

National Research University Higher School of Economics

as a manuscript

Mariya Khudyakova

**METHODS FOR ANALYSIS AND ASSESSMENT OF SPOKEN
DISCOURSE BY PEOPLE WITH SPEECH-LANGUAGE DISORDERS**

Dissertation Summary
for the purpose of obtaining
academic degree Doctor of Philosophy in Philology and Linguistics

Academic Supervisor:
Olga Dragoy
Doctor of Science

Moscow 2023

The dissertation was prepared at the HSE University.

Publications

Four publications were selected for the defense. In these publications, the authors describe the development of methods for analysis and assessment of spoken discourse by people with language impairments. The author of the dissertation is the first and corresponding author of the article (Khudyakova et al., 2023) in a Scopus-indexed journal and the sole author of the article (Khudyakova, 2020) published in a journal included in the list of high-level journals prepared by the HSE University. The author of the dissertation is the fourth and corresponding author in the article (Linnik et al., 2021) published in a journal included in Q1 of WoS and Scopus databases and the eighth author in the article (Ivanova et al., 2021) indexed in Scopus (Q1).

1. Ivanova, M. V., Akinina, Y. S., Soloukhina, O. A., Iskra, E. V., Buivolova, O. V., Chrabaszc, A. V., Stupina, E. A., Khudyakova, M. V., Akhutina, T. V., & Dragoy, O. (2021). The Russian Aphasia Test: The first comprehensive, quantitative, standardized, and computerized aphasia language battery in Russian. *PLOS ONE*, *16*(11), e0258946.
2. Linnik, A., Bastiaanse, R., Stede, M., & Khudyakova, M. (2021). Linguistic mechanisms of coherence in aphasic and non-aphasic discourse. *Aphasiology*, *36*(2), 123-146. <https://doi.org/10.1080/02687038.2020.1852527>
3. Khudyakova, M., Antonova, N., Nelubina, M., Surova, A., Vorobyova, A., Minnigulova, A., Gronskaya, N., Yashin, K., Medyanik, I., Shishkovskaya, T., Ryazanskaya, G., Zuev, A., & Dragoy, O. (2023). Discourse Diversity Database (3D) for Clinical Linguistics Research: Design, Development, and Analysis. *Bakhtiniana: Revista de Estudos Do Discurso*. <https://doi.org/10.1590/2176-4573e55885>
4. Khudyakova, M. (2020). Effect of Speaker's Fatigue on Features of Spoken Discourse. *The Russian Journal of Cognitive Science*, *7*(3), 78–88. <https://doi.org/10.47010/20.3.5>

Conference presentations and public demonstrations of the results

The main results and conclusions were presented in 5 conferences in oral and poster presentations:

1. General Theoretical and Typological Problems of Linguistics (Russia, online, 2022). Oral presentation: Corpus analysis of speech by patients before and after brain tumor resection.

2. The International Online Workshop on Language in Healthy and Pathological Aging (Spain, online, 2021). Oral presentation: Effect of age on lexical diversity in three discourse genres.
3. Linguistic Forum 2020: Language and Artificial Intelligence (Russia, online, 2020). Oral presentation: Speech characteristics depend on discourse elicitation task.
4. The Fifth Saint Petersburg Winter Workshop on Experimental Studies of Speech and Language (Night Whites, Russia, 2019). Poster presentation: Effect of speaker's fatigue and discourse genre on speech characteristics: a pilot study.
5. XI International interdisciplinary congress: Neuroscience for medicine and Psychology (Russia, 2015). Oral presentation: Developing a standardized test for language assessment in aphasia.

Introduction

The dissertation includes articles on the development and use of various methods of discourse assessment in clinical and research practice. One of the articles describes the development and standardization process of the Discourse Comprehension and Discourse Production subtests of the Russian Aphasia Test – a standardized language assessment tool for clinical practice. In the second article, a new scale for assessment of four aspects of discourse coherence is introduced, and the contribution of micro- and macrolinguistic parameters into coherence is investigated. The third article describes the rationale behind the data collection protocol of the Discourse Diversity Database (3D) corpus, and the clinical and normative sections of the corpus. The fourth article focuses on effects of fatigue on spoken discourse features.

