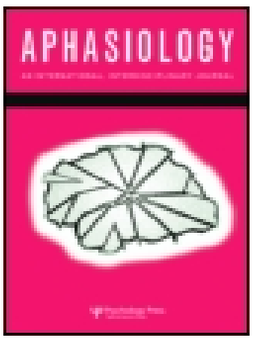

Aphasiology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/paph20

Linguistic mechanisms of coherence in aphasic and non-aphasic discourse

Anastasia Linnik , Roelien Bastiaanse , Manfred Stede & Mariya Khudyakova

To cite this article: Anastasia Linnik , Roelien Bastiaanse , Manfred Stede & Mariya Khudyakova (2021): Linguistic mechanisms of coherence in aphasic and non-aphasic discourse, Aphasiology, DOI: 10.1080/02687038.2020.1852527

To link to this article: https://doi.org/10.1080/02687038.2020.1852527

APHASIOLOGY
https://doi.org/10.1080/02687038.2020.1852527

Routledge
Taylor & Francis Group

Linguistic mechanisms of coherence in aphasic and non-aphasic discourse

Anastasia Linnik[a], Roelien Bastiaanse[b,c], Manfred Stede[d] and Mariya Khudyakova[b]

[a]International Doctorate for Experimental Approaches to Language and Brain (IDEALAB, (IDEALAB, Universities of Trento, Groningen, Potsdam, Newcastle and Macquarie University), Potsdam, Germany; [b]Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia; [c]Center for Language and Cognition Groningen (CLCG), University of Groningen, Groningen, The Netherlands; [d]Applied Computational Linguistics, University of Potsdam, Potsdam, Germany

ABSTRACT

Background: Coherence is the quality that distinguishes discourse from a random collection of sentences. People with aphasia have been reported to produce less-coherent discourse than non-language-impaired speakers. It is largely unclear how coherence is established in natural language and what leads to its impairment in aphasia.

Aims: This paper presents a cross-methodological investigation on coherence in the discourse of Russian native speakers with and without aphasia. The purpose of this study was to examine the connection between language impairments in aphasia and different aspects of discourse coherence in order to determine the linguistic mechanisms that could be involved in establishing and maintaining it.

Methods & Procedures: Coherence was operationalised as a combination of four aspects: informativeness, clarity, connectedness, and understandability. Twenty participants were asked to retell the content of a short movie. The retellings were annotated using Rhetorical Structure Theory (RST), a formalistic framework for discourse-structure analysis. Next, they were evaluated for coherence on a four-point scale by trained raters. The ratings were compared between groups. A classification analysis was performed to determine whether the ratings could be predicted based on the macrolinguistic variables collected from the RST annotations and several microlinguistic variables previously linked to coherence.

Results: Retellings produced by speakers with aphasia received lower ratings than those of control participants on all aspects of coherence. The results indicate that different combinations of microlinguistic and discourse-structure variables play a role in establishing each of the coherence aspects.

Conclusions: Our results provided supporting evidence on coherence impairment in aphasia. Perception of a discourse as more or less coherent was associated with both microlinguistic and macrolinguistic variables, with different combinations of variables relevant for each of the aspects. Furthermore, we found that discourse structure plays an important role, especially for understandability. We speculate that pragmatic knowledge shared by interlocutors may boost the coherence of aphasic discourse.

ARTICLE HISTORY
Received 2 December 2019
Accepted 15 November 2020

KEYWORDS
Aphasia; coherence; discourse production; discourse structure; rhetorical structure theory

CONTACT Mariya Khudyakova mariya.kh@gmail.com Center for Language and Brain, National Research University Higher School of Economics, 3 Krivokolenny Pereulok, Room 3-308,101000 Moscow, Russia

Introduction

Coherence is an intrinsic property of discourse reflecting its semantic and pragmatic unity. It is vital for communication, as it allows a listener to reconstruct discourse as the speaker had it in mind, or the mental representation of it (Gernsbacher & Givon, 1995). Although language production in aphasia at the discourse level has received considerable attention, the mechanisms underlying coherence remain relatively poorly understood. While some studies reported discourse coherence to be impaired in aphasia, others found it to be relatively well preserved (cf. reviews by Armstrong, 2000; Ellis et al., 2016; Pritchard et al., 2017). The existing disparity in results has been attributed, among other reasons, to different frameworks and definitions used by researchers, as well as different language profiles of PWA (Linnik et al., 2016).

Previously used definitions of coherence centred around the notions of thematic relatedness and topic maintenance (Agar & Hobbs, 1982; Glosser & Deser, 1991, Wright & Capilouto, 2012), a more general notion of semantic and pragmatic unity, or the degree to which discourse as a whole makes sense (Olness & Ulatowska, 2011; Ulatowska et al., 2004). Other studies addressed coherence as the well-formedness in terms of propositional content of discourse (e.g., Andreetta et al., 2012; Christiansen, 1995; H.K. Ulatowska et al., 1981; Marini et al., 2011; Ulatowska et al., 1983). It also has become common to distinguish between local coherence, occurring between consecutive sentences or propositions, and global coherence, referring to the semantic unity of a discourse as a whole (e.g., Ellis et al., 2016; Pritchard et al., 2017).

It has been argued that a single operationalisation of coherence was not well suited to reflect the complex phenomenon behind it (e.g., Linnik et al., 2016; Olness et al., 2005). In the present study, coherence is operationalised as a complex phenomenon combining different local and global aspects of discourse production. Specifically, we consider coherence to be established through *informativeness, clarity, understandability*, and *connectedness*. The four aspects refer to the definitions and operationalisations of coherence which have been introduced in previous research. However, while each of them has been studied before, they have not been addressed together as different aspects of the same phenomenon.

- *Informativeness* represents the amount and contextual appropriateness, or relevance, of information, or propositional content, in a discourse (e.g., Gleason et al., 1980; H.K. Ulatowska et al., 1981).
- *Clarity* refers to the semantic unity of discourse reached through the organisation of its parts with respect to the overall purpose, goal, or topic (e.g., Olness & Ulatowska, 2011; Andreetta & Marini, 2015; Christiansen, 1995; Glosser & Deser, 1991; Marini et al., 2011; Wright & Capilouto, 2012).
- *Understandability* refers to the subjective evaluation of communicative success achieved by the act of producing discourse. The concept is related to that of "functional communication" (Holland, 1982), in that the focus is not on linguistic performance in isolation, but on language in use, as one of the foremost means of communication. Communicative success certainly does not rely on linguistic performance alone. However, for H.K. Ulatowska et al. (1981), (1983)) include such evaluation questions as "Do you know what is happening in the story?" and "Do you as a

listener know what the subject is talking about?" in their coherence evaluation, and the study by Olness et al. (2005) addressed overall narrative quality in aphasia. The general idea is that interlocutors could rely on pragmatics (common ground, context) in the presence of language impairments to achieve coherence. *Understandability* aims to take into consideration the interactive nature of coherence.
- *Connectedness* reflects the local semantic and pragmatic connections between elements of discourse (Andreetta et al., 2012; Glosser & Deser, 1991; 1991).

As operationalisations varied, so did the methods used to study coherence and the outcomes. Holland (1982) observed that speakers with aphasia retained the ability to successfully communicate their meaning in spite of various language deficits. Similar results were reported in a study on conversational discourse in aphasia conducted by Ulatowska et al. (1992). However, Armstrong (1987) investigated listener ability to "make sense" of discourse and reported low rating scores for the narratives produced by PWA compared to those of non-brain-damaged (NBD) speakers. Ulatowska and colleagues used a similar scale to assess overall clarity and content, and while the narratives in the aphasia group received lower scores than those in the control group on average, the discourse by PWA was found to be well-formed in terms of superstructure (H.K. Ulatowska et al., 1981; Ulatowska et al., 1983).

Glosser and Deser (1991) developed a rating scale to evaluate the relatedness of every utterance to the content of the preceding one (local coherence) and to the overall topic (global coherence) and reported comparable performance for their PWA and NBD groups. Conversely, several studies using similar rating scales reported low scores in their aphasia groups (e.g., Coelho & Flewellyn, 2003; Wright & Capilouto, 2012). A number of investigations using methods other than rating scales produced further evidence suggesting that coherence was impaired in aphasia. For example, Christiansen (1995) demonstrated that speakers with fluent and non-fluent aphasia produced different patterns of propositional coherence violations (information gaps, progression, and relevance). Andreetta et al. (2012) found a difference in global-coherence error rate between PWA with anomia and NBD participants, but not in the local-coherence error rate.

