):370–382. [PubMed: 7442197]
- Glosser G, Deser T. Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language*. 1991; 40(1):67–88. [PubMed: 2009448]
- Grosz BJ, Sidner CL. Attention, intentions, and the structure of discourse. *Computational Linguistics*. 1986; 12(3):175–204.
- Halliday MAK, Hasan R. *Cohesion in English* London: Longman; 1976
- Hobbs J. On the coherence and structure of discourse 1985 Retrieved from <http://www.isi.edu/~hobbs/ocsd.pdf>
- Iruskieta M, da Cunha I, Taboada M. A qualitative comparison method for rhetorical structures: identifying different discourse structures in multilingual corpora. *Language Resources and Evaluation*. 2014; :1–47. DOI: 10.1007/s10579-014-9271-6
- Kibrik AA, Podlesskaya VI. Night dream stories: A corpus study of spoken Russian discourse [Korpus ustnoj russkoj rechi “Rasskazy o snovidenijax”. *Jazyki slavjanskij kul'tur*; Moscow. 2009
- Kintsch W. Text comprehension, memory, and learning. *American Psychologist*. 1994; 49(4):294. [PubMed: 8203801]
- Kintsch W, Van Dijk TA. Toward a model of text comprehension and production. *Psychological Review*. 1978; 85(5):363–394.
- Kong APH, Law SP. The Cantonese AphasiaBank 20016 <http://www.speech.hku.hk/caphbank/>
- Korpijaakko-Huuhka AM, Lind M. The impact of aphasia on textual coherence: Evidence from two typologically different languages. *Journal of Interactional Research in Communication Disorders*. 2012; 3(1):47–70.
- Labov W. *Language in the Inner City* Philadelphia: University of Pennsylvania Press; 1972
- Lascarides A, Asher N. *Computing meaning* Springer; Netherlands: 2007 Segmented discourse representation theory: Dynamic semantics with discourse structure; 87124
- Li EC, Williams SE, Della Volpe A. The effects of topic and listener familiarity on discourse variables in procedural and narrative discourse tasks. *Journal of Communication Disorders*. 1995; 28(1):39–55. [PubMed: 7790537]
- Linnik A, Bastiaanse R, Höhle B. Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology*. 2016; 30(7):765–800.
- Longacre RE. *The grammar of discourse* New York: Plenum Press; 1996
- Luzzatti C, Raggi R, Zonca G, Pistarini C, Contardi A, Pinna GD. Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and language*. 2002; 81(1):432–444. DOI: 10.1006/brln.2001.2536 [PubMed: 12081411]
- MacWhinney B. *The CHILDES Project: Tools for analysing talk 3*. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 2000
- MacWhinney B, Fromm D, Forbes M, Holland A. AphasiaBank: Methods for studying discourse. *Aphasiology*. 2011; 25(11):1286–1307. DOI: 10.1080/02687038.2011.589893 [PubMed: 22923879]
- Mann WC, Thompson SA. Rhetorical structure theory: Toward a functional theory of text organization. *Text*. 1988; 8(3):243–281.
- Marcu D. Discourse trees are good indicators of importance in text. *Advances in automatic text summarization*. 1999:123–136.
- Marcu D. The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*. 2000; 26(3):395–448.
- Marcu D. Discourse structure: Trees or graphs? 2003 <http://www.isi.edu/~marcu/discourse/Discourse%20structures.htm>
- Marcu D, Amorrortu E, Romera M. Experiments in constructing a corpus of discourse trees. *Proceedings of the ACL'99 Workshop on Standards and Tools for Discourse Tagging*; 1999 4857
- Marini A, Andreetta S, del Tin S, Carlomagno S. A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*. 2011; 25(11):1372–1392.