In clinical linguistics, analysis of spoken discourse is considered an essential part of language assessment among populations with brain damage in language dominant hemispheres, since it allows us to evaluate language on both micro- (phonetic, lexical, syntactic) and macrolinguistic (discourse and pragmatics) levels (Bryant et al., 2017; Prins & Bastiaanse, 2004). Also, understanding coherent speech and telling stories or giving instructions are a fundamental part of human communication (Mar, 2004; Schank, 1995). That is why discourse comprehension and production subtests are included in many language assessment batteries (e.g. CAT, Swinburn, Porter, Howard, 2004; BDAE, Goodglass, Kaplan, Barresi, 2001; QAB, Wilson et al., 2018).

Development of discourse production and comprehension subtests can be challenging for several reasons. First, comprehension of a coherent text involves not only phonological, lexical-semantic, and syntactic processing, but also requires constructing inferences and understanding the connections between discourse elements, and creates a load on working memory (for a

review see Carpenter et al., 1995; Linda E Nicholas & Brookshire, 1995). That is why the text length, and the number of questions should be limited, while allowing for the task to discriminate between different levels of severity. For creation of a discourse production assessment task, it is important to choose the evaluation method (annotation scheme or a rating scale) that would be low time-consuming on the one hand, while remaining objective and yielding high test-retest and inter-rater reliability. The first chapter of the thesis describes the Discourse Comprehension and Discourse Production subtests of the Russian Aphasia Test (RAT), as well as its psychometric properties.

Rating scales are a common method for evaluation of spoken discourse, especially its macrolinguistic properties, both in clinical practice and fundamental research. However, we still lack understanding of the connection between assessment and interpretation of discourse as a whole by the reader or listener, and the quantitative characteristics of speech on micro- and macro-linguistic levels (e.g. number of errors, lexical diversity, etc.). Chapter 2 describes the study on the connection of different aspects of discourse coherence measured with a newly developed rating scale and a set of linguistic features extracted from annotated narratives by people with aphasia and neurologically healthy individuals (NHI).

For multidimensional analysis of spoken discourse by people with language impairments, annotated corpora are a valuable and important instrument. Corpus analysis allows to investigate different sources of variability in speech features in various clinical populations and healthy speakers. In clinical corpus linguistics, the standards for the collection and analysis of speech samples are established and many corpora of speech by people with various neurological and psychiatric disorders exist, for example, the most well-known corpora from the TalkBank collection (<https://talkbank.org>, (MacWhinney, 2007), the Cambridge Cookie-Theft Corpus [Williams et al, 2010], the Greek Corpus of Aphasic Discourse (Varlokosta, 2016) etc. However, there is a lack of large clinical corpora for Russian language. In Chapter 3, we describe the newly developed Discourse Diversity Database (3D), the rationale behind the data collection procedure and the different sub-sections of the corpus. Also, we present the study of effects of fatigue on speech features based on the analysis of one of the 3D sub-sections.

The aim of the thesis is to describe a set of new methods for assessing discourse comprehension and production abilities for clinical and research purposes. The studies collected in the current thesis describe a range of instruments that can be applied to speech by people with language impairments in order to provide evaluation of discourse abilities as a whole or focus on specific features of speech. The lack of modern standardized and normed methods for discourse assessment, as well as large clinical corpora, especially for Russian, determines the *relevance* of the study. The *object of the study are* discourse production and comprehension abilities of

people with various neurological and psychiatric disorders. The *subject of the study* is the assessment of discourse production and comprehension with newly developed rating scales and the analysis of micro- and macrolinguistic parameters of speech in various clinical populations.

Research *novelty*:

- RAT is the first standardized comprehensive test in Russian with subtests for assessment of discourse comprehension and discourse production, developed based on the psycholinguistic parameters.
- The new rating scale allowed to analyze the concept of discourse coherence and its connection with the micro- and macrolinguistic parameters of discourse, and to address the contradictory findings on discourse coherence in aphasia in the literature.
- Discourse Diversity Database (3D) is the first large corpus in Russian containing discourse samples by different discourse types by people with various neurological and psychiatric disorders, and healthy speakers in different functional states.