The connection between different linguistic variables and coherence has been addressed in several previous works on discourse production in aphasia. Comprehensive multi-level procedures combining word- and sentence-level measures, information content, fluency, global and/or local coherence assessment have been developed, for example, by Andreetta and colleagues (Andreetta et al., 2012; Andreetta & Marini, 2015; Marini et al., 2011) and Wright and Capilouto (2012). Both groups found a connection between lexical-retrieval deficits (number of lexical information units and lexical diversity) and global-coherence scores in people with different types of aphasia. In the present study, we examine different aspects of coherence in connection with various linguistic parameters in order to obtain a comprehensive picture of discourse coherence in aphasia.

In the research on discourse of NBD individuals, several frameworks have been developed based on the idea that discourse coherence is established through its internal organization, or structure (e.g., Grosz & Sidner, 1986; Taboada & Mann, 2006b; Wolf & Gibson, 2005). Discourse structure is constructed of elementary discourse units (EDU), such as clauses or utterances, and the semantic and pragmatic relations connecting them

(e.g., causal, elaborative, or contrastive), also called "discourse", "coherence", or "rhetorical" relations. In Rhetorical Structure Theory (RST), one of the formal frameworks commonly used for discourse-structure analysis, the structure is represented as a tree, each element of which belongs with the rest and can be meaningfully connected to another element or a larger span (Mann & Thompson, 1988; Taboada & Mann, 2006a; 2006b). While RST is a powerful tool for discourse analysis that has been widely applied to natural-language data of different modalities and genres, very few studies so far have investigated the differences in rhetorical structures of narratives produced by neurologically healthy and clinical populations (Kong et al., 2018 for aphasia, Abdalla et al., 2018 for Alzheimer's disease, Kibrik & Podlesskaja, 2009 for neurosis).

In this paper, we present a further investigation of the relationship between the perception of coherence and linguistic well-formedness of discourse, both on microlinguistic and macrolinguistic levels. In previous research, the term "coherence" was used to refer to different aspects of language production (e.g., Linnik et al., 2016). Building on this observation, we accept that coherence is a complex phenomenon and consider these different aspects as its constituting parts. Our initial assumption is that narratives produced by PWA differ from those of NBD speakers in different aspects of coherence depending on specific language impairments speakers present with. The role of discourse structure was of particular interest, as it has been emphasised in research on discourse by non-language-impaired speakers, but has scarcely been studied in the context of language production in aphasia. The two research questions raised in this study thus are (RQ1) whether retellings produced by PWA differ on the four aforementioned aspects of coherence from those produced by non-brain-damaged participants, and if so, (RQ2) whether there is a connection between these aspects and linguistic variables previously linked to coherence.

We considered a set of variables reflecting various linguistic parameters of discourse, which were previously demonstrated to be related to coherence, such as lexical diversity, propositional content, and percentage of morphosyntactic errors (e.g., Marini et al., 2011; Wright & Capilouto, 2012). Next, we enriched the set with a number of macrolinguistic variables based on RST annotations in order to investigate the relationship between discourse structure and coherence. Discourse structure has been considered in very few studies on speech production in aphasia, mainly for selected genres and at higher levels of language organization, such as story grammar and episode structure (e.g., Coelho et al., 1994; Olness & Ulatowska, 2011; Ulatowska et al., 1983). It could, however, "be utilized to evaluate multiple levels of language" (Ulatowska et al., 1999, p. 4). We selected RST, as it provides a possibility to analyse discourse structure from the local between-clause level to the higher-level organization of larger episodes.

Methods

Participants

Ten people with aphasia resulting from stroke (mean age 56.4 years, range 40–73 years; 6 male and 4 female) and 10 non-brain-damaged participants (mean age 58.7 years, range 42–84 years; 5 male and 5 female), all native Russian speakers, participated in the study. Participants with aphasia were recruited at the Centre for Speech Pathology and

Neurorehabilitation in Moscow, Russia. Five of the participants with aphasia were diagnosed with fluent, five others with non-fluent aphasia using Luria's Neuropsychological Investigation (Akhutina, 2016; Luria, 1966), the severity level varying from mild to severe and time post-onset ranging from 3 months to 7 years and 9 months. Participants without brain damage reported no history of neurological impairments or psychiatric disorders. Further information about participants is provided in Tables A1 and A2 of Appendix A. The participants signed an informed consent form that was formulated according to the Declaration of Helsinki.

Elicitation stimulus and procedure

The *Pear Film* (Chafe, 1980), a six-minute speechless movie, was used to elicit spoken discourse samples. Following Chafe's original procedure, the participants were instructed to watch the movie and retell the story to someone who had not seen the movie before. Audio and/or video recordings of the retellings made with the consent of the participants were added to the Russian Clinical Pear Stories Corpus (Russian CliPS; Khudyakova et al., 2016).

Linguistic analysis

Linguistic analysis of the narratives was performed by two of the authors (AL and MK) unless specified otherwise. The linguistic variables used in the analysis are listed in Table 1. Information on inter-rater agreements is presented in Table B1 of Appendix B.

Coherence evaluation

Coherence was operationalised through the ratings on four central aspects of discourse production that have been addressed in previous research on this phenomenon: informativeness, clarity, understandability, and connectedness. We designed rating scales for the four aspects and used them to rate the collected discourse samples. The instruction specified that the ratings would not be summed up into a total score.

All audio-recordings were independently evaluated by three trained raters (AL, MK, and an independent third rater). The raters were familiar with the story and are professional linguists with prior experience of working with aphasic speech. Table 2 provides the

Table 1. Micro- and macrolinguistic variables used for the analysis.

Set	Features	Source
Microlinguistic variables	CIUs per minute(N/min)	CHAT annotation
	Word-level errors (N/N_{words})	CHAT annotation
	Fillers (N/N_{words})	CHAT annotation
	Ungrammatical EDUs (N/N_{EDUs})	CHAT annotation
	Lexical diversity (MATTR)	Lemmatized CHAT annotation
Macrolinguistic variables	Main events (N): information content	Annotation of main events
	Structural disfluencies (N/N_{EDUs}): false starts, corrections, restatements, and retracings	RST annotation
	Relation set size: (N of relations of different types used to build discourse structure)	RST annotation
	Meta-comments (N/N_{EDUs})	RST annotation

Table 2. Evaluation for the four selected aspects of coherence.

Informativeness. Informativeness refers to the amount of relevant information content in the retold story.	
Informative	(4) All the essential information is conveyed
	(3) Most of the essential information is conveyed
Uninformative	(2) Some of the essential information is conveyed; there are omissions that affect the understanding of the retold story
	(1) The information provided in the retelling is incomplete, a lot of information is missing
Clarity. The term clarity is used to describe the overall meaningfulness of discourse achieved through the appropriateness of its elements with respect to the overall topic of the story and narrative structure. Clarity is the degree to which the story as a whole, hangs together or makes sense.	
Clear	(4) The story is well structured, every episode of the story is appropriately placed, the narrative is the speaker does not deviate from the storyline too much
	(3) The story is structured well; there are occasional comments and deviations from the storyline, but they do not disrupt the sequence of elements/episodes of the story
Unclear	(2) The storyline is hard to follow because of multiple deviations or omissions from the storyline or confused order of episodes
	(1) The retelling is hard or impossible to follow because the storyline is confused and/or there were serious omissions
Understandability. Understandability reflects how well the meaning intended by the speaker could be perceived and/or interpreted by the listener. This is not a measure of information content, although omitted information may influence the understandability of a retelling.	
Understandable	(4) It is easy to understand the plot of the story, I understood everything well
	(3) It is possible to understand the overall plot of the story, although some parts are somewhat fuzzy
Not understandable	(2) It is hard to understand much of the story from the retelling; often it is not clear what the speaker is talking about
	(1) It is almost impossible to understand anything from the retelling
Connectedness. Connectedness touches upon the relationship between consecutive elements of the retelling and appropriate organization of its parts into a unified whole through the use of semantic and pragmatic relations.	
Connected	(4) Transitions between sentences and larger parts of the story are smooth
	(3) Transitions between parts of the story are mostly smooth, although some transitions between episodes and within them are missing or too abrupt
Disconnected	(2) The flow of the story is often disrupted, many transitions between sentences and larger parts of the story are missing or too abrupt
	(1) The retelling is "chunky", telegraphic, it is a disconnected description of separate episodes; sometimes it is hard to tell whether the speaker is continuing with the same story

definition and an overview of the coherence ratings used for each aspect of coherence. The median of the ratings across the three raters was taken for each of the coherence aspects for every sample.