- Marini A, Boewe A, Caltagirone C, Carlomagno S. Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research*. 2005; 34(5):439–463. DOI: 10.1007/s10936-005-6203-z [PubMed: 16177935]
- Marini A, Fabbro F. Psycholinguistic models of speech production in bilingualism and multilingualism. In: Ardila A, Ramos E, editors *Speech and language disorders in Bilinguals* New York: Nova Science Publishers; 2007 4767
- Marshall RC, Tompkins CA. Verbal self-correction behaviors of fluent and nonfluent aphasic subjects. *Brain and Language*. 1982; 15:292–306. DOI: 10.1016/0093-934X(82)90061-X [PubMed: 7074346]
- Moore JD, Pollack ME. A problem for RST: The need for multiple-level discourse analysis. *Computational Linguistics*. 1992; 18(Appelt 1985):537–544.
- Moser M, Moore JD. Towards a synthesis of two accounts of discourse structure. *Computational Linguistics*. 1996; 22(1992):410–419.
- O'Brien EJ, Albrecht JE. Updating a mental model: Maintaining both local and global coherence. *Journal of experimental psychology: Learning, memory, and cognition*. 1993; 19(5):1061–1070. DOI: 10.1037/0278-7393.19.5.1061
- O'Donnell M. RST-Tool: An RST Analysis Tool. *Proceedings of the 6th European Workshop on Natural Language Generation*; Duisburg, Germany: Gerhard-Mercator University; 1997
- Olness GS. Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*. 2006; 20:175–187. DOI: 10.1080/02687030500472710
- Olness GS, Ulatowska HK. Personal narratives in aphasia: Coherence in the context of use. *Aphasiology*. 2011; 25(August 2012):1393–1413. DOI: 10.1080/02687038.2011.599365
- Packard JL. *The morphology of Chinese: A linguistic and cognitive approach* Cambridge University Press; 2000
- Passonneau RJ, Litman DJ. Discourse segmentation by human and automated means. *Computational Linguistics*. 1997; 23(1):103–139.
- Prasad R, Dinesh N, Lee A, Miltsakaki E, Robaldo L, Joshi A, Webber B. The Penn Discourse TreeBank 2.0. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*; 2008 14 doi:10.1.1.165.9566
- Pritchard M, Dipper L, Morgan G, Cocks N. Language and iconic gesture use in procedural discourse by speakers with aphasia. *Aphasiology*. 2015; 29(7):826–844. [PubMed: 25999636]
- Saffran EM, Berndt RS, Schwartz MF. The quantitative analysis of agrammatic production: Procedure and data. *Brain and language*. 1989; 37(3):440–479. [PubMed: 2804622]
- Sherratt S. Multi-level discourse analysis: A feasible approach. *Aphasiology*. 2007; 21(3–4):375–393.
- Stede M. Subordination versus Coordination in Sentence and Text 2008 RST revisited: Disentangling nuclearity; 3358
- Stine-Morrow EAL, Miller LMS, Leno R. Patterns of on-line resource allocation to narrative text by younger and older readers. *Aging, Neuropsychology, and Cognition*. 2001; 8(1):36–53.
- Taboada MT. *Building coherence and cohesion: Task-oriented dialogue in English and Spanish* Vol. 129. John Benjamins Publishing; 2004
- Taboada M. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*. 2006; 38(4):567–592. DOI: 10.1016/j.pragma.2005.09.010
- Taboada M, Mann WC. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies*. 2006; 8:423–459. DOI: 10.1177/1461445606061881
- Taboada M, Mann WC. Applications of Rhetorical Structure Theory. *Discourse Studies*. 2006; 8:567–588. DOI: 10.1177/1461445606064836
- Tavakoli P, Skehan P. Strategic planning, task structure, and performance testing. In: Ellis R, editor *Planning and task performance in a second language* Amsterdam: Benjamins; 2005 239277
- Thompson CK, Shapiro LP, Li L, Schendel L. Analysis of verbs and verb-argument structure: A method for quantification of aphasic language production. *Clinical Aphasiology*. 1995; 23(1):121–140.