The *theoretical significance* of the study:

- We found that all aspects of coherence are significantly lower in film retellings by people with aphasia, and that different sets of micro- and macrolinguistic parameters extracted from the discourse samples contribute to the coherence ratings in different aspects.
- We found that temporal characteristics of speech are affected by the level of the speaker's fatigue.

The *practical significance* of the study:

- Discourse Comprehension and Discourse Production subtests of RAT were created and standardized.
- Rating scales for the aspects of discourse coherence were created.
- The annotated 3D corpus was developed with protocols for data collection in different clinical and healthy populations.

The main results of the study and provisions for the defense:

1) Discourse Comprehension and Discourse Production subtests of RAT were created according to modern psycholinguistic theory and standardized a PWA group and a control group. The results of the study showed that this instrument meets psychometric standards and makes it possible to distinguish between people with and without aphasia.

2) A new rating scale for four aspects of discourse coherence revealed lower coherence scores in the PWA group. Different sets of micro- and macrolinguistic properties contribute to each of the coherence aspects.

3) The Discourse Diversity Database (3D) is a corpus of speech samples by people with various neurological and psychiatric disorders. For elicitation of three discourse types, three types of tasks were selected. The annotation scheme allows to extract micro- and macrolinguistic parameters.

4) The results of the pilot study revealed the variability of phonetic and temporal characteristics of speech in different functional states and depending on the discourse type.

1. Development and standardization of Discourse Comprehension and Discourse Production subtests of the Russian Aphasia Test

Paper selected for the defence: (Ivanova et al., 2021)

The Russian Aphasia Test (RAT) is a standardized test for assessment of speech production and comprehension. Unlike screening tools such as the Aphasia Rapid Test (ART) (Buivolova et al., 2020) or Token Test (Bastiaanse et al., 2015), RAT is a comprehensive tool that allows to assess the severity of speech deficit at various language levels. RAT consists of 13 subtests assessing speech comprehension (Pseudo-word Comprehension, Lexical Decision, Noun Comprehension, Verb Comprehension, Sentence Comprehension, Discourse Comprehension), repetition (Pseudo-word Repetition, Word Repetition, Sentence Repetition) and speech production (Object Naming, Action Naming, Sentence Production, Discourse Production). Stimuli for each subtest were created based on modern psycholinguistic theories.

In this study, we present the experience of developing and standardizing subtests for assessment of discourse production and comprehension. The Discourse Comprehension subtest consists of a short story about a cat and 16 statements that match or do not match the content of the text. The text is pre-recorded and presented by ear, then the follow-up statements are presented one at a time, and the participants responds whether this statement corresponds to the story. The Cat Story was created specifically for the RAT and does not repeat well-known plots from cinema, literature, folk culture, etc., so that the participants cannot rely on heuristics and general knowledge about the world when answering questions (Caplan & Evans, 1990; Ferstl et al., 2005). The Cat Story is 156 words long (mean content word frequency = 261.32 ipm; median = 94.7 ipm) and contains 31 clauses (clause length in words: M = 5.03, min = 2, max = 10), all in canonical word order; 4 clauses are relative object clauses. The follow-up statements are in a pseudo-randomized order and are organized in pairs. Each pair consists of two statements relating to the same fact from history, one of which is true, and the other is false. Each pair of statements refers to either a main plot or a detail, and either explicit or implicit information;

similar question types are presented in the Discourse Comprehension Test (DCT; Nicholas & Brookshire, 1993). A point is given for each pair of responses to statements if both answers are correct; thus, the maximum score for the subtest is 8 points.

The Discourse Production subtest contains one elicitation stimulus, that is, the single Bike picture designed specifically for the test. Similarly to the Cat Story, the plot of the Bike picture does not resemble any well-known story plot. The instruction motivates the patient to produce a narrative, that is, a structured discourse where the units are organized by timeline, and not a set of descriptions of the elements on the picture (Olness, 2006). The picture-based story is evaluated on four rating scales (informativeness, fluency, grammaticality, paraphasias) from 1 (severe deficit) to 5 (no deficit). Using several rating scales is considered to be the preferable solution for discourse assessment in clinical practice since it allows to evaluate different linguistic levels while remaining relatively low time-consuming (Prins & Bastiaanse, 2004). The total score for the subtest is the sum of scores on each of the four scales. The score of 0 is given if the patient failed to produce at least five relevant content words in one minute.