Transcription

The discourse samples were transcribed using the Codes for the Human Analysis of Transcripts (CHAT) format (MacWhinney, 2000) by one of the authors (AL). The CHAT transcription format has been developed for the analysis of spoken discourse and further adjusted for aphasic speech (MacWhinney et al., 2011).

Segmentation

The segmentation criteria were based on the guidelines developed by Carlson and Marcu (2001) for the RST annotation of written discourse and adapted for the analysis of spoken discourse based on the works by Kibrik and Podlesskaja (2009), and MacWhinney (2000). A standard EDU is a syntactic clause with the non-clausal modifiers of its constituents, that has one of the commonly recognized uninterrupted intonation language-dependent

contour, delimited by boundary pauses, and represents one complete act or event. It should be noted that syntactic criteria for segmenting into EDUs are different from the ones for segmenting into utterances in CHAT-format. While an utterance includes a main clause with all associated dependent clauses, an EDU constitutes a single clause. Although the original EDU definition is based on syntactic structure, the guidelines for spoken-language segmentation suggest a combination of phonological, syntactic, and semantic criteria to identify EDUs (cf. Marini et al., 2011).

Microlinguistic variables

The microlinguistic variables we used in the classification analysis were calculated from the CHAT transcriptions using the CLAN Tool. Word-level phonological, semantic, and morphological errors were collapsed into a single variable, "word-level errors". Fillers and empty words (e.g., "em", "hm", "like", "what's its name", "you know"), which are frequently produced when word-finding problems arise, were counted separately. We also calculated ungrammatical EDUs, which are effectively agrammatic and paragrammatic EDUs (marked as [+gram] in CHAT; for definition see MacWhinney, 2000, p. 112). In order to adjust for narrative length, the counts of word-level errors and fillers were divided by the number of words, and the number of ungrammatical EDUs was divided by the total number of EDUs. Lexical diversity was measured using MATTR (Moving-average type-token ratio method; Covington & McFall, 2010) on automatically lemmatized transcripts (LemmaGen lemmatizer; Juršic et al., 2010) with manual correction. Correct information units (CIUs) were annotated in the transcript following the procedure described by Nicholas and Brookshire (1993), and the rate of CIUs per minute was used as a variable in the analysis.

Discourse structure

In RST, the interconnected EDUs form larger spans that are in turn linked to each other, building up into a tree representation. The connection between the nodes is described in terms of semantic and pragmatic relations, also called "discourse" or "coherence" relations. In every pair of spans, both units are assigned a nuclearity status: the so-called nucleus contains the more important part of the discourse and is independently comprehensible, while the satellite either elaborates the information presented in the nucleus, or is less pertinent to the core of the discourse, or the purpose of its author. Structures with two or more nodes equally central to the speaker's purpose and of equal importance in a discourse are called multinuclear (cf. Figure 1, where the multinuclear relation *Joint* connects EDUs (18) and (19). In mononuclear relations, one EDU elaborates or complements the other (cf. Figure 1, where the mononuclear relation *correction* connects satellite (17) with the nucleus (16), and the mononuclear relation *restatement* connects satellite (18–19) with the nucleus (16–17). Multinuclear constructions do not require assigning prominence to one of the EDUs or groups of EDUs or other manipulations with information structure. They could thus be easier for speakers with aphasia to produce, but their effect on coherence perception is unclear.

A detailed description of the basic principles of RST can be found in the original papers by Mann and Thompson (1985; Mann & Thompson, 1988), as well as in a series of works by Taboada (e.g., Taboada, 2004), and the reviews of Taboada and Mann (2006a; 2006b).

Discourse-structure annotation was performed using the software RST-Tool (O'Donnell, 2000). We used the relation set suggested by Carlson and Marcu (2001) with three additional relations for the impaired- and spontaneous-language phenomena: *retracing, correction*, and *false-start* (cf. similar to Kibrik & Podlesskaja, 2009; Stent, 2000). The definitions for these relations were adapted from the CHAT annotation scheme. A retracing is annotated when a speaker abandons an utterance, sometimes without completing it, and repeats it after providing additional information, for example, to clarify the circumstances or to insert a left-out element of a sequence. Corrections, or revisions, are retracings or reformulations with changes to the content of an utterance. False starts indicate incomplete utterance, after which a new tangent starts. Two other relations, *word-finding comment* and *meta-comment*, were added to mark deviations from the main storyline of the retellings. Word-finding comments are EDUs describing the process of finding or inability to remember a word or phrase. Meta-comments are the comments speakers made outside of the story, for example, about their discourse abilities, or to share a related memory. The relation set used in the study with definitions can be found in Table C1 of Appendix C.

Three macrolinguistic parameters were extracted from the RST annotations of discourse structure: the size of relation set (the number of different relation types), the number of meta-comments, divided by the total number of EDUs to adjust for the narrative length, and the rate of structural disfluencies (false-starts, restatements, retracings, word-finding comments, and corrections). For the latter we calculated the number of EDUs in the satellite node of the relation, e.g., in Figure 1 EDUs (17), (18), and (19) are labelled as structural disfluencies. Then, the number of EDUs was divided by the total number of EDUs to adjust for the narrative length.

Main events

The method for analysing the "number of main events" developed by Wright et al. (2005) was used to evaluate information content in the samples. A set of main events in the stimulus story was identified through the comparison of the events included by all the participants in the NBD group. The list of the main events used for the count is listed in Appendix D.

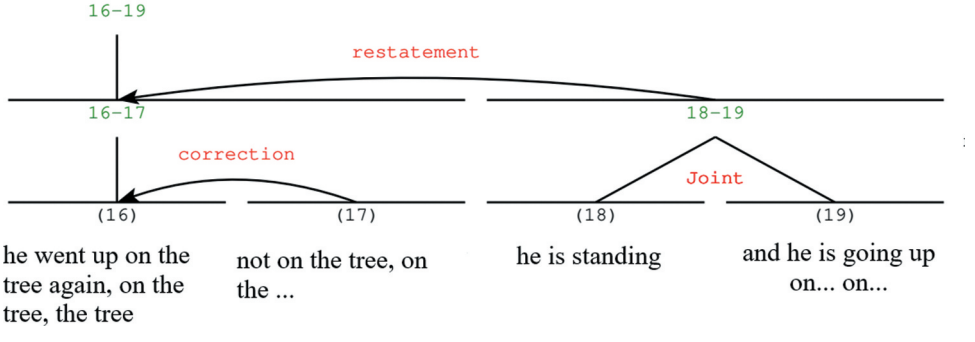

Figure 1. Rhetorical tree example (English translation).

Statistical analysis

The between-group comparisons of the four-point ratings were performed for each of the four aspects of coherence separately. As the distribution of the data did not meet the assumption of normality, one-way permutation tests of independence for ordinal data were performed for each aspect with an adjustment for multiple comparisons. Effect sizes were estimated with Vargha and Delaney's A (VDA, 2002), where values range from 0 to 1, with 0.5 indicating stochastic equality.

The classification analysis was performed using random forests (Breiman, 2001), an ensemble learning method for classification and regression. The algorithm constructs a collection of decision trees based on randomly selected observations and predictors out of a set of given variables (the sampling method is referred to as bootstrap aggregation, or "bagging"). The output class is then predicted for unseen observations based on the average of the predictions from all trees in the collection (in case of regression) or the majority class (in case of classification). Experimental datasets often suffer from ceiling/floor effects, have non-normal distributions and relatively small numbers of observations. The random-forests method has several advantages crucial for this study: it does not have any formal distribution assumptions, works on small data samples, and is generally robust against overfitting (Breiman, 2001; Segal, 2004). In RFs, each node of each tree is constructed by selecting a single predictor and cutpoint for it, allowing the inclusion of collinear predictors (Tomaschek et al., 2018). It also uses feature selection, a mechanism for determining variable importance.