- Ulatowska HK, Reyes B, Santos TO, Garst D, Vernon J, McArthur J. Personal narratives in aphasia: Understanding narrative competence. *Topics in stroke rehabilitation*. 2013; 20(1):36–43. [PubMed: 23340069]
- Ulatowska H, Streit Olness G, Wertz R, Samson A, Keebler M, Goins K. Relationship between discourse and Western Aphasia Battery performance in African Americans with aphasia. *Aphasiology*. 2003; 17(March 2015):511–521. DOI: 10.1080/0268703034400102
- Ulatowska HK, North AJ, Macaluso-Haynes S. Production of narrative and procedural discourse in aphasia. *Brain and Language*. 1981; 13(2):345–371. [PubMed: 7260577]
- Ulatowska HK, Weiss-Doyell A, Freedman-Stern R, Macaluso-Haynes S. Production of narrative discourse in aphasia. *Brain and Language*. 1983; 19(2):317–334. DOI: 10.1016/0093-934X(83)90074-3 [PubMed: 6192869]
- Ulatowska HK, , Allard L, , Chapman SB. *Discourse ability and brain damage* Springer; New York: 1990 *Narrative and procedural discourse in aphasia*; 180198
- Van Dijk TA. *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition* Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1980
- Vermeulen J, Bastiaanse R, Van Wageningen B. Spontaneous speech in aphasia: A correlational study. *Brain and Language*. 1989; 36(2):252–274. [PubMed: 2465808]
- Walker MA. Limited attention and discourse structure. *Computational Linguistics*. 1996; 22(2):255–264.
- Wang WSY, , Sun C. *The Oxford Handbook of Chinese Linguistics* Oxford University Press; 2015
- Wolf F, Gibson E. Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics*. 2005; 31(October 2004):249–287. DOI: 10.1162/0891201054223977
- Wright H, Capilouto GJ. Manipulating task instructions to change narrative discourse performance. *Aphasiology*. 2009; 23:1295–1308.
- Wright H, Capilouto GJ. Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology*. 2012a; 26(5):656–672. DOI: 10.1080/02687038.2012.676855 [PubMed: 22962509]
- Wright HH, , Capilouto GJ. Considering cognition, age, and discourse task on global coherence ability; *Cognitive Aging Conference*; Atlanta, GA. 2012b
- Yiu EML. Linguistic assessment of Chinese-speaking aphasics: Development of a Cantonese aphasia battery. *Journal of Neurolinguistic*. 1992; 7(4):379–424. DOI: 10.1016/0911-6044(92)90025-R
- Yiu EML, Worrall L. Patterns of grammatical disruption in Cantonese aphasic subjects. *Asia Pacific Journal of Speech, Language and Hearing*. 1995; 1(2):105–126.