In order to test the psychometric properties of RAT, the validity and reliability of the test were tested, and the norms were collected. 106 neurologically healthy individuals (77 females; 19-86 years old; mean age – 49.9, SD = 18.4), and 85 people with aphasia (26 females; 25-80 years old; mean age – 57.6, SD = 12.1) participated in the study. For evaluation of test-retest reliability, a group of 20 additionally recruited patients (11 females, 39-82 years old; mean age – 58.8, SD = 12.9) completed the Discourse Production and Discourse Comprehension subtests. Inter-rater reliability was calculated for the Discourse Production subtest based on 20 picture-based stories rated by two independent raters, both speech-language pathologists. Internal reliability was calculated for the Discourse Production subtest. The results are presented in Table 1.

Table 1. Psychometric parameters of the Discourse Comprehension and Discourse Production subtest of RAT

		Discourse Comprehension	Discourse Production
Internal reliability	Cronbach alpha	0,83	-
	95% CI	[0,776, 0,885]	-
	ICC	-	0,833
Inter-rater reliability	95% CI	-	[0,609-0,932]
	Rater 1 (Mean)	-	57,0%
	Rater 2 (Mean)	-	62,5%
Test-retest reliability	ICC	0,784	0,71

95% CI	[0,53, 0,909]	[0,394-0,875]
Test 1 (Mean)	70,0%	75,5%
Test 2 (Mean)	76,9%	75,2%

The scores for the subtests were calculated in two age groups: younger adults (18-59 years old) and older adults (60+ years old), see Table 2 for details. In the NHI group, the two age cohorts performed comparably in the Discourse Production subtest, while the in the Discourse Comprehension subtest the older NHI participants obtained significantly lower scores (independent-samples two-tailed Welch t-tests, $t = -3.47$, $p_{adj} < .05/13 = .0038$). PWA obtained significantly lower scores compared to the NHI group in both age cohorts and in both subtests (based on independent-samples two-tailed Welch t-tests with $p_{adj} < .05/26 = .0019$).

Table 2. Scores for the Discourse Comprehension and Discourse Production subtests across participant groups

Subtest	Age	Group	N	Range	Mean (SD)	Median (IQR)
Discourse Comprehension	Young	NHI	69	62.5 - 100	94.36 (8.74)	100 (12.5)
		PWA	40	0 - 100	63.44 (31.82)	75 (37.5)
	Elderly	NHI	37	50 - 100	86.15 (14.06)	87.5 (25)
		PWA	38	0 - 100	56.58 (32.33)	62.5 (50)
Discourse Production	Young	NHI	69	75 - 100	90.65 (6.3)	90 (5)
		PWA	38	0 - 95	46.97 (27.67)	50 (33.75)
	Elderly	NHI	35	75 - 100	91.14 (5.83)	90 (5)
		PWA	38	0 - 80	45 (30.78)	57.5 (70)

2. Linguistic mechanisms of discourse coherence in aphasia

Paper selected for the defence: (Linnik et al., 2021)

Discourse units are connected with each other in a certain manner and are organized according to a certain hierarchy and rules. One of the important features of discourse is coherence, that is, the degree to what the text is comprehensible to the reader or listener. This is a mental phenomenon which includes establishing relations between the speaker and the addressee (whether a real person or a mental representation of them) and establishing a common context,

reference, and thematic structure (Jucker, 1997). The linguistic mechanisms of coherence impairment in aphasia and their connection with deficits on other linguistic levels are yet understudied. One of the major reasons for that is the variety of definitions of coherence and annotation methods used by different researchers, as well as the great involvement of the listener (or reader) in the interpretation of coherence. In this study we designed rating scales for evaluation of the four aspects of coherence in order to answer two research questions: 1) is discourse coherence impaired in aphasia? 2) what micro- and macrolinguistic parameters affect these aspects of coherence?