Results

Group comparison

The ratings for informativeness, clarity, understandability, and connectedness are presented in Table E1 of Appendix E, the distribution of ratings by groups is presented in Figure 2. We found significant effects of the group on all four aspects. P-values, VDA statistics, and differences in median scores between groups are presented in Table 3.

Classification

In the classification analysis, the ratings were considered independently of the participant group (PWA/NBD) in order to eliminate any intrinsic group effects. Due to the small size of the sample, the four-point ratings were converted into binary ones (high/rather high vs. low/rather low or present/mostly present vs. absent/mostly absent). Although binary ratings are potentially less informative than the four-point ones, we considered them sufficient for the purposes of the present investigation. The data set used for classification thus comprised only the binary ratings for each aspect of coherence and the chosen continuous and discrete linguistic variables (see Table 1). As the aphasia group had diverse impairment profiles, we expected that various linguistic deficits would occur in their discourse samples, as well as similar impairments with different severity. The diversity fortunately provided us with numerous combinations of higher and lower indicators for the chosen linguistic variables and different coherence ratings. Data on all variables are presented in Tables E2 and E3 of Appendix E.

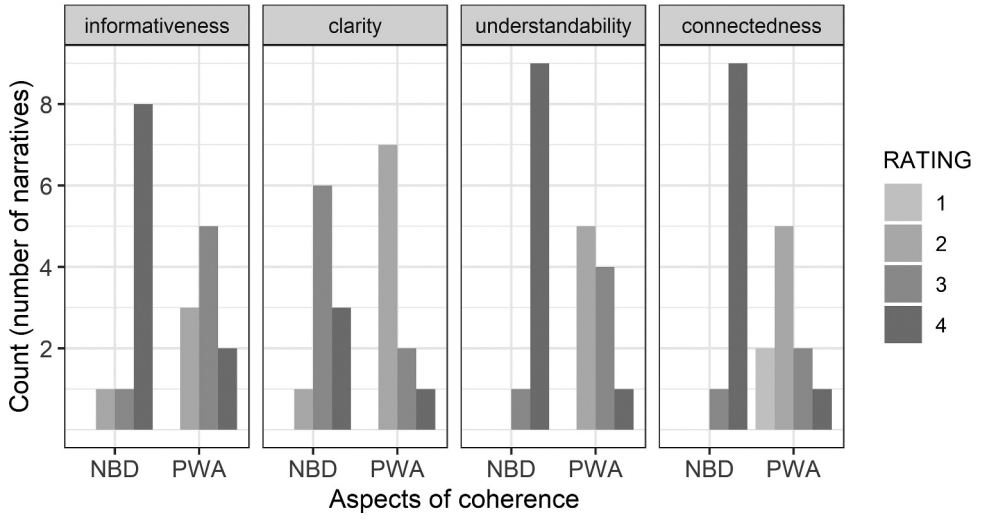

Figure 2. Distribution of ratings of different aspects of coherence.

Table 3. Group comparison results.

	Aspect of coherence			
	Informativeness	Clarity	Understandability	Connectedness
p_{adj}	0.04*	0.04*	0.002**	0.002**
VDA	0.790	0.800	0.925	0.935
Difference in median scores	1	1	1.5	2

*Significant values p <0.05; **very significant value p < 0.01. VDA values range from 0 to 1, VDA < 0.5 indicates higher scores in PWA group, VDA >0.5 indicates higher scores in NBD group.

A random-forest model was trained to classify discourse samples into the lower-coherence (scores 1 and 2) or the higher-coherence (scores 3 and 4) group based on the selected linguistic variables. Repeated three times 10-fold cross-validation was performed with values for the m_{try} parameter ranging between 1 and 10, indicating the number of randomly sampled variables, and n_{tree}, number of decision trees, set to 1000. A description of the in-built feature-selection mechanisms for random forests can be found in the documentation of R package *caret* (Kuhn, 2008). The top predictors (with variable importance above 70%) for each aspect of coherence and variable importance estimates are presented in Figure 3. Average classification accuracy was 0.84 for informativeness (Kappa = 0.355; OOB Error = 15%; m_{try} = 3), 0.8 for clarity (Kappa = 0.57; OOB Error = 20%; m_{try} = 3), 0.8 for understandability (Kappa = 0.34; OOB Error = 20%; m_{try} = 2), and 0.95 for clarity (Kappa = 0.89; OOB Error = 5%; m_{try} = 1).

Discussion

Coherence in aphasia

In this study, coherence was operationalised through a combination of four aspects of discourse production: informativeness, clarity, connectedness, and understandability, as

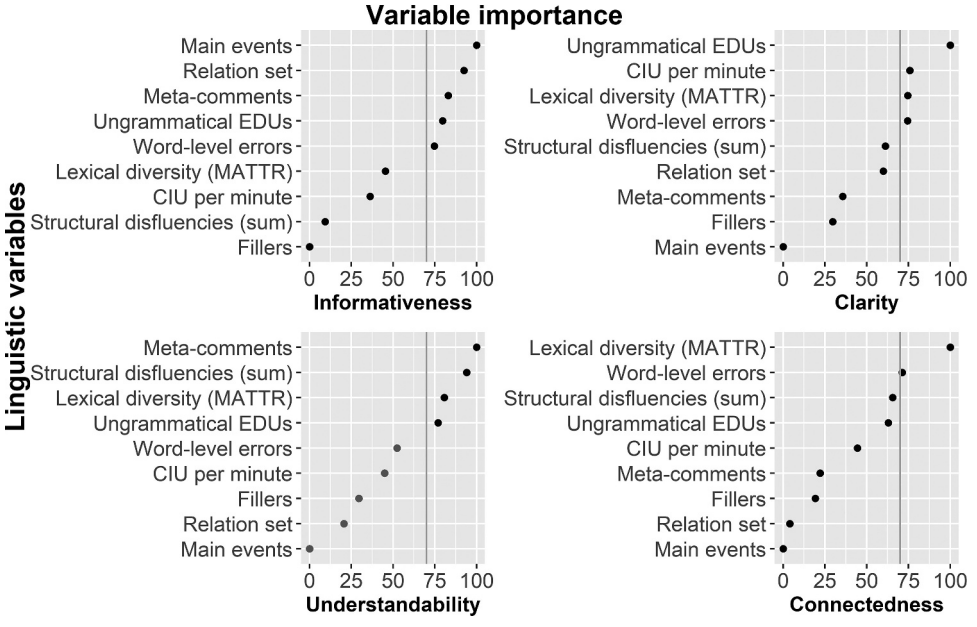

Figure 3. Random-forest feature selection plot.

perceived by the listener. We refrained from collapsing these ratings into a single coherence score, as the contribution of each aspect to the overall perception of coherence remains to be determined. Instead, we accepted that different coherence deficits may exist, in which one or more of its aspects are impaired while the others remain intact. The scores on all four aspects of coherence in the PWA group were on average lower than those in the NBD group. Individual performance varied, and on some aspects of coherence, speakers with aphasia received higher ratings than neurologically healthy participants (see Table E1). This observation is in line with the existence of a continuum between disturbed and non-impaired linguistic performance, as suggested by H.K. Ulatowska et al. (1981). We have noted that the disparity in previously reported results on communicative abilities and coherence in aphasia resulted from the differences in aphasia profiles of study participants as well as in definitions and operationalisations of the term. The novel view we proposed was intended to capture the complex multifaceted nature of coherence, which had been referred to as a monolithic concept in previous studies, and to shed light on the linguistic mechanisms involved in communication of people with aphasia. Our analysis indicated that results may indeed vary depending on the aspect of coherence being considered, as different combinations of linguistic variables were found to be relevant for each of them.

Aspects of coherence

Informativeness

One could argue that the main goal of discourse production is the transfer of information between interlocutors. Informativeness reflects whether a speaker was successful in

achieving this goal. The variables identified as the most likely predictors of informativeness were the information content (main events), the set of discourse relations used to signal connections between different parts of discourse (e.g., causal, explanatory, or evaluative), and the ratio of deviations from the main storyline (meta-comments). The relationship between thematic informativeness and discourse coherence has been addressed in previous research (e.g., Andreetta et al., 2012; Capilouto et al., 2006; Gleason et al., 1980). In our study, the amount of information content (number of main events) was a strong predictor for informativeness and was less relevant to the other aspects of coherence.