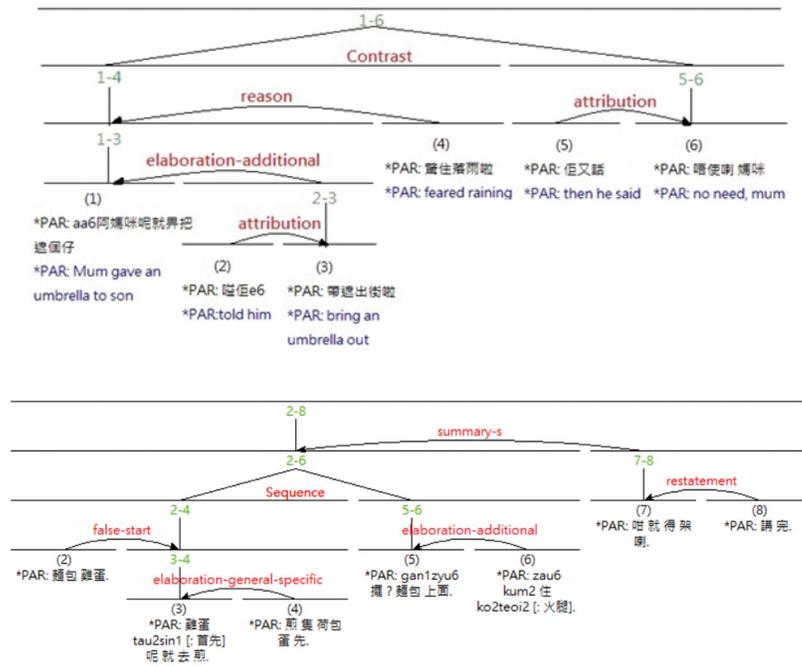


Figure 1. Example of the Rhetorical Structure Theory (RST) analysis in Cantonese transcripts of (a) a control participant (PAR) in the story-telling task of the refused umbrella story and (b) a participant (PAR) with aphasia in the procedural description task of making an egg and ham sandwich. Each numbered utterance represents an Elementary Discourse Unit (EDU), which is the smallest semantic entity with specific syntactic and phonological criteria. A text is first segmented into EDUs. The hierarchical and connected structure among EDUs in a text is then formed.

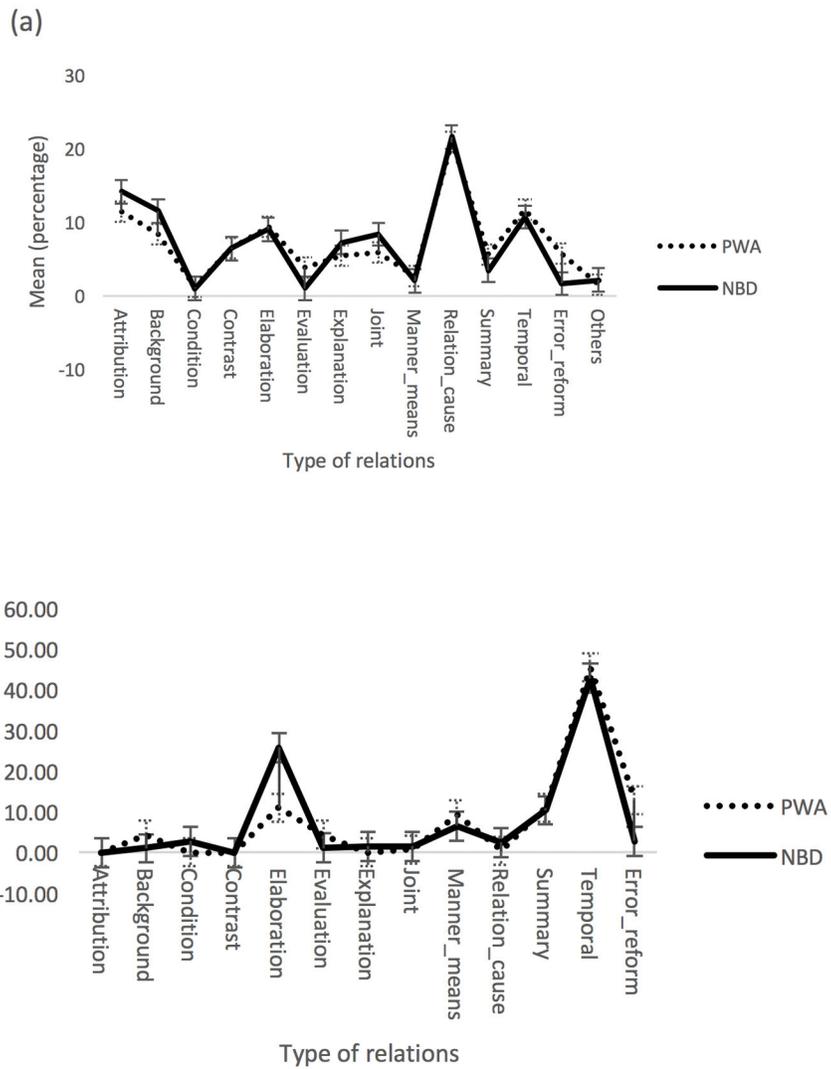


Figure 2. Distribution of Rhetorical Structure Theory (RST) relations in the (a) story-telling and (b) procedural description tasks for persons with aphasia (PWA) and non-brain-damaged (NBD) participants