The materials for the study were the retellings of the Pear film (Chafe, 1980). Two groups of people participated in the study: ten PWA (4 females; 40-73 years old, mean age – 56.4) and ten NHI (5 females; 42-84 years old; mean age – 58.7). The audio recordings were rated by three raters according to a specifically designed scale. Each of the four aspects of coherence – informativeness, clarity, understandability, connectedness – were rated from 1 (completely incoherent) to 4 (completely coherent). The retellings were annotated in CHAT format (Codes for the Human Analysis of Transcripts; (MacWhinney, 2010)) and segmented into elementary discourse units (EDUs) according to criteria formulated by (Carlson & Marcu, 2001). In addition, the transcripts were annotated according to the Rhetorical Structure Theory (Mann & Thompson, 1987), and a number of mentioned main events was calculated for each story (Wright et al., 2010).

Nine linguistic features were extracted from the annotations: five micro-linguistic parameters (correct information units per minute, word-level error ratio, filler ratio, ratio of ungrammatical EDUs, lexical diversity) and four macro-linguistic ones (number of main events, ratio of structural disfluencies, rhetorical relation set, ratio of meta-comments). For the purposes of classification analysis, ratings on the four-point scale were transformed into binary ratings (0 – incoherent discourse, 1 – coherent discourse). Classification was performed with the random forests method (Breiman, 2001) for all the 20 discourse samples.

Between-group comparison with one-way permutation tests revealed significantly lower scores on all aspects of coherence in PWA group as compared to the NHI group. Results of the between-group and classification analyses, and the top predictors are presented in Table 3.

In this study coherence was considered as a quality of language in use, co-constructed by a speaker and the addressee rather than solely the internal property of a discourse. The results showed that different combinations of micro- and macrolinguistic variables were relevant for different aspects of discourse.

Table 3. Discourse coherence aspects: results of the between-group and classification analyses

Between-group comparison			
	p_{adj}	Median difference in rating	
Informativeness	0.04	1	
Clarity	0.04	1	
Understandability	0.002	1.5	
Connectedness	0.002	2	

Classification analysis				
	Accuracy	Kappa	OOB error	Top predictors
Informativeness	0.84	0.355	15%	<ul style="list-style-type: none"> • Main events • Relation set • Meta-comments • Ungrammatical EDUs • Word-level errors • Ungrammatical EDUs
Clarity	0.8	0.57	20%	<ul style="list-style-type: none"> • CIU per minute • Lexical diversity • Word-level errors • Meta-comments
Understandability	0.8	0.34	20%	<ul style="list-style-type: none"> • Structural disfluencies • Lexical diversity • Ungrammatical EDUs
Connectedness	0.95	0.89	5%	<ul style="list-style-type: none"> • Lexical diversity • Word-level errors

3. Discourse Diversity Database

3D contains speech samples of three different types: picture-based narratives, personal stories, and picture-based instructions. For the elicitation of narratives, we chose three comics by Herluf Bidstrup; for personal story elicitation we created three questions about the best or the most memorable trip, gift, or party; for picture-based instruction elicitation we chose three IKEA furniture assembly instructions each containing eight pictures and no text.

3.1 Discourse Diversity Database (3D) for clinical linguistics research

Paper selected for the defense: (Khudyakova et al., 2023)

The 3D corpus contains several sections with discourse samples by adults with neurologic and psychiatric diagnoses, and neurotypical adults. The collection consists of discourse samples from two clinical groups: patients with brain tumors and people with

psychiatric disorders, and three normative sub-sections: self-reported NHI, ages 18-80, neurologically healthy adults with data from psychiatric questionnaires, and self-reported NHI recorded at two time points: in an active state and in the state of fatigue. Data collection is ongoing. The summary of the subsections is presented in Table 4.

Table 4. Sections of the 3D corpus

	Sections				
	Clinical		Age-balanced self-reported norm	Normative	
	Neurosurgery	Psychiatric		Psychiatric norm	Functional states
N participants	87	107	86	76	10
N time points	3	1	1	1	2
Age	M=49.7, SD=14.6	M=28.8, SD=4.3	18-29 y.o. (M = 21.2, SD = 2.6); 30-49 y.o. (M = 38.1, SD = 6.6); 50-64 y.o. (M= 57, SD = 3.8); 65+ y.o. (M = 72, SD = 7.0)	M=23.9, SD=4.3	M=28.80, SD=2.86
Diagnoses	Brain tumors	Schizophrenia spectrum disorders, affective disorders	no	possibly	no
Meta-data	MRI; RAT; CETI	ICD-10; PANSS; HDRS		SCL-90-R; AMSR; QIDS-16SR	Test of differential self-evaluation of one's functional state