Our analysis also indicated that syntactic and word-level (phonological, semantic, and morphological) errors had an impact on the perception of informativeness. The term "syntactic errors" is used here broadly to refer to clause-level, or rather EDU-level, paragrammatic and agrammatic events, such as omissions or substitution errors, confused or incomplete use of grammatical structures (ungrammatical EDUs). During language production, lexical entries are supplemented with morpho-syntactic information to signal relations between them. Together they are integrated with information available at the discourse level (e.g., Avrutin, 2006; Creswell, 2003). Syntactic form may be constrained by discourse (e.g., use of pronouns in absence of a clear referent) and, conversely, discourse may constrain the choice of syntactic structure (e.g., using clefts typically assigns prominence to a constituent). Violations of these constraints and morphosyntactic errors, such as omission of a subject or incorrect inflections, may lead to confusion regarding information structure and coreference, which, in turn, can damage the overall informativeness of a discourse sample.

Connectedness

Connectedness refers to the unity of discourse established through the semantic and pragmatic relations between its elements on a local level (e.g., Glosser & Deser, 1991; Sanders & Canestrelli, 2012). Here as well we found the effects of lexical-semantic and morphological levels (lexical diversity, word-level errors), which can result in breakdowns of essential referential ties and an overall lack of continuity in a narrative.

Clarity

Clarity is the aspect of coherence that refers to the semantic unity of discourse achieved through the appropriate organization of its parts with respect to the overall topic and goal. The best predictors of clarity were syntactic errors (ungrammatical EDUs), efficiency (CIU/min), and, as in the case of connectedness, lexical diversity, and word-level errors. Reduced efficiency accompanied by numerous microlinguistic errors makes discourse fragmented and hard to follow, decreasing its overall semantic unity. This outcome is in line with previous reports on the effects of microlinguistic impairments, and lexical and syntactic deficits in particular, on coherence (e.g., Wright & Capilouto, 2012; Andreetta & Marini, 2015 ; Andreetta & Marini, 2015).

The overlap in microlinguistic predictors between connectedness, clarity, and informativeness is not unexpected. Our group comparison demonstrated that PWA received lower ratings, on average, on all four aspects of coherence than the control participants. While people with aphasia commonly have microlinguistic deficits, this type of errors is scarcely produced by non-brain-damaged speakers. Consequently, microlinguistic errors

become an extremely good predictor of group and by transition of coherence ratings, without necessarily being one. Nevertheless, this consideration does not per se mean that the effect of microlinguistic deficits should be ignored or diminished. It simply is an indication that the perception of discourse produced by PWA in general is influenced by microlinguistic impairments.

Understandability

Several studies called for the need to consider coherence as a product of interaction between interlocutors (e.g., Christiansen, 1995; Olness & Ulatowska, 2011). Understandability reflects a subjective assessment of how well a discourse produced by a speaker is understood by the listener. While some of the parameters that led to reduced clarity were also among the selected predictors of decreased understandability (lexical diversity, ungrammatical EDUs), the strongest identified predictors of this aspect were structural disfluencies and meta-comments. This outcome indicates that understandability is achieved on multiple levels. It is likely that discourse with low clarity due to lexical-retrieval and/or syntactic issues may still be understandable, as microlinguistic deficits and lower-level connectivity issues can be compensated through context, common ground, or even world knowledge. It may be harder to compensate for numerous discourse-structure impairments (structural disfluencies) or challenges such as multiple and/or extensive deviations from the main storyline (meta-comments). While structural disfluencies, such as false-starts, corrections, and retracings, are not uncommon in spontaneous discourse production, they are used excessively in aphasic speech in order to repair linguistic deficits. This observation supports the assumption that discourse structure is relevant for coherence, which has previously not been substantiated by language-impairment data (e.g., Capilouto et al., 2006; Giora, 1997; Hobbs, 1978; Redeker, 1990; Ulatowska & Olness, 2000).

General discussion

The focus of this study was the relationship between perceptual ratings of discourse coherence and a number of micro- and macrolinguistic variables that have been reported to have an effect on or to be correlated with coherence in aphasic discourse (e.g., Andreetta & Marini, 2015; Marini et al., 2011; Wright & Capilouto, 2012). Our aim was to advance the understanding of linguistic mechanisms involved in establishing and maintaining discourse coherence by identifying linguistic variables linked to its decrease. Coherence was defined as the overall semantic unity of discourse and operationalised through a combination of ratings on four aspects: informativeness, clarity, connectedness, and understandability. While these aspects of coherence have been investigated previously, they were studied separately and using different methods. As a result, conclusions on coherence in aphasia varied. On average, discourse samples produced by our aphasia group received lower ratings than those of the non-language-impaired group on all four aspects of coherence, providing further evidence of coherence decrease/impairment in aphasia (research question 1).

Based on the recent findings demonstrating a connection between micro- and macrolinguistic processes and coherence in discourse production (e.g., Andreetta & Marini, 2015; Wright & Capilouto, 2012), we made an assumption that different aspects of

coherence would not be associated with the same linguistic variables and deficits (research question 2). Drawing on previous research on discourse production by non-language-impaired individuals (e.g., Mann & Thompson, 1985), we also wanted to examine the contribution of discourse structure to coherence in aphasia. The conducted classification analysis included microlinguistic variables, such as word-level error rate and proportion of fillers, as well as macrolinguistic ones – the number of main events mentioned in the retellings and the proportion of digressions from the storyline (meta-comments). We used a formal approach to analyse the discourse structure of the samples (RST) and included the following parameters from this analysis into the classification: structural disfluencies (e.g., corrections, restatements) and the number of types of discourse (or rhetorical) relations used to express connections between different, sometimes distant, parts of discourse. The obtained results indicated that different combinations of variables were relevant for each of the rated aspects of coherence and that a connection existed between coherence and discourse structure.

Together, the outcomes for the four aspects paint the following picture of how coherence may be constructed and impaired. Lexical-semantic deficits commonly observed in aphasia lead to a decrease in connectedness, a local aspect of coherence. The impact of lexical-semantic deficits can, of course, extend beyond the local level, which is why lexical diversity and word-level errors were among the top predictors for all aspects of coherence. Aside from that, different predictors were selected for each aspect of coherence. Morphosyntactic deficits (word-level errors, ungrammatical EDUs) coupled with low efficiency (CIU/min) decrease the clarity of discourse, and in combination with low information content (main events) and its poor presentation (relation set) result in lower informativeness. Lexical and syntactic deficits also had an effect on the overall understandability of a discourse. However, our results indicated that understandability was affected more by macrolinguistic issues, such as discourse-structure impairment through numerous revisions, replanning (structural disfluencies), and/or departures from the storyline (meta-comments). Possibly, microlinguistic deficits can be compensated through the scaffolding provided by context and pragmatic information shared between speakers, while structural problems are harder for speakers to overcome and for listeners to process. The detrimental effects of discourse-structure impairments have also been observed in other populations with neurological and mental disorders, such as schizophrenic speakers and people with right-hemisphere brain damage whose microlinguistic deficits are generally mild (Johns et al., 2008; Marini et al., 2005; Martin & McDonald, 2003; Kuperberg, 2010a; 2010b).

While the classification accuracy was high overall, elevated levels of classification errors were observed for the samples with low informativeness and understandability ratings. It is possible that the linguistic variables crucial for these aspects were not considered in this study or that the models overfit due to an imbalance in the informativeness and understandability groups (4/16 and 5/15 observations with low/high scores, respectively). Another possibility is that the key to a better understanding of coherence lies in the pragmatic aspect of discourse production, or the ability of a listener to make sense of a discourse. It has been noted before that people "naturally assume coherence and tend to interpret text [discourse] in light of this assumption" (Brown & Yule, 1983, p. 66; Asher & Lascarides, 2003). In conversations with people with aphasia, listener ability to bridge the gaps created by linguistic disturbances becomes paramount. Classification errors could

thus indicate that the scores for informativeness and understandability were raised through the use of context and/or pragmatic information shared between interlocutors. While pragmatic effects are out of the scope of this study, we propose that there may be a pragmatic component to coherence, which would explain why discourse can be sufficiently coherent to be understood in spite of poor linguistic form. A follow-up investigation of the effects of shared pragmatic knowledge between speakers and listeners, such as, for example, familiarity with the stimulus, could shed light on the proposed explanation of the results of this study (e.g., Olness et al., 2005).