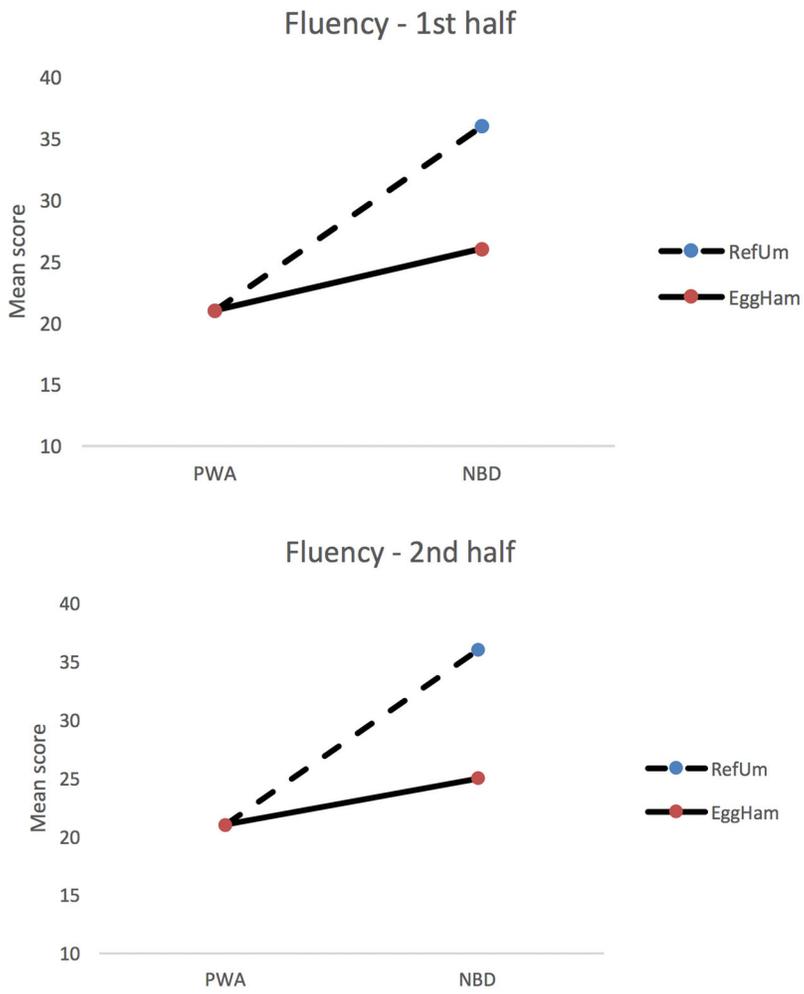


Figure 3. Interaction between genre type and speaker group on fluency of Elementary Discourse Unit (EDU), which is the smallest semantic entity with specific syntactic and phonological criteria, production in the (a) first and (b) second half of the sample in the story-telling task of the refused umbrella story (RefUm) and procedural description task of making an egg and ham sandwich (EggHam).

Table 1
Descriptive statistics of linguistic measures based on Rhetorical Structure Theory (RST) relations