Annotation of the narratives was performed in ELAN (Wittenburg et al., 2006) on multiple tiers. The *Transcript* tier is aligned with the media files and contains an orthographic transcript of the recorded speech and pause annotation. The discourse is segmented into *elementary discourse units* (EDUs), where an EDU roughly equals a clause. Utterances include a main clause with all its subordinate clauses. Lemmatization and part-of-speech information, as well as annotation of errors, false-starts, repetitions, semantically empty words and automatized expressions are presented on several tiers. Each EDU is annotated as one of five macrocomponents, either

containing on-topic information (Mainline, Background and Comment EDUs), or being off-topic (Meta-comments and Regulatory EDUs).

3.2 Effect of speaker's fatigue on speech parameters

Paper selected for the defense: (Khudyakova, 2020)

Spontaneous speech analysis is a common method for assessing language impairments in clinical practice and fundamental research. However, there is evidence that many speech parameters, such as speech rate and voice characteristics vary depending on the level of stress, fatigue, or cognitive load (Quatieri et al., 2015; Rao et al., 2020; Sloboda et al., 2018). In order to assess the severity of language deficit in case of language impairment, it is necessary to know the variability of speech parameters in healthy speech. We ran a pilot study of speech variability depending on the level of fatigue on speech samples that were later included in the 3D corpus.

Ten participants took part in the online study (8 females; age range – 23-33 years old; mean age – 28.8). Each participant completed two sessions of the study – one in an active state and one in a tired state. The participants completed the Test of differential self-evaluation of one's functional state (Doskin et al., 1973), and then recorded a picture-based narrative, a personal story and an instruction (the stimuli were balanced across two experimental lists).

We performed a mixed effect model analysis of speech parameters with fixed effects of functional state, discourse type and their interaction and random effect of participant. We found significant effects of the speaker's functional state on speech rate ($p_{adj} = 7.20e-06$), and significant effects of discourse type on speech sample length in seconds ($p_{adj} = 3.57e-04$) and words ($p_{adj} = 0.038$), speech rate articulation rate ($p_{adj} = 1.23e-07$), words per EDU ($p_{adj} = 0.006$), lexical diversity ($p_{adj} = 0.042$), and content word ratio ($p_{adj} = 0.014$). No interactions were significant. The results of the pilot study revealed the variability of phonetic and temporal characteristics of speech in different functional states and depending on the discourse type.

4. Conclusion

The articles included in the dissertation are united by the topic of development of new methods and tools for research and assessment of speech comprehension and production at discourse level in people with neurological and psychiatric disorders. The methods presented in the articles can be used for language assessment in clinical practice and as instruments of fundamental research in clinical linguistics.

In Chapter 1, the Discourse Comprehension and Discourse Production subtests of RAT were described, as well as the rationale behind the tasks. Results of the norming and reliability measurement showed that both subtests have the necessary psychometric properties and can

distinguish between people with and without aphasia.

Chapter 2 focused on the study of linguistic mechanisms of discourse coherence. We presented a new rating scale for assessment of four aspects of coherence and demonstrated that all the four aspects are impaired in aphasia. Also, we analyzed the contribution of different micro- and macrolinguistic parameters extracted from the annotation of the discourse samples into the coherence aspect scores. The results showed that the scores on each of the aspects of coherence are influenced by different combinations of the parameters. These findings can explain the previous contradictory findings on coherence of discourse in aphasia.

In Chapter 3, the 3D corpus was introduced in its current state. We described the three elicitation tasks and the data collection procedure, and the clinical and normative sub-sections of the corpus. Also, we presented the results of the pilot study on one of the sections that proved significant effects of discourse type on speech parameters. Also, we found the effect of speaker's fatigue on speech rate, which showed the variability of speech in healthy speakers.