Limitations

The following considerations outline the limitations of our study. Most importantly, the small sample size limits the generalizability of the described findings. The assumption that discourse samples can be classified into just two groups by coherence (lower vs. higher coherence) may not reflect the natural distribution of coherence ratings. Coherence evaluation was performed by trained raters familiar with the story and with aphasic speech. While ratings from naive listeners could provide invaluable insights into coherence and the role of pragmatic knowledge, naive raters could only rate a single retelling, making it impossible to calculate an inter-rater agreement. We focused on one genre of spontaneous discourse, video-induced story retelling. This restriction was dictated mainly by the comparability requirements and the choice of methodology, since Rhetorical Structure Theory has not been widely used to study language-impairments (Abdalla et al., 2018; Kibrik & Podlesskaja, 2009). It would be insightful to include other genres and types of narratives, such as personal stories and conversations, in further investigations, as outcomes may differ depending on the genre.

Conclusion

The outcomes of this study extend the body of work on coherence in discourse production by people with aphasia and non-language-impaired speakers. We addressed coherence as a complex multifaceted phenomenon operationalised through the combination of four aspects: informativeness, clarity, connectedness, and understandability. These aspects have previously only been addressed separately, which resulted in disparate findings on coherence in aphasia. We argue that coherence should not be considered solely as the internal connectedness of a discourse, or the reflection of appropriate information content. Rather, coherence is a quality of language in use, co-constructed by a speaker and the intended listener or reader. Using classification analysis, we discovered that different combinations of micro- and macrolinguistic variables were relevant for each of them. Most of the variables we considered had been linked to (one of the aspects of) coherence in previous research on discourse production in aphasia. However, we also used a formal framework to analyse discourse structure and factored it into the classification analysis. Although the connection between discourse production in aphasia and discourse structure has barely been explored, numerous studies of non-impaired discourse suggested that coherence was achieved through the construction of discourse structure. Our results indicated that a well-formed discourse structure was especially important for two aspects of coherence – informativeness and understandability. Though pragmatic effects were not

examined directly, we suggest that shared pragmatic knowledge could boost coherence, as it helps interlocutors to compensate for language deficits.

From a clinical perspective, our findings could inform aphasia assessment. According to Bryant et al. (2016), 89% of the speech pathologists use a judgement-based analysis in clinical practice to evaluate the discourse abilities of people with aphasia. While this does not necessarily relate to coherence, it is likely that ratings may be preferred over linguistic analyses for coherence evaluation in clinical settings. Our findings indicate that a rating-based assessment of each aspect of coherence may provide more accurate information. Understanding the connection between various linguistic deficits and coherence could also inform the choice of a treatment program, though these results have to be validated on a larger, more representative data set.

The data tables, R code, and supplementary materials are available to speech and language researchers at https://osf.io/xnr5w/?view_only=25f23fad1795488ca9d12d3cb2c59269

Acknowledgments

This work was supported within the framework of a subsidy by the Russian Academic Excellence Project '5-100'; Erasmus Mundus International Doctorate for Experimental Approaches to Language and Brain (IDEALAB) under grant number 2012-1713, Framework Partnership Agreement number 2012-0025.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was funded within the framework of a subsidy by the Russian Academic Excellence Project '5-100' and by the Erasmus Mundus Joint Doctorate (EMJD) program IDEALAB (International Doctorate for Experimental Approaches to Language And Brain; Macquarie University, Newcastle University, University of Groningen, University of Trento and University of Potsdam, FPA 2012-0025, grant nr. 2012-1713).

ORCID

Manfred Stede http://orcid.org/0000-0001-6819-2043
Mariya Khudyakova http://orcid.org/0000-0002-5293-3991

References

Abdalla, M., Rudzicz, F., & Hirst, G. (2018). Rhetorical structure and Alzheimer's disease. *Aphasiology*, *32*(1), 41–60. https://doi.org/10.1080/02687038.2017.1355439

Agar, M., & Hobbs, J. R. (1982). Interpreting discourse: Coherence and the analysis of ethnographic interviews. *Discourse Processes*, *5*(1), 1–32. https://doi.org/10.1080/01638538209544529

Akhutina, T. (2016). Luria's classification of aphasias and its theoretical basis. *Aphasiology*, *30*(8), 878–897. https://doi.org/10.1080/02687038.2015.1070950

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia*, *50*(8), 1787–1793. https://doi.org/10.1016/j.neuropsychologia.2012.04.003

Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology*, *29*(6), 705–723. https://doi.org/10.1080/02687038.2014.979394
Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology*, *29*(6), 705–723. doi: 10.1080/02687038.2014.979394
Armstrong, E. (1987). Cohesive harmony in aphasic discourse and its significance in listener perception of coherence. In R. H. Brookshire (Ed.), *Clinical aphasiology conference proceedings* (Vol. 17, pp. 210–215). BRK Publishers.
Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*(9), 875–892. https://doi.org/10.1080/02687030050127685
Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
Avrutin, S. (2006). Weak syntax. *Broca's Region*, 49–62. doi:10.1093/acprof:oso/9780195177640.003.0004
Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.
Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, *30*(7), 489–518. doi: 10.3109/02699206.2016.1145740
Capilouto, G. J., Wright, H. H., & Wagovich, S. A. (2006). Reliability of main event measurement in the discourse of individuals with aphasia. *Aphasiology*, *20*(2–4), 205–216. https://doi.org/10.1080/02687030500473122
Carlson, L., & Marcu, D. (2001). *Discourse tagging reference manual* (Information Science Institute Technical Report ISI-TR-545). University of South Carolina.
Chafe, W. (Ed). (1980). *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Ablex.
Christiansen, J. A. (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*, *51*(2), 291–317. https://doi.org/10.1006/brln.1995.1062
Coelho, C., & Flewellyn, L. (2003). Longitudinal assessment of coherence in an adult with fluent aphasia: A follow-up study. *Aphasiology*, *17*(2), 173–182. https://doi.org/10.1080/729255216
Coelho, C. A., Liles, B. Z., Duffy, R. J., Clarkson, J. V., & Elia, D. (1994). Longitudinal assessment of narrative discourse in a mildly aphasic adult. *Clinical Aphasiology, 22*, 145-155. Retrieved from: http://aphasiology.pitt.edu/164/
Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100. https://doi.org/10.1080/09296171003643098
Creswell, C. (2003). *Syntactic form and discourse function* [Doctoral dissertation]. University of Pennsylvania.
Delaney, H. D., & Vargha, A. (2002). Comparing several robust tests of stochastic equality with ordinally scaled variables and small to moderate sized samples. Psychological Methods, 7(4), 485–503. doi: 10.1037/1082-989X.7.4.485
Ellis, C., Henderson, A., Wright, H. H., & Rogalski, Y. (2016). Global coherence during discourse production in adults: A review of the literature. *International Journal of Language & Communication Disorders*, *51*(4), 359–367. https://doi.org/10.1111/1460-6984.12213
Gernsbacher, M. A., & Givon, T. (1995). Introduction: Coherence as a mental entity. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text* (pp. 7–10). John Benjamins Publishing.
Giora, R. (1997). Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory. *Journal of Pragmatics*, *27*(1), 17–34. https://doi.org/10.1016/0378-2166(95)00065-8
Gleason, J. B., Goodglass, H., Obler, L., Green, E., Hyde, M. R., & Weintraub, S. (1980). Narrative strategies of aphasic and normal-speaking subjects. *Journal of Speech, Language, and Hearing Research*, *23*(2), 370–382. https://doi.org/10.1044/jshr.2302.370
Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language*, *40*(1), 67–88. https://doi.org/10.1016/0093-934X(91)90117-J

Grosz, Barbara J. and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. Computational Linguistics 12(3): 175–204. Retrieved from: http://www.aclweb.org/anthology-new/J/J86/J86-3001.pdf

Hobbs, J. (1978). *Why is discourse coherent?* (Technical report 176, SRI International). Artificial Intelligence Center, SRI International.