	Anomic aphasia		Non-brain-damaged		Anomic aphasia		Non-brain-damaged	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
	Story-telling task				Procedural discourse task			
RST Measures								
Fluency-1 st half	22.6 (13.2)	7.5–54.0	38.0 (9.6)	21.4–60.0	16.2 (8.0)	4.6–30.0	23.3 (6.4)	15.0–36.0
Fluency-2 nd half	22.1 (10.2)	6.8–42.9	36.8 (6.8)	21.4–50.0	21.4 (8.2)	4.3–35.0	28.4 (7.7)	16.9–50.0
Total # of EDUs	14.6 (5.2)	7.0–26.0	19.2 (4.7)	10.0–26.0	8.4 (4.1)	4.0–19.0	14.0 (6.3)	7.0–23.0
Size of the relation set	11.5 (5.8)	4.0–24.0	17.7 (4.5)	9.0–24.0	4.8 (2.2)	1.0–8.0	8.8 (4.7)	3.0–16.0
Relation type frequency	8.4 (4.1)	3.0–15.0	10.9 (1.9)	7.0–13.0	3.6 (1.5)	1.0–6.0	5.5 (3.2)	2.0–11.0
Depth of discourse structure	5.3 (1.8)	2.0–8.0	6.8 (1.4)	5.0–10.0	2.8 (1.0)	1.0–4.0	4.8 (2.5)	2.0–9.0
% of structural expansions	1.8 (3.9)	0.0–12.5	0.7 (1.8)	0.0–5.0	3.0 (6.5)	0.0–20.0	1.1 (2.2)	0.0–6.3
% of mononuclear relations	74.3 (11.7)	54.6–100.0	75.1 (11.0)	44.4–90.5	56.8 (25.6)	0.0–87.5	60.3 (22.3)	20.0–90.9
% of incomplete EDUs	5.4 (10.3)	0.0–36.4	1.6 (3.3)	0.0–11.8	11.3 (14.2)	0.0–50.0	2.9 (5.1)	0.0–14.3
% of failed EDUs	6.5 (19.3)	0.0–75.0	0.0 (0.0)	0.0–0.0	4.8 (10.3)	0.0–33.3	0.0 (0.0)	0.0–0.0
Micro-linguistic Measures								
% of errors	4.9 (4.4)	0.0–15.4	0.2 (0.6)	0.0–1.7	6.7 (5.8)	0.0–18.2	0.2 (0.7)	0.0–2.5
Type-token ratio	0.6 (0.1)	0.4–0.8	0.6 (0.1)	0.5–0.7	0.7 (0.1)	0.6–0.9	0.6 (0.1)	0.5–0.9
% of function words	48.3 (8.0)	34.0–58.7	52.3 (4.8)	47.5–64.3	50.2 (8.5)	32.0–64.4	53.8 (6.4)	40.7–64.7

Table II

Statistical comparisons between performances of persons with aphasia and non-brain-damaged speakers

	Effect of group	Effect of genre	Interaction effect
	Two way ANOVA, F(1, 58)		ANOVA, F(3,56)
Fluency-1st half	14.81 ***	28.50 ***	258.30 *
Fluency-2nd half	18.15 ***	7.23 *	220.19 *
Total # of EDUs	11.74 **	14.53 **	10.97, p=0.73
Size of the relation set	5.554 *	48.56 ***	22.33, p=0.66
Depth of discourse structure	11.62 **	21.09 ***	14.45, p=0.59
Type-token ratio	8.124 *	3.295, p=0.07	4.85, p=0.15
	Mann-Whitney W	Wilcoxon Signed-Rank V	---
% of incomplete EDUs	340, p=0.17	25, p=0.112	---
% of failed EDUs	330 *	13, p=0.67	---
% of structural expansions	423, p=0.76	13, p=0.57	---
% of errors	105 ***	37, p=0.34	---
% of mononuclear relations	509.5, p=0.76	383.5 **	---

Note:

*
p < .05;**
p .01;***
p .001.

P-values were adjusted using the Holm-Bonferroni method. EDU = Elementary Discourse Unit.

Table III

Descriptive statistics of coherence ratings by naïve listeners

	Anomic aphasia		Non-brain-damaged	
	Mean (SD)	Range	Mean (SD)	Range
Story-telling task				
Understandability (1–9)	4.6 (1.6)	1.4–6.7	8.0 (0.5)	7.2–9.0
Completeness (%)	42.7 (28.2)	5.0–95.0	91.0 (14.2)	50.0–100.0
Order of events (%)	86.7 (19.5)	25.0–100.0	98.7 (3.0)	90.0–100.0
Disruptions of clarity (%)	65.0 (29.3)	10.0–100.0	9.4 (10.2)	0.0–30.0
Procedural discourse task				
Coherence rating (from 1–9)	4.7 (1.7)	2.05–6.6	7.3 (0.9)	6.3–9.30
Completeness (%)	50.7 (34.9)	0.0–90.0	79.2 (22.0)	25.0–100.0
Order of events (%)	81.6 (14.5)	60.0–100.0	94.9 (7.3)	73.7–100.0
Disruptions of clarity (%)	55.0 (28.6)	5.0–100.0	15.2 (16.3)	0.0–60.0