This thesis was motivated by the need for various methods for assessment of discourse in clinical linguistics. We presented several methods for clinical discourse assessment in Russian, as well as the annotated corpus of speech. Further research of the author of the thesis will include the two questions that were started in this thesis: 1) the relation between discourse metrics extracted from the annotation of speech and evaluation of discourse properties by a listener on a rating scale; and 2) exploration of the variability of discourse parameters depending on the speaker's functional state in people with and without neurological and psychiatric disorders. The dissertation is interdisciplinary and it connects the traditions of speech language pathology, clinical linguistics and corpus linguistics.

References

- Bastiaanse, R., Raaijmakers, S., Satoer, D., & Visch-Brink, E. (2015). The Multilingual Token Test. *Http://Dx.Doi.Org/10.1080/02687038.2015.1121710*, 30(4), 508–508.
<https://doi.org/10.1080/02687038.2015.1121710>
- Breiman, L. (2001). Random Forests. *Machine Learning 2001 45:1*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, 31(10), 1105–1126.
<https://doi.org/10.1080/02687038.2016.1239013>
- Buivolova, O., Vinter, O., Bastiaanse, R., & Dragoy, O. (2020). The Aphasia Rapid Test: adaptation and standardisation for Russian.
Https://Doi.Org/10.1080/02687038.2020.1727836, 35(5), 730–744.

<https://doi.org/10.1080/02687038.2020.1727836>

- Caplan, D., & Evans, K. L. (1990). The effects of syntactic structure on discourse comprehension in patients with parsing impairments. *Brain and Language*, 39(2), 206–234. [https://doi.org/10.1016/0093-934X\(90\)90012-6](https://doi.org/10.1016/0093-934X(90)90012-6)
- Carlson, L., & Marcu, D. (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 2, 1–87. <ftp://128.9.176.20/isi-pubs/tr-545.pdf>
- Carpenter, P. A., Miyake, A., & Just, M. A. (1995). Language comprehension: sentence and discourse processing. *Annual Review of Psychology*, 46(1), 91–120. <https://doi.org/10.1146/annurev.psych.46.1.91>
- Chafe, W. (1980). *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production* (W. Chafe (ed.)). Ablex.
- Doskin, V. A., Lavrent'eva, N. A., Miroshnikov, M. P., & Sharai, V. B. (1973). A test of differential self-evaluation of one's functional state. *Voprosy Psichologii*.
- Ferstl, E. C., Walther, K., Guthke, T., & von Cramon, D. Y. (2005). Assessment of Story Comprehension Deficits After Brain Damage. *Journal of Clinical and Experimental Neuropsychology*, 27(3), 367–384. <https://doi.org/10.1080/13803390490515784>
- Goodglass, H., Kaplan, E., & Barresi, B. (2001). *BDAE-3: Boston Diagnostic Aphasia Examination* (Third Edit). Lippincott Williams & Wilkins.
- Ivanova, M. V., Akinina, Y. S., Soloukhina, O. A., Iskra, E. V., Buivolova, O. V., Chrabaszcz, A. V., Stupina, E. A., Khudyakova, M. V., Akhutina, T. V., & Dragoy, O. (2021). The Russian Aphasia Test: The first comprehensive, quantitative, standardized, and computerized aphasia language battery in Russian. *PLOS ONE*, 16(11), e0258946. <https://doi.org/10.1371/JOURNAL.PONE.0258946>
- Jucker, A. H. (1997). Coherence in spontaneous text. *Journal of Pragmatics*, 27(3), 400–402. [https://doi.org/10.1016/S0378-2166\(97\)83633-8](https://doi.org/10.1016/S0378-2166(97)83633-8)
- Khudyakova, M. (2020). Effect of Speaker's Fatigue on Features of Spoken Discourse. *The Russian Journal of Cognitive Science*, 7(3), 78–88. <https://doi.org/10.47010/20.3.5>
- Khudyakova, M., Antonova, N., Nelubina, M., Surova, A., Vorobyova, A., Minnigulova, A., Gronskaya, N., Yashin, K., Medyanik, I., Shishkovskaya, T., Ryazanskaya, G., Zuev, A., & Dragoy, O. (2023). Discourse Diversity Database (3D) for Clinical Linguistics Research: Design, Development, and Analysis. *Bakhtiniana: Revista de Estudos Do Discurso*. <https://doi.org/10.1590/2176-4573e55885>
- Linnik, A., Bastiaanse, R., Stede, M., & Khudyakova, M. (2021). Linguistic mechanisms of coherence in aphasic and non-aphasic discourse. *Aphasiology*, 36(2), 123–146. <https://doi.org/10.1080/02687038.2020.1852527>