Holland, A. L. (1982). Observing functional communication of aphasic adults. *Journal of Speech and Hearing Disorders*, *47*(1), 50–56. doi: 10.1044/jshd.4701.50

Iruskieta, M., Da Cunha, I., & Taboada, M. (2015). A qualitative comparison method for rhetorical structures: Identifying different discourse structures in multilingual corpora. *Language Resources and Evaluation*, *49*(2), 263–309. https://doi.org/10.1007/s10579-014-9271-6

Johns, C. L., Tooley, K. M., & Traxler, M. J. (2008). Discourse impairments following right hemisphere brain damage: A critical review. *Language and Linguistics Compass*, *2*(6), 1038–1062. https://doi.org/10.1111/j.1749-818X.2008.00094.x

Juršic, M., Mozetic, I., Erjavec, T., & Lavrac, N. (2010). Lemmagen: Multilingual lemmatisation with induced ripple-down rules. *Journal of Universal Computer Science*, *16*(9), 1190–1214.doi:10.3217/jucs-016-09-1190

Khudyakova, M., Bergelson, M., Akinina, Y., Iskra, E., Toldova, S., & Dragoy, O. (2016). Russian CliPS: A corpus of narratives by brain-damaged individuals. In *Proceedings of LREC 2016 Workshop. Resources and processing of linguistic and extra-linguistic data from people with various forms of cognitive/psychiatric impairments (RaPID-2016), 2016 (No. 128), Portorož, Slovenia*. Linköping University Electronic Press.

Kibrik, A. A., & Podlesskaja, V. I. (Eds). (2009). *Rasskazy o snovidenijax: Korpusnoe issledovanie ustnogo russkogo diskursa [Night Dream Stories: A corpus study of spoken Russian discourse]*. Jazyki slavjanskix kul'tur.

Kong, A. P. H., Linnik, A., Law, S. P., & Shum, W. W. M. (2018). Measuring discourse coherence in anomic aphasia using Rhetorical Structure Theory. *International Journal of Speech-language Pathology*, *20*(4), 406–421. https://doi.org/10.1080/17549507.2017.1293158

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuperberg, G. R. (2010a). Language in schizophrenia part 1: An introduction. *Language and Linguistics Compass*, *4*(8), 576–589. doi:10.1111/j.1749-818X.2010.00216.x

Kuperberg, G. R. (2010b). Language in schizophrenia Part 2: What can psycholinguistics bring to the study of schizophrenia … and vice versa? *Language and Linguistics Compass*, *4*(8), 590–604. https://doi.org/10.1111/j.1749-818X.2010.00217.x

Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology*, *30*(7), 765–800. https://doi.org/10.1080/02687038.2015.1113489

Luria, A. R. (1966). *Higher cortical functions in man*. Basic Books.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

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

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Palgrave Macmillan.

Mann, W. C., & Thompson, S. A. (1985,). Assertions from discourse structure. In *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistic Societyl*, 11, 245–258. Berkeley Linguistic Society, Berkeley. doi: 10.3765/bls.v11i0.1886

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, *8*(3), 243–281. https://doi.org/10.1515/text.1.1988.8.3.243

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

Marini, A., Carlomagno, S., Caltagirone, C., & Nocentini, U. (2005). The role played by the right hemisphere in the organization of complex textual structures. *Brain and Language*, *93*(1), 46–54. https://doi.org/10.1016/j.bandl.2004.08.002

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, *85*(3), 451–466. https://doi.org/10.1016/S0093-934X(03)00070-1

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*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

O'Donnell, M. (2000). RSTTool 2.4 – A markup tool for rhetorical structure theory. In *Proceedings of the international natural language generation conference* (pp. 253–256). Mitzpe Ramon, Israel.

Olness, G. S., & Ulatowska, H. K. (2011). Personal narratives in aphasia: Coherence in the context of use. *Aphasiology*, *25*(11), 1393–1413. https://doi.org/10.1080/02687038.2011.599365

Olness, U., Carpenter, W.-H., & Dykes. (2005). Holistic assessment of narrative quality: A social validation study. *Aphasiology*, *19*(3–5), 251–262. https://doi.org/10.1080/02687030444000723

Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language & Communication Disorders*, *52*(6), 689–732. https://doi.org/10.1111/1460-6984.12318

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, *14*(3), 367–381. https://doi.org/10.1016/0378-2166(90)90095-U

Sanders, T. J., & Canestrelli, A. R. (2012). The processing of pragmatic information in discourse. In H.-J. Schmid (Ed.), *Cognitive Pragmatics*, (pp. 201–232). De Gruyter Mouton. doi: 10.1515/9783110214215.201

Segal, M. (2004). *Machine learning benchmarks and Random Forest regression* (Center for Bioinformatics and Molecular Biostatistics, Technical report). University of California San Francisco.

Stent, A. (2000). Rhetorical structure in dialog. In *INLG'2000 proceedings of the first international conference on natural language generation* (pp. 247–252). Mitzpe Ramon, Israel.

Taboada, M., & Mann, W. C. (2006a). Applications of rhetorical structure theory. *Discourse Studies*, *8*(4), 567–588. https://doi.org/10.1177/1461445606064836

Taboada, M., & Mann, W. C. (2006b). Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies*, *8*(3), 423–459. https://doi.org/10.1177/1461445606061881

Taboada, M. T. (2004). Building coherence and cohesion: Task-oriented dialogue in English and Spanish (Vol. 129). Amsterdam/Philadelphia: John Benjamins Publishing.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, *71*, 249–267. https://doi.org/10.1016/j.wocn.2018.09.004

Ulatowska, H. K., Allard, L., & Chapman, S. B. (1990). Narrative and procedural discourse in aphasia. In Y. Joanette & H. H. Brownell (Eds.), *Discourse ability and brain damage* (pp. 180–198). Springer. https://doi.org/10.1007/978-1-4612-3262-9_8

Ulatowska, H. K., Allard, L., Reyes, B. A., Ford, J., & Chapman, S. (1992). Conversational discourse in aphasia. *Aphasiology*, *6*(3), 325–330. doi: 10.1080/02687039208248602.

Ulatowska, H. K., Chapman, S., Johnson, J., & Branch, C. (1999). Macrostructure and inferential processing in discourse of aphasic patients. *Psychology of Language and Communication, 3*(2), 3-20. Retrieved from https://www.researchgate.net/publication/254947976_Macrostructure_and_inferential_processing_in_discourse_of_aphasic_patients

Ulatowska, H. K., Freedman-Stern, R., Doyel, A. W., Macaluso-Haynes, S., & North, A. J. (1983). Production of narrative discourse in aphasia. *Brain and Language*, *19*(2), 317–334. https://doi.org/10.1016/0093-934X(83)90074-3

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, *13*(2), 345–371. https://doi.org/10.1016/0093-934X(81)90100-0

Ulatowska, H. K., & Olness, G. S. (2000). Discourse revisited: Contributions of lexico-syntactic devices. *Brain and Language*, *71*(1), 249–251. https://doi.org/10.1006/brln.1999.2261

Ulatowska, H. K., Olness, G. S., & Williams, L. J. (2004). Coherence of narratives in aphasia. *Brain and Language*, *91*(1), 42–43. https://doi.org/10.1016/j.bandl.2004.06.024

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101–132.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational linguistics*, *31*(2), 249–287. doi: 10.1162/0891201054223977.

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

Wright, H. H., Capilouto, G., Wagovich, S., Cranfill, T., & Davis, J. (2005). Development and reliability of a quantitative measure of adults' narratives. *Aphasiology*, *19*(3–5), 263–273. https://doi.org/10.1080/02687030444000732

Appendices

Appendix A. Participant data

Table A1. Participant data: aphasic speakers

ID	Sex	Age	Education	Aphasia type	Severity	Time post-onset	Etiology	Lesion location
P11	F	51	Vocational	Dynamic, complex motor	Severe	5 y, 6 m	CVA	IMCA
P12	F	47	Higher (unfinished)	Dynamic, complex motor	Moderate	2 y, 3 m	CVA	IMCA
P13	M	56	Vocational	Dynamic, complex motor	Moderate	1 y, 2 m	CVA	IMCA
P14	M	73	Higher	Sensory	Severe	1 y, 4 m	CVA	IMCA
P15	M	50	Secondary	Complex motor	Mild	7 y, 9 m	CVA	IMCA
P16	M	70	Unknown	Sensory, acoustic-mnestic	Moderate	4–5 m	CVA	Unknown
P17	M	52	Higher	Complex motor	Moderate	3 m	CVA	IMCA
P18	F	40	Vocational	Sensory, acoustic-mnestic	Moderate	3 m	CVA	IMCA
P19	F	67	Higher	Sensory, acoustic-mnestic	Moderate	2 y, 6 m	CVA	IMCA
P20	M	58	Secondary	Sensory	Moderate	1 y, 9 m	CVA	IMCA

"Higher" education status stands for a completed college/university degree, "higher (unfinished)" – for an uncompleted college/university degree.