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table IV

Comparisons of speaker groups in terms of perceptual judgment by naïve listeners

	Mann-Whitney U			
	Understandability	Completeness	Order of events	Disruptions of clarity
Story-telling task	0.0 *	11.5 *	34.5 *	12.0 *
Procedural discourse task	8.5 *	51.5	50.0 *	26.0 *

Note:

*
p 0.00625

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Correlation between macro- and micro-linguistic parameters and subjective coherence ratings

Table V

	Story-telling task						Procedural discourse task					
	Understand-ability	Complete-ness	Order of events	Disruptions of clarity	Understand-ability	Complete-ness	Order of events	Disruptions of clarity	Understand-ability	Complete-ness	Order of events	Disruptions of clarity
Fluency-1 st half	0.43 ^{***}	0.41 ^{**}	0.30 [*]	-0.43 ^{***}	0.26 [*]	0.10	0.13	-0.20	0.16	0.21	0.15	-0.15
Fluency-2 nd half	0.54 ^{***}	0.49 ^{***}	0.38 ^{**}	-0.48 ^{***}	0.36 ^{**}	0.31 [*]	0.20	-0.13	0.25	0.27 [*]	0.13	-0.09
Total # of EDUs	0.47 ^{***}	0.48 ^{***}	0.29 [*]	-0.31 [*]	0.34 ^{**}	0.30 [*]	0.17	-0.14	0.43 ^{**}	0.36 ^{**}	0.21	-0.29 [*]
# of types of relation	0.56 ^{***}	0.51 ^{***}	0.37 [*]	-0.40 ^{**}	0.42 ^{**}	0.30 [*]	0.17	-0.14	0.43 ^{**}	0.36 ^{**}	0.21	-0.29 [*]
Size of the relation set	0.57 ^{***}	0.58 ^{***}	0.42 ^{**}	-0.39 ^{**}	0.42 ^{**}	0.30 [*]	0.17	-0.14	0.43 ^{**}	0.36 ^{**}	0.21	-0.29 [*]
Depth of the resulting discourse structure	0.47 ^{***}	-0.45 ^{***}	-0.36 [*]	-0.30 [*]	0.43 ^{**}	0.36 ^{**}	0.21	-0.29 [*]	0.43 ^{**}	0.36 ^{**}	0.21	-0.29 [*]
% of structural expansions	-0.02	0.01	0.07	-0.08	-0.02	-0.11	-0.06	0.10	-0.02	-0.11	-0.06	0.10
% of incomplete EDUs	-0.03	0.02	0.03	0.16	-0.45 ^{***}	-0.29 [*]	-0.40 ^{**}	0.52 ^{***}	-0.45 ^{***}	-0.29 [*]	-0.40 ^{**}	0.52 ^{***}
% of failed EDUs	-0.39 ^{**}	-0.36 [*]	-0.49 ^{**}	0.40 ^{**}	-0.37 [*]	-0.30 [*]	-0.32 [*]	0.40 ^{**}	-0.37 [*]	-0.30 [*]	-0.32 [*]	0.40 ^{**}
% of errors	-0.71 ^{***}	-0.73 ^{***}	-0.46 ^{**}	0.56 ^{***}	-0.62 ^{***}	-0.46 ^{**}	-0.38 [*]	0.56 ^{***}	-0.62 ^{***}	-0.46 ^{**}	-0.38 [*]	0.56 ^{***}
Type-token ratio	-0.21	-0.28 [*]	-0.02	0.19	-0.26 [*]	-0.15	-0.05	0.08	-0.26 [*]	-0.15	-0.05	0.08
% of function words	0.19	0.10	0.05	-0.06	0.24	0.16	0.12	-0.13	0.24	0.16	0.12	-0.13

Note:

* p < .05;

** p < .01;

*** p < .001.