- MacWhinney, B. (2007). The Talkbank Project. *Creating and Digitizing Language Corpora*, 163–180. https://doi.org/10.1057/9780230223936_7
- MacWhinney, B. (2010). Part 1: The CHAT Transcription Format. In *The CHILDES Project: Tools for Analyzing Talk*. <https://doi.org/10.1111/1460-6984.12101/abstract>
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: Description and construction of text structures*. Springer Netherlands.
- Mar, R. A. (2004). The neuropsychology of narrative: story comprehension , story production and their interrelation. *Neuropsychologia*, 42(10), 1414–1434. <https://doi.org/10.1016/j.neuropsychologia.2003.12.016>
- Nicholas, L E, & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36(2), 338–350.
- Nicholas, Linda E, & Brookshire, R. H. (1995). Comprehension of Spoken Narrative Discourse by Adults With Aphasia, Right-Hemisphere Brain Damage, or Traumatic Brain Injury. *American Journal of SpeechLanguage Pathology*, 4(3), 69–81. <http://ajslp.asha.org/cgi/content/abstract/4/3/69>
- Olness, G. S. (2006). Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology*, 20(2/3/4), 175–187. <https://doi.org/10.1080/02687030500472710>
- Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075–1091. <https://doi.org/10.1080/02687030444000534>
- Quatieri, T. F., Williamson, J. R., Smalt, C. J., Patel, T., Perricone, J., Mehta, D. D., Helfer, B. S., Ciccarelli, G., Ricke, D., Malyska, N., Palmer, J., Heaton, K., Eddy, M., & Moran, J. (2015). *Vocal biomarkers to discriminate cognitive load in a working memory task*.
- Rao, H. M., Smalt, C. J., Rodriguez, A., Wright, H. M., Mehta, D. D., Brattain, L. J., Edwards, H. M., Lammert, A., Heaton, K. J., & Quatieri, T. F. (2020). Predicting Cognitive Load and Operational Performance in a Simulated Marksmanship Task. *Frontiers in Human Neuroscience*, 14, 222. <https://doi.org/10.3389/fnhum.2020.00222>
- Schank, R. C. (1995). *Tell me a story: Narrative and intelligence*. Northwestern University Press.
- Sloboda, J., Lammert, A., Williamson, J., Smalt, C., Mehta, D. D., Curry, C. I., Heaton, K., Palmer, J., & Quatieri, T. (2018). Vocal Biomarkers for Cognitive Performance Estimation in a Working Memory Task. *Interspeech 2018, 2018-Septe*, 1756–1760. <https://doi.org/10.21437/Interspeech.2018-2418>
- Swinburn, K., Porter, G., & Howard, D. (2004). *CAT: comprehensive aphasia test*.
- Varlokosta, S. (2016). A Greek Corpus of Aphasic Discourse: Collection, Transcription, and

Annotation Specifications. *LREC 2016 Workshop Resources and Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments (RaPID-2016)*, May, 14–21.

- Williams, C., Thwaites, A., Buttery, P., Geertzen, J., Randall, B., Shafto, M., Devereux, B., & Tyler, L. K. (2010). The Cambridge Cookie-Theft corpus: A corpus of directed and spontaneous speech of brain-damaged patients and healthy individuals. *LREC 2010. Proceedings of the 7th International Conference on Language Resources and Evaluation, November 2015*, 2824–2830. <http://www.lrec-conf.org/proceedings/lrec2010/summaries/327.html>
- Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLOS ONE*, *13*(2), e0192773. <https://doi.org/10.1371/journal.pone.0192773>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556–1559. <https://tla.mpi.nl/tools/tla-tools/elan/>
- Wright, H. H., Capilouto, G. J., Wagovich, S. A., Cranfill, T. B., & Davis, J. E. (2010). Development and reliability of a quantitative measure of adults' narratives. <Http://Dx.Doi.Org/10.1080/02687030444000732>, *19*(3–5), 263–273. <https://doi.org/10.1080/02687030444000732>