CVA, cerebrovascular accident, IMCA, left middle cerebral artery; y, years; m, months.

Table A2. Participant data: NBD speakers

ID	Sex	Age	Education
P1	M	42	Higher
P2	M	63	Higher
P3	M	62	Higher
P4	F	50	Higher
P5	F	53	Higher
P6	M	49	Higher
P7	F	75	Higher
P8	F	66	Higher
P9	M	85	Higher
P10	F	45	Higher

"Higher" education status stands for a completed college/university degree.

Appendix B. Inter-rater agreement

Table B1. Inter-rater agreement

Linguistic analysis	Inter-rater agreement measure	N_{raters}	$N_{narratives}$	N_{EDUs}	Inter-rater agreement
Coherence evaluation					
Informativeness 0.715	Kendall's coefficient of concordance	2	20	1192	(100%)
Clarity					0.710
Understandability					0.648
Connectedness	0.763				
Annotation					
CIUs	Inter-rater disagreement were calculated for each narrative as the N_{CIUs} annotated by rater 1 minus N_{CIUs} annotated by rater 2. Then the absolute values of differences were averaged. Agreement was calculated as 1 –average $N_{diagreements}$	2	4	176	(15%)
98%					
Segmentation	The proportion of segment boundaries identified by both raters out of the sum of all independent boundary placement decisions made by the two raters	2	2	93	(8%)
90%					
RST	Following Iruskieta et al. (2015), the inter-annotator agreement was assessed in terms of the four types of decision involved in the annotation, that is, the choice of a discourse relation; the attachment point of the EDU; the constituent which it is attached to; and nuclearity.	2	2	74	(6%)
77%					
Main events 96%	Percentage of shared judgments	2	20	1192	(100%)

Appendix C. List of rhetorical relations

Table C1. Rhetorical relations used in RST annotation

Class	Multinuclear relation	Mononuclear relation	Source
Attribution		Attribution	(Carlson & Marcu, 2001)
		Attribution-negative	
Background		Background	
		Circumstance	
Cause	Cause–result	Cause	
		Result	
	Consequence	Consequence	
Comparison	Analogy	Analogy	
	Comparison	Comparison	
		Preference	
	Proportion		
Condition		Condition	
		Contingency	
		Hypothetical	
	Otherwise	Otherwise	
Contrast	Contrast	Antithesis	
		Concession	
Elaboration		Elaboration-additional	
		Elaboration-general-specific	
		Elaboration-part-whole	
		Elaboration-process-step	
		Elaboration-object-attribute	
		Elaboration-set-member	
		Example	
		Definition	
Enablement		Enablement	
		Purpose	
Evaluation		Comment	
	Conclusion	Conclusion	
	Evaluation	Evaluation	
	Interpretation	Interpretation	
Explanation		Evidence	
		Explanation-argumentative	
	Reason	Reason	
Joint	Disjunction		
	List		
Manner-means		Manner	
		Means	
Topic-comment	Comment-topic		
	Problem–solution	Problem–solution	
	Question–answer	Question–Answer	
	Statement–response	Statement–Response	
	Topic-comment		
		Rhetorical-question	
Summary		Summary	
Temporal	Inverted sequence		
	Sequence		
		Temporal-after	
		Temporal-before	
	Temporal-same-time	Temporal-same-time	
Topic change	Topic-drift	Topic-drift	
	Topic-shift	Topic-shift	
Spoken-language phenomena		Correction	Linnik et al., based on CHAT
		False-start	
		Restatement	
		Retracing	
Deviations from the main storyline		Meta-comment	Linnik et al.
		Word-finding comment	

Appendix D. List of Main events

(1) The gardener is gathering pears
(2) The boy arrives and steals a basket of pears
(3) The boy falls down; the pears are on the grounded
(4) Three boys help the boy to pick up the pears
(5) The boy gives the three boys pears
(6) The three boys pass the gardener while eating the pears
(7) The gardener is confused

Appendix E. Coherence rating scores and micro- and macrolinguistic parameters of the narratives

Table E1. Coherence ratings of the narratives

ID	Group	Informativeness	Clarity	Understandability	Connectedness
P1	NBD	4	4	4	4
P2	NBD	4	4	4	4
P3	NBD	4	3	4	4
P4	NBD	4	3	4	4
P5	NBD	4	4	4	4
P6	NBD	4	3	4	4
P7	NBD	3	3	4	3
P8	NBD	4	3	4	4
P9	NBD	2	2	3	4
P10	NBD	4	3	4	4
P11	PWA	3	2	2	1
P12	PWA	2	2	2	2
P13	PWA	3	3	3	3
P14	PWA	4	4	4	4
P15	PWA	4	2	3	1
P16	PWA	2	2	3	3
P17	PWA	3	2	2	2
P18	PWA	3	3	2	2
P19	PWA	3	2	2	2
P20	PWA	2	2	3	2

Table E2. Microlinguistic parameters of the narratives

ID	Group	CIUs per minute (*N*/min)	Word-level errors (N/N_{words})	Fillers (N/N_{words})	Ungrammatical EDUs (N/N_{EDUs})	Lexical diversity (MATTR)
P1	NBD	78.4	0.004	0.000	0.000	0.78
P2	NBD	78.7	0.012	0.004	0.000	0.83
P3	NBD	76.3	0.010	0.000	0.019	0.75
P4	NBD	82.1	0.012	0.000	0.000	0.77
P5	NBD	74.7	0.003	0.000	0.000	0.80
P6	NBD	84.0	0.007	0.000	0.000	0.78
P7	NBD	64.6	0.004	0.012	0.034	0.74
P8	NBD	66.7	0.020	0.005	0.022	0.76
P9	NBD	48.2	0.026	0.000	0.040	0.83
P10	NBD	64.9	0.010	0.000	0.000	0.77
P11	PWA	5.1	0.116	0.401	0.235	0.52
P12	PWA	8.6	0.120	0.168	0.097	0.66
P13	PWA	14.9	0.104	0.022	0.027	0.71
P14	PWA	44.7	0.067	0.034	0.037	0.72
P15	PWA	9.0	0.084	0.132	0.070	0.68
P16	PWA	17.8	0.036	0.009	0.093	0.71
P17	PWA	31.0	0.078	0.000	0.071	0.67
P18	PWA	26.3	0.074	0.068	0.095	0.68
P19	PWA	33.3	0.080	0.004	0.045	0.69
P20	PWA	26.2	0.127	0.020	0.093	0.64

Table E3. Macrolinguistic parameters of the narratives

ID	Group	Main events (*N*)	Structural disfluencies (N/N_{EDUs})	Relation set size (*N*)	Meta-comments (N/N_{EDUs})
P1	NBD	7	0.041	16	0.061
P2	NBD	7	0.018	16	0.036
P3	NBD	7	0.019	17	0.189
P4	NBD	7	0.103	18	0.118
P5	NBD	7	0.027	21	0.095
P6	NBD	5	0.039	22	0.078
P7	NBD	5	0.288	16	0.085
P8	NBD	7	0.078	26	0.289
P9	NBD	5	0.040	13	0.280
P10	NBD	7	0.027	23	0.411
P11	PWA	7	0.471	9	0.000
P12	PWA	4	0.290	8	0.097
P13	PWA	7	0.027	13	0.027
P14	PWA	7	0.074	10	0.037
P15	PWA	7	0.268	15	0.085
P16	PWA	3	0.209	11	0.279
P17	PWA	7	0.262	13	0.000
P18	PWA	7	0.324	16	0.019
P19	PWA	7	0.270	19	0.018
P20	PWA	7	0.140	12	0